Assessment of the low visibility predictions At Schiphol Airport

TU Delft
Msc. Thesis
Environmental Engineering
S.D.A. van Beek

TUDelft

knowledge &
development
centre
Mainport Schiphol

Performance of the Terminal Aerodrome Forecast Guidance

Assessment of the low visibility predictions at Schiphol Airport

by

S.D.A. van Beek

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Tuesday October 12, 2021 at 09:00 AM.

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

TUDelft

Abstract

Airport operations are highly dependent on safe weather conditions for their daily operations. Especially the visibility conditions are crucial. Low Visibility Procedures (LVP phase M, A, B and C) are defined at Schiphol Airport, resulting in capacity and runway use restrictions. These Low Visibility Procedures are based on a post-processing model of a Numerical Weather Prediction (NWP): the Terminal Aerodrome Forecast Guidance (TAFG). This model provides a probability for the occurrence of a given LVP phase, based on thresholds of the Runway Visual Range (visibility including back luminescence at airports) and the ceiling (cloud base height). Currently, the TAFG is operational for three NWP models (HARMONIE, HIRLAM and ECMWF). Since November 2020, HARMONIE TAFG is used which has a higher horizontal resolution resulting in a more detailed spatial variability and a better representation of convective clouds than its predecessor HIRLAM TAFG. ECMWF TAFG is available with more mid- to long-range forecasts, but this model is not often used by the aviation meteorologists. Inaccuracies remain in the LVP forecasts by NWP models due to unknowns in the physics of fog and observations.

This research aims to study the performance of the low visibility forecasts of the three TAFG models used at Schiphol Airport. This research found that the HIRLAM TAFG has a better performance than the HARMONIE TAFG. This result is caused by the larger training data-set that is used for the HIRLAM TAFG (20 years) compared the training data-set of for the HARMONIE TAFG (3 years). The ECMWF TAFG has a better performance for the severe LVP phases B and C. The ECMWF TAFG also has a better performance for all LVP phases from a lead time of 8 hours or higher, compared to the HIRLAM TAFG and HARMONIE TAFG. This is remarkable, since ECMWF is not often used for airport operations. However, the standard error of the performance indicator of the ECMWF TAFG is higher compared to the other models and, therefore, a larger data-set should be studied to confirm these findings.

Different locations at Schiphol Airport and the influence of their spatial variability of the surface characteristics on the TAFG performances are studied. The furthest location, the Polderbaan, has more water availability due to the grass surrounding the runway compared to the concrete surface characteristics of the other locations. This resulted in a higher fog occurrence. All the TAFG models have a lower performance in forecasting fog and LVP phases for this location. The spatial variability of the surface characteristics of the other observation locations did not result in different performances of the TAFG models.

The 2 meter temperature appeared to be a potential cause for the inaccuracies in the performance of the forecasts of the LVP phases by the TAFG models. This was due to an overestimation in temperature leading to an underestimation of fog and LVP occurrence. However, one should keep in mind that, even if the temperature, humidity and wind conditions are simulated correctly, this does not guarantee correct fog simulation. Fog forms due to microphysical interactions that are complex and the formation is also very dependent on the aerosol composition of the air, which are not studied.

The main limitation of this research is the size of the TAFG output data-set. This was caused by a limited operational use of the HARMONIE TAFG from November 2020. A longer data-set should be studied in order to draw further conclusions on the performance of the TAFG in forecasting Low Visibility Procedures at Schiphol Airport. Further research into the performance of the ECMWF TAFG would be especially interesting. This model showed the most promising results for severe fog occurrence and long lead times and could result in improvement in fog and LVP phase forecasts that are of great value for operational use at Schiphol Airport.

List of Figures

List of Tables

List of Acronyms

LVP   Low Visibility Procedure

NWP   Numerical Weather Prediction

HARMONIE HIRLAM ALADIN Reseach on Mesoscale Operational NWP in Euromed

HIRLAM   High-Resolution Limited-Area Model

ECMWF   European Centre for Medium-Range Weather Forecasts

SKV   Schiphol Kans Verwachting

TAF   Terminal Aerodrome Forecast

TAFG   Terminal Aerodrome Forecast Guidance

CCN   Cloud Condensation Nuclei

LWC   Liquid Water Content

METAR   Meteorological Aviation Routine Weather Reports

MOR   Meteorological Optical Range

RVR   Runway Visual Range

KNMI   Dutch meteorological institute

BSS   Brier Skill Score

ROC   Receiver Operating Characteristic

rhs   right hand side

lhs   left hand side

Contents

1

Introduction

An aircraft is dependent on safe weather conditions in order to land and takeoff safely. These conditions are, among others, wind, fog and rainfall. Aircraft operations are based on thresholds for these weather conditions, to minimize the risk of accidents (van Dinther et al., 2015) and to maximize the efficiency of airport operations.

Fog is the main cause for reduced visibility, with the highest occurrence and highest severity of low visibility (Izett et al., 2019). Low visibility conditions have a consequence on the operational capacity of an airport (Van Der Velde et al., 2010). It can lead to delays while travelling, but can also lead to disasters in aviation (Herman and Schumacher, 2016). To minimize these risks and lower operational losses, it is important to have an accurate forecasting of the low visibility conditions at Schiphol Airport.

At Schiphol Airport there are five phases of low visibility procedures (LVPs). These visibility conditions are determined by either the clarity of the atmosphere (visibility) or the height of the cloud-base (ceiling). The severity of the forecasted low visibility procedure, determined by the two thresholds, will result in adapted airport traffic operations (de Rover et al., 2008). The adapted airport traffic operations as a result of the occurrence of a certain LVP phase can be seen in Table 1.1. It can be seen that a LVP phase can result in a restriction of the number of aircraft movements per hour (capacity) and on the number of runways that can be used at the airport (flow restrictions).

LVP phase	Visibility		Cloud base (ceiling)	Capacity restrictions [Movements/hour]	Flow restrictions
Good	$VIS \geqslant 5km$	and	$CLB \geqslant 1000ft$	max 104/108	No restrictions
Phase M	$1.5km \leqslant VIS \leqslant 5km$	or	$300 \leqslant CLB \leqslant 1000ft$	max 104/108	No restrictions
Phase A	$550m \leqslant VIS \leqslant 1500m$	or	$200 \leqslant CLB \leqslant 300ft$	max 70	In general no restrictions
Phase B	$350m \leqslant VIS \leqslant 550m$	or	$CLB \leqslant 200ft$	max 60-70	Restrictions
Phase C	$200m \leqslant VIS \leqslant 350m$			max 47	Restrictions
Phase D	$VIS \leqslant 200m$			max 36	Restrictions

Table 1.1: The table shows the determination of the LVP phases that follow from the combination of visibility and cloud base thresholds. The capacity and flow restrictions are in force with the LVP phase prediction and are therefore included in this table.

Statistics show that visibility is the second largest, after extreme wind conditions, cause for accidents in airport operations (van Dinther et al., 2015). Reduced visibility is often caused by fog, dust and ash (Tardif and Rasmussen, 2007; Hodges and Pu, 2016). This research will focus on fog based low visibility procedures, since this is the largest cause for low visibility procedures at Schiphol airport.

Fog consists of suspended water particles (droplets) that cause a reduction of the horizontal visibility to below 1 km (OFCM, 2017). There are different types of fog events leading to reduced visibility. Research by Tardif and Rasmussen (2007) reduces the fog classifications to five main types: lowering stratus event, radiation fog, advection fog, precipitation fog and morning evaporation fog. Research

at Charles de Gaulle Airport in Paris, shows that 78% of the low visibility procedures are caused by the first three events: lowering stratus event, radiation fog and advection fog (Roquelaure et al., 2009). These three types of fog are also the most common in The Netherlands (Izett et al., 2019) and the types of fog considered in this research. A lowering stratus event is the lowering of a cloud. Radiation fog is a result of radiative cooling of the lower part of the atmosphere, leading to condensation of water vapor above the surface. Advection fog forms when warm moist air advects over a cold surface, resulting in cooling of the air mass and condensation of water vapor.

A large amount of research has been done on fog, in the understanding of the formation and microstructure of fog, many uncertainties still remain. The first cause is the inability to fully comprehend the microphysics of the formation of fog. When these microphysics are not fully understood, it can result in improper parameterization. An example of this is the inaccurate vertical structure of fog as predicted by numerical weather models (Gultepe et al., 2007). A second cause is the difficulty to accurately observe fog. Fog is difficult to measure due to the small droplet size of the liquid water in fog and the mixed phases of water present in fog formations (Gultepe et al., 2008). It is difficult to verify model output, without accurate measurements.

Also, much research has been performed on visibility forecasting models including fog formation processes. Still a lot of challenges in the numerical weather predictions (NWP) when evaluating the forecast performances remain. The model previously used for the forecasting at Schiphol Airport, High-Resolution Limited-Area Model (HIRLAM), was able to model the beginning of fog formation in the case of severe radiation fog. However, HIRLAM was not sufficient for the lowest heights in the atmosphere and therefore the fog in the model was dispersed, leading to underestimations in the forecasts (Van Der Velde et al., 2010) compared to observations. The HARMONIE NWP model (HIRLAM ALADIN Research on Mesoscale Operational NWP in Euromed), currently used for fog predictions at Schiphol Airport, also has low forecasting performance when forecasting radiation fog. This is partly due to the inability to simulate low surface temperatures because of the inability to capture the mesoscale phenomena or sub-gridscale processes (Román-Cascón et al., 2019). The various models are still not able to provide an accurate fog forecast. This is due to uncertainties within the physics of the model and the model algorithms (Gultepe et al., 2008; Roquelaure and Bergot, 2007).

Schiphol Airport uses a numerical guidance system to extend numerical weather prediction model outputs and correct for their inaccuracies, called Terminal Aerodrome Forecast Guidance (TAFG). This TAFG converts the deterministic variables of the NWP output to probabilistic forecasts if an event will happen between 0 and 100. This way, the uncertainties involving forecasts are better captured, resulting in more informative forecasts for aviation meteorologists and users. The aviation meteorologist uses the TAFG as input for the final Schiphol Kans Verwachting (SKV). The SKV is the probabilistic forecast on which the airport operations are based. The steps involved to produce the SKV are visible in Figure 1.1
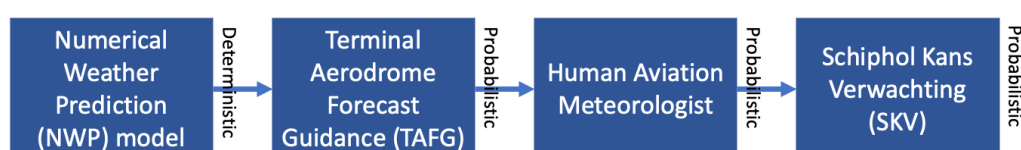
Figure 1.1: Realization of the Schiphol Kans Verwachting (SKV). It shows the general process of the NWP deterministic output on the left to the probabilistic output of the SKV on the right.

The TAFG model uses on sight and upstream observation data in a stepwise forward multiple regression algorithm to minimize the root mean square error of the relevant deterministic NWP output for visibility predictions like 2 meter temperature, dewpoint temperature and wind speed. After the correction of the NWP output, the TAFG model computes probabilistic aviation relevant meteorological parameters that are not in the NWP model output, like Runway Visual Range (RVR). The RVR is the

visibility including the back luminescence of the aircrafts and runway lighting. The probabilistic forecasts of RVR and ceiling are combined to compute probabilistic forecasts of the LVP phases M, A, B and C. The output of the TAFG therefore includes deterministic corrected variables (e.g. dewpoint temperature) and probabilistic variables (e.g. LVP phase C). The steps involved in the TAFG are visible in Figure 1.2
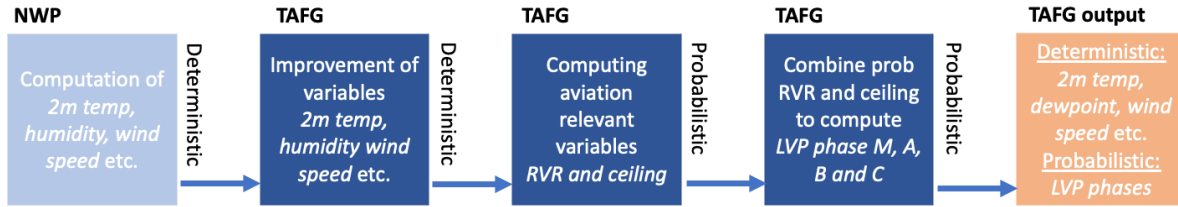
Figure 1.2: Representation of the realization of the TAFG output. The improvement and conversion of raw NWP model output in the left of the figure, with the intermediate steps, to the probalistic [0-100] TAFG output at the right.

An example of the final SKV on which the aircraft operations are based is visible in Figure 1.3. It can be seen that the SKV does not only include the LVP phases, but also remarks at the bottom of the forecast for the air traffic controller from the aviation meteorologist to include all knowledge.

Figure 1.3: Example of a SKV forecast made by the KNMI. The top five rows show the probabilistic forecasts for the LVP phases for the lead times in the columns, in which the colors indicate when there is severe low visibility prediction. From the sixth row, the SKV shows the deterministic forecasts for other relevant meteorological parameters like wind speed. The bottom box of the figure shows the remarks made by the aviation meteorologist.

In 2007, the KNMI performed research to improve the probabilities of poor visibility and ceiling (de Rover et al., 2008). The HIRLAM TAFG model performance of predicting low visibility procedures (LVP) at Schiphol Airport was investigated and improvements were made. Air Traffic Control The Netherlands (LVNL) shifted to the newer HARMONIE TAFG in November 2020, which is also being used for normal weather predictions of the KNMI and has a higher horizontal resolution resulting in a more detailed spatial variability and a better representation of convective clouds. The ECMWF, which is available with more mid- to long-range forecasts, TAFG model is also operational but not often used by the aviation meteorologists. No information on the performance of using these models as input for the TAFG is available.

Schiphol Airport has a wide set-up covering almost 3000 hectares. The furthest runway, the Polderbaan, is at a distance of 4 km from the terminal buildings. The surroundings of Schiphol Airport is agriculture, with a local highly urbanized surface due to the buildings and runways. Bergot et al. (2015) showed that these different surface characteristics can have an effect on the fog formation at an airport. The effects of the heterogeneous surfaces at Schiphol airport on the performance of the fog forecasting has not been investigated.

Izett et al. (2019) describes fog as a "secondary process" for which a correct simulation depends on a correct simulation of the temperature, humidity and wind conditions. When incorrectly simulating these variables, it can have a significant impact on the performance of predicting low visibility. For now, it is still unclear what the performance of temperature, dewpoint temperature which represents the humidity and the wind speed of the TAFG models is and how this can influence the performance of the visibility forecast.

1.1. Research question

The numerical guidance system, TAFG, computes and improves predictions of low visibility at Schiphol Airport and converts deterministic to probabilistic forecasts. The TAFG has recently shifted from input NWP model from HIRLAM to HARMONIE, with the ECMWF TAFG also available but not in operational use. The difference in the performances of these three TAFGs has not been investigated. Furthermore, different surface characteristics of Schiphol Airport can lead to local differences in fog occurrence. This can have an effect on the performance of the TAFG performance for low visibility. Also, the causes for inaccuracies in performance should be investigated using the variables temperature, dewpoint temperature and windspeed that are the basis of good low visibility simulation. This leads to the following research question:

What is the performance of the low visibility forecast of the TAFG at Schiphol Airport?

To answer the main question, four subquestions are defined and answered:

1. What are the performances of the low visibility forecasts at Schiphol Airport with HARMONIE, HIRLAM and ECMWF as an input model for the TAFG compared to the observations?
2. What effect does the spatial variability in surface characteristics between four observations locations at Schiphol Airport have on the performance of the TAFG?
3. Are the temperature, humidity and wind speed a cause for the inaccuracies in the low visibility forecasts of the TAFG?
4. What are possible solutions to improve the low visibility forecasts of the TAFG at Schiphol Airport?

Chapter 2 will summarize the relevant background theory on the fog formation process and the forecasting methodologies to help understand the analysis performed in this research. After this, the methods used in the this research are explained in chapter 3. Chapter 4 gives the results of the methods explained in the previous research to answer the sub question 1,2 and 3 subsequently. Lastly, conclusions are drawn and recommendations are given for both the practical improvement of the TAFG as well as suggestions for further research to improve visibility and ceiling predictions at Schiphol Airport.
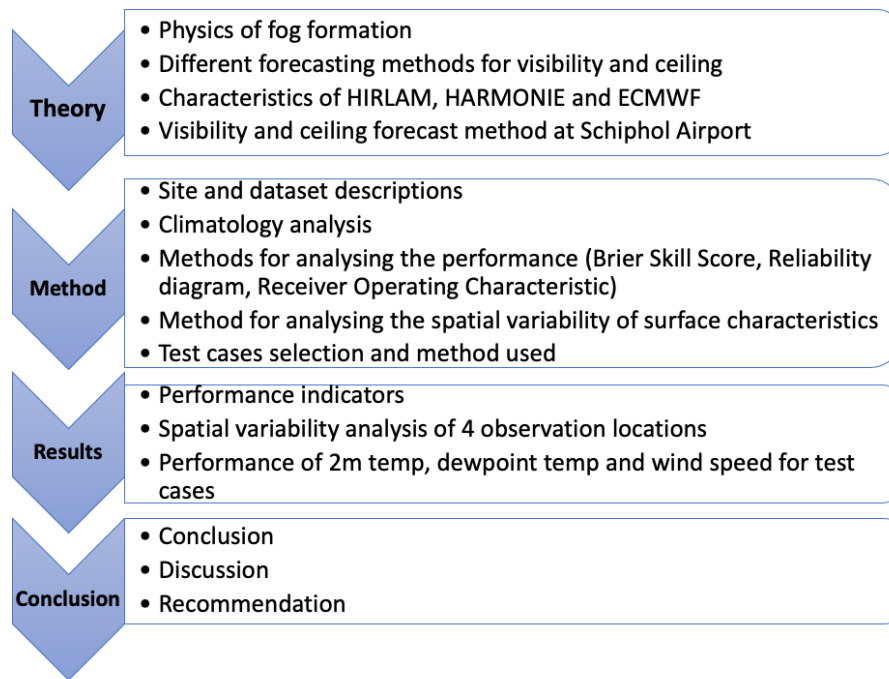
Figure 1.4: Elaboration of the methods used in this master thesis to answer the research questions and to form a conclusion. The four consecutive chapters are visible in the blue arrows and their content is specified in the boxes on the right.

2

Fog forecasting in aviation

This chapter will summarize the relevant background theory on the fog formation and forecasting process at Schiphol airport, that will be helpful to understand the analyses done in this research. This chapter will first describe the physics of fog formation, after which the different techniques used for predicting fog in literature are discussed and lastly the forecasting process at Schiphol Airport is discussed.

2.1. Physics of fog formation

This research focuses on reduced visibility caused by fog, since this is the most occurring and severe phenomena that causes reduced visibility (Izett et al., 2019). Fog consists of suspended water particles (droplets) that cause a reduction of the horizontal visibility to below 1 km (OFCM, 2017). These suspended water particles have a diameter size generally from 5 to 50 μm. It is often referred to as a surface cloud.

The air can contain an amount of water vapour, which is the gaseous phase of water. The amount of water vapour air can hold, before the air becomes saturated, is dependent on the temperature and pressure. Whether the air is saturated is often measured as the relative humidity (RH), in which a 100% relative humidity indicates saturated air. When the air is saturated, the water vapour will condensate into the liquid phase resulting in fog (Izett et al., 2019). The water holding capacity of the atmosphere is described with the Clausius-Clapeyron equation with T in Degrees Celsius:

$$e_s(T) = 6.1094 \exp \frac{17.625T}{T + 243.04} [hPa] \tag{2.1}$$
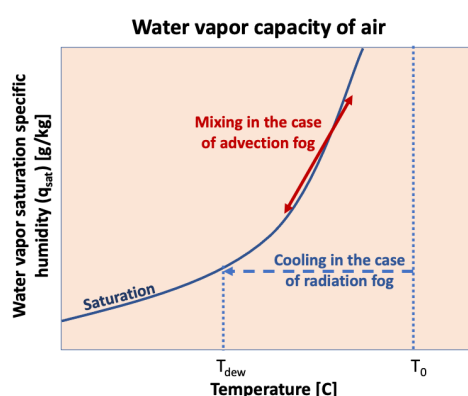
Figure 2.1: The water capacity of air. The x-axis represents the dependency of temperature and y-axis the water vapor saturation specific humidity. The blue arrow is the cooling of air to dewpoint temperature, as in the process of radiation fog. The red arrow is the mixing of air masses with different q_{sat} as in the process of advection fog.

The Clausius-Clapeyron equation defined by Alduchov and Eskridge (1997) gives the rate of increase in the vapor pressure per unit of increase in temperature. The Figure 2.1 visualises the relation of increase of temperature resulting in increase in saturation pressure of the water vapor. The saturation specific humidity is the mass of water vapor in a given mass of saturated air, which is dependent on the temperature of the system. So when this number is exceeded, condensation takes place. The figure shows the dewpoint temperature, at which the air is saturated and the relative humidity is 100%. Saturation and therefore condensation of the water vapour into fog can be reached in two ways. Firstly, the air can cool with a constant q_v, reaching the dewpoint temperature and therefore RH of 100%. Secondly, more water vapour can be added (Izett et al., 2019), for example by mixing of air masses with different humidity.

6

Fog formation is always a result of the condensation of water, but there are different processes that onset this condensation, which distinguishes the fog types. Research at Charles de Gaulle Airport in Paris, show that 78% of the low visibility procedures are caused by three types of fog: lowering stratus event, radiation fog and advection fog (Roquelaure et al., 2009). Also, research by Izett et al. (2019) shows that these are the most common in the Netherlands and therefore these are the ones considered in this research.

Radiation fog usually forms at nights with clear sky and low wind speed conditions, with occurrence of strong near-surface cooling. These are stable boundary layer conditions, at which the radiative cooling of the surface leads to saturation at dew point temperature and condensation of water vapor of the near-surface air mass (Gultepe et al., 2007). Deepening of the fog layer will occur with continuing near-surface radiative cooling and radiative cooling at the top of the fog layer. Turbulent mixing of the stable boundary layer will lead to both the deepening of a fog layer (Gultepe et al., 2007) or to dew deposition and thinning of the fog layer (Bergot and Guedalia, 1994). Wind and increase in radiation by sunlight are other causes for fog dissipation, with wind dissipating the fog and higher temperatures due to sunlight counter-acting the radiative cooling. The characteristics of radiation fog can be very variable; short-lived and persistent, not- and well-mixed, shallow and very deep fog event (Román-Cascón et al., 2016). The formation process is visible in Figure 2.2.

Advection fog forms when warm moist air advects over a cold surface and the mixing of the air masses results in condensation. This phenomenon does often occur at coastal regions and in spring conditions due to warming of land and cold sea air (Gultepe et al., 2007). The formation process is visible in Figure 2.3.

Lowering stratus event is the lowering of cloud base all the way to the surface. Various different processes have been identified to lead to cloud base lowering. The different formation processes leading to lowering stratus events have not been studied excessively, also there has not been many observational analyses so data is also lacking for lowering stratus events (Román-Cascón et al., 2019). Due to this complex phenomenon and lack of previous research, this event is not further considered.
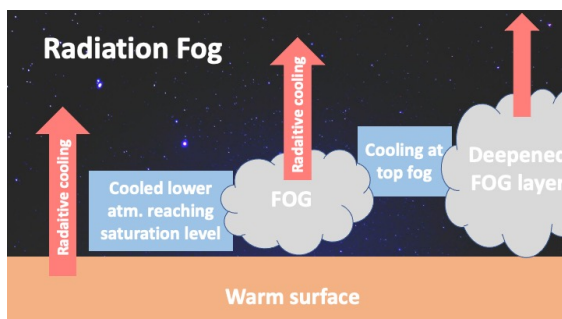
Figure 2.2: The process of radiative cooling of the surface leading to a cooled lower part of the atmosphere, at which saturation is reached and condensation will take place forming a fog layer. Deepening of the fog later is caused by further radiative cooling at the top of the fog layer.
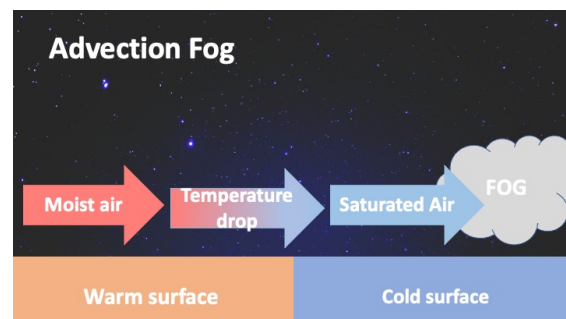
Figure 2.3: Advection of warm moist air over a cold surface, leading to a temperature drop and eventually saturated air at which condenstion can take place and a fog layer will be formed. Deepening of the fog layer is also here caused by further radiative cooling at the top of the fog layer.

Condensation of water vapour, and therefore fog formation, does not always occur when the air is saturated. Water can deposit onto a surface when the air is not 100% saturated. Aerosols in the air provide these deposition surfaces on which water droplets can occur, therefore they are called cloud condensation nuclei (CCN) (Izett et al., 2019). CCNs have a size of typically 0.2 µm on which the water vapor will condense. CCN can be marine aerosols like salt or continental aerosols like desert dust and industrial emissions. Air pollution results in more CCN leading to more condensation. Therefore, the likelihood of fog formation does not only depend on temperature and moisture content, but also on the atmospheric composition.

Fogs have a liquid water content (LWC) of 0.01-0.4 g/m^3, which is low compared to most clouds. The lower liquid water content is due to lack of strong updrafts (Gultepe and Isaac, 2007). The cloud droplet Number Concentration ranges from only a few droplets per cm^{-3} to a 100 per cm^{-3} (Gultepe et al., 2008). The cloud droplet number concentration is dependent on many other variables like aerosol size distribution, temperature and the updrafts in the air mass (McFiggans et al., 2006). The reduced visibility is caused by the forward scattering of the light by the droplets and the reduction in brightness contrast (Gultepe et al., 2007). A relation between the two microphysical characteristics of fog LWC and Number Concentration (N_d) has been determined called the Fog Index (Gultepe et al., 2008). This relation shows that when there is more available water in the air mass and the density of these water droplets is high there will be low visibility due to more forward scattering of light by the droplets and a reduction of brightness contrast.

$$Vis_{obs} = \frac{1.002}{(LWCN_d)^{0.6473}}[m] \tag{2.2}$$

Due to a high variability in the LWC [g/m^3] and N_d [cm^-3], there is also a high variability in the reduced visibility caused by the fog.

The high variability in visibility is due to the various processes shaping, developing, dispersing and decaying the microstructure of fog. The most important processes, defined by Duynkerke (1991), are radiative cooling of moist air, mixing of air masses with different heat and moisture contents, horizontal and vertical wind, horizontal advection, topographic effects and surface characteristics like vegetation leading to different heat and moisture transport in and at the surface. Duynkerke (1991) also describes the variables on which the further development of a fog layer depends on. These are the longwave radiative cooling of the air mass at the top of the fog layer leading to more condensation, the micro-physics of the fog layer and shortwave radiation which can lead to more turbulent fluxes and therefore dissipation.

In order to simulate all these processes of fog formation and development, many models use governing equations for atmospheric dynamics, which are the conservation of momentum, flux and energy and the ideal gas law. These form a basis for the simulation of formation, development and dispersion of a fog layer and are described in Appendix A. The state of the atmosphere changes due to diabatic processes, but also due to advection which includes turbulent motions. The Reynolds decomposition of the governing equations are derived to split the variables in a fluctuation part (ϕ'), which is associated with turbulent motions, and the mean ($\overline{\phi}$) part visible in Equation 2.3. With the decomposition, the equations describe the turbulent flows and this form is often used in weather models.

$$\phi = \overline{\phi} + \phi' \tag{2.3}$$

The Reynolds decomposition results in the turbulent motions (at the right hand side of the equations) and large-scale advection (from the second term at the left hand side of the equations). The large scale advection also includes the large scale subsidence in the air mass. All the decompositions of the governing equations are described in Appendix A. This chapter will include the two most relevant in this research: the heat and water budget.

Reynolds decomposition of the heat budget
For the heat budget, horizontally homogeneous turbulence conditions are assumed. An overbar is the mean and the prime is the fluctuation:

$$\frac{\partial \overline{\theta_l}}{\partial t} + \overline{u}\frac{\partial \overline{\theta_l}}{\partial x} + \overline{v}\frac{\partial \overline{\theta_l}}{\partial y} + \overline{w}\frac{\partial \overline{\theta_l}}{\partial z} = -\frac{1}{\overline{\rho}c_p}\frac{\partial \overline{F_j}}{\partial x_j} - \frac{\partial \overline{w'\theta_l'}}{\partial z} + S_{\theta_l} \tag{2.4}$$

θ_l is the liquid potential temperature defined in Appendix A, c_p is the specific heat of dry air at constant pressure, F_j is the added heat (e.g. radiation) and w' is the vertical velocity fluctuation.
The first term of the right hand side (rhs) of the equation represents the added radiation. This is both longwave as shortwave radiation including the surface heat fluxes. The second term of the rhs represent the vertical motions of a parcel. The third term of the rhs is the sink or source term which is

precipitation. The turbulent motions in the heat in the air mass are dependent on all three components.

Reynolds decomposition of the water budget
For the water budget an overbar is the mean and the prime is the fluctuation:

$$\frac{\partial \overline{q_t}}{\partial t} + \overline{u}\frac{\partial \overline{q_t}}{\partial x} + \overline{v}\frac{\partial \overline{q_t}}{\partial y} + \overline{w}\frac{\partial \overline{q_t}}{\partial z} = -\frac{\partial \overline{w'q_t'}}{\partial z} + S_{q_t} \tag{2.5}$$

w' is the vertical velocity fluctuation and q_t the total water specific humidity. The turbulent motions in the rhs of Equation 2.5 represents the vertical motions and the sink and source term. This sink source term represents the precipitation.

The accurate simulation of temperature, humidity and wind speed is needed to simulate the formation, dispersion and development of fog. Turbulent motions influence the water and heat budget and therefore the temperature and humidity as seen in Equation 2.4 and Equation 2.5 and should therefore be well represented in a model Román-Cascón et al. (2019). These turbulent fluxes can be the result of different small-scale processes like vertical displacement of the air visible in the rhs of the Reynolds decomposition of the heat and water budget. Too much turbulence will result in no fog formation or dispersion of the fog and no turbulence results in fog formation or deepening of the cloud layer, when there is enough water vapor available (Steeneveld and de Bode, 2018). Even though turbulence is an important factor considering airport meteorology, NWP models are too coarse for resolving the turbulence (Gultepe et al., 2019) and therefore parameterizations are used.

The land-surface coupling is also an important process. Formation of fog is highly dependent on the surface fluxes of heat and moisture, which are represented in the right hand side of Equation 2.4 and Equation 2.5. These surface fluxes are dependent on the soil properties, like vegetation. High uncertainties in calculating surface fluxes still remain (Gultepe et al., 2007). The thermal conductivity is dependent on the soil type and more specific moisture content. A dry soil leads to faster cooling of the ground, which results in earlier on-set of fog (Bergot and Guedalia, 1994). As mentioned before radiative cooling of the ground can lead to fog formation. Dew deposition, occurring when soil temperatures are below dewpoint temperature, is often a key ingredient for fog formation which follows after dew has formed (Hodges and Pu, 2016). Soil temperatures are also dependent on soil properties. Therefore, heterogeneity of the soil is also key in good representation of fog processes.
A significant variability in the microphysics of fog has also been observed. For example, the droplet size distribution has a high spatial variability in a fog layer. This is due to the high sensitivity of the different processes shaping the microstructure of fog like turbulence and radiation (Gultepe et al., 2007).

The reduction of visibility by fog is due to the absorption of light by the droplets and the scattering of the light along the optical path. Therefore, visibility and fog can be measured by using the light transmission properties (Izett et al., 2019). There are many techniques for field observations of fog formation, development and dissipation. Climatological data (Tardif and Rasmussen, 2007), in-situ observations (Gultepe and Isaac, 2007) and remote sensing data are techniques used. There is also a lot of uncertainty in the measurements of fog. Due to the small particle sizes and sometimes mixed-phase structure of fog, difficulties occur when obtaining measurements that are reliable (Gultepe et al., 2007). Also difficulties arise due to the three-dimensional characteristics of a fog layer, that is hard to accurately map without a large number of sensors (Izett et al., 2019). More and extensive measurements are needed for different conditions of fog to better understand and parameterize the microphysics of fog (Bergot et al., 2005). More so, local measurements are needed to determine adequate initial model conditions and verification of model outputs (Rémy and Bergot, 2009). Uncertain observations, used as input for the NWP predictions can lead to significant forecast issues (Gultepe et al., 2019).

The physics of fog are still challenging due to the difficulties in obtaining reliable measurements of fog and the unknowns in the processes shaping the micro-structure of fog and therefore difficult to accurately parameterize in weather models.

2.2. Fog forecasting models

Different fog forecasting models have been developed and studied. The main difference between models is the number of spatial dimension of the models that correspond to the focus on different processes. The two types that are used are the one-dimensional model which corresponds to a specific point in space and the three dimensional model such as NWP models which represents a larger area. Radiative cooling is often well represented in one-dimensional models due to a higher vertical resolution compared to three-dimensional models. When advective processes contribute to the formation of fog, three-dimensional models are often suggested for the forecasting (Müller et al., 2010). Especially for heterogeneous and complex terrains, the use of three-dimensional models is advised. Often these models can also be linked to each other, for example the one-dimensional models often use the three-dimensional model output as a model input. The input models for the TAFG are three-dimensional models and therefore these models will be discussed.

NWP models are three-dimensional forecasting models. Although these models solve horizontal processes, there are still many difficulties in the accurate forecasts of fog. This leads to both under-(Westerhuis et al., 2020) and overestimations (Román-Cascón et al., 2019) of the fog in forecasts. The NWP models have both a vertical and horizontal resolution that are not small enough to capture the small processes that are most important in the formation, development and dissipation of fog (Román-Cascón et al., 2019). This results in many parameterizations schemes within all NWP models. Main parameterizations in three-dimensional, as well as for the one-dimensional, models are the turbulent transport of heat and moisture and the microphysics (Gultepe et al., 2007). Different parameterization schemes result in sometimes contradicting results of the NWP models (Szintai et al., 2015; Román-Cascón et al., 2016; Westerhuis et al., 2020; Steeneveld and de Bode, 2018).

To decrease the number of parameterization schemes, solutions are sought in increasing the vertical and horizontal resolution. The fine grid resolution will lead to better resolving skills of the model for the physical processes near the ground which lead to better predictions of fog formation (Gultepe et al., 2007). Increasing the vertical resolution was effective in better capturing the onset of fog formation (Gultepe et al., 2007), but not a better representation of the dissipation phase of fog forecasting (Westerhuis et al., 2020). Increasing the horizontal resolution will capture the heterogeneities of the soil and her vegetation leading to better predictions of fog formation (Gultepe et al., 2007). However it has no impact on the dissipation that is modelled (Westerhuis et al., 2020).
Reducing the parameterization schemes does not have the desired impact on the results of the fog forecasting to be accurate enough to use for aviation operations. Also, the increase of resolutions is very costly and not feasible in most forecasting systems (Boutle et al., 2016).

2.2.1. NWP models used as TAFG input

The TAFG at Schiphol Airport uses three NWP models as input; HIRLAM, HARMONIE and ECMWF.
HIRLAM (High Resolution Limited Area Model) is a meso-scale weather forecast model. It was previously used in European countries for operational use, but development has stopped since 2006 and therefore operational use has decreased. It is a hydrostatic model, initialized every 6 hours.
The HIRLAM cooperation shifted to a newer model, which was a collaboration between different consortia. HARMONIE-AROME, or HARMONIE, is used for short-range meso-scale weather predictions in Europe. It is a cooperation of in total 26 countries. The wide use of this model resulted in detailed updates to the model's physical parameterizations. HARMONIE has a small horizontal grid spacing and is a non-hydrostatic weather model. Due to this, the vertical convective motions can be resolved.
ECMWF stands for European Centre for Medium-Range Weather Forecasts. As the name suggests, the ECMWF NWP aims to compute medium-range global weather forecasts from 15 days to even more seasonal predictions of 12 months ahead. ECMWF is a cooperation of 23 countries. It is a hydrostatic weather model with a high number of vertical levels and long lead times.
HIRLAM and HARMONIE are also embedded in ECMWF. More model specifics of HIRLAM, HARMONIE and ECMWF are visible in Table 2.1.

Model Specifics			
	HARMONIE	**HIRLAM**	**ECMWF**
Vertical levels [nr]	65	60	137
Vertical levels in lowest 3000m [nr]	33	27	32
Initialization frequency [hours]	3	6	6
Leadtime maximum [hours]	48	48	72
Grid spacing [km]	2.5	11	9
Time step [min]	1	4	5
Other	Non-hydrostatic	Hydrostatic	Hydrostatic

Table 2.1: Technical specification parameters from the HIRLAM, HARMONIE and ECMWF models as the NWP models which are used as input for the TAFG at Schiphol Airport.

The NWP models as input for the TAFG also have difficulties with accurate fog forecasting. HIRLAM can't represent fog under the lowest model layer. This leads to early dissipation of fog and limits the deepening and growth of a fog layer. However, HIRLAM is able to model the onset of fog for radiation fog in the Netherlands (Van Der Velde et al., 2010).
In the Netherlands there is an overestimation of dense fog by HARMONIE (de Rooy, 2014). This was mostly due to an overestimation of the radiative cooling.
ECMWF is able to capture the fog onset, but the development of the simulated fog layer is not well simulated. This is due to an overestimation of the 2 meter temperature and the dewpoint temperature (Belo-Pereira and Santos, 2016). This study was performed in Lisbon, Portugal and no recent study of the performance of fog predictions in The Netherlands are found.
All three NWP models used as input for the TAFG have difficulties in forecasting the development of a fog layer. Possible causes could be the inaccurate prediction of temperature and dewpoint temperature.

Although all the research addresses different causes for inaccurate fog predictions for NWP models, consensus has been reached that more research in high-resolution NWP models to better represent fog is needed (Westerhuis et al., 2020; Tardif and Rasmussen, 2007; Román-Cascón et al., 2016). The current NWP models still have difficulties in accurate fog forecasting and have contradicting results. Therefore further processing of these forecast model outputs is needed for the use in airport operations.

2.3. Aviation Forecasts

In order to get fog forecasts and spatial and temporal scales that are needed for aviation meteorology, several solutions to improve NWP output are available. Firstly, the integration of super-sites (meteorological observatory) in NWP models is used. Short-term forecasts are based on both NWP model outputs as well as in-situ and remote sensing observations measured at these super-sites (Gultepe et al., 2019). This is also used at Schiphol Airport.

Another solution to improve NWP output is the post-processing of NWP model output. Statistical post-processing is often used to combine observations and the deterministic output of the NWP models into a probabilistic forecast (Jacobs and Maat, 2005). These improved weather forecasts used for airport operations are produced in a specific format: the Terminal Aerodrome Forecasts (TAF). This format contains detailed forecasts of meteorological conditions such as wind, visibility, clouds, ceiling height and precipitation for the surrounding of an airport (Jacobs and Maat, 2005). This TAF includes 6 short-term forecasts every hour, followed by 10 long-term forecasts every three hours with a lead time of 8 to 36 hours. Meteorological aviation routine weather reports (METARs) are the type of observations used in aviation meteorology. They are obtained every 30 minutes, and contain information among others on the temperature, visibility, pressure and wind speed. The TAF and METARs are in the same universal format, which makes it easy applicable for air traffic controllers worldwide.
TAF output contains probabilistic forecasts of the occurrence of visibility conditions. This is in the form of a probability of occurrence indicator for the likelihood of a change in weather. Only likelihoods of 30% and 40% are used (Gultepe et al., 2019). All values below the threshold of 30% are insignificant in the TAF and will not have a change indicator. All values greater than 40% always have a change indicator, which indicates that an event will happen.

The visibility in the TAF output is a combined probabilistic variable of the Meteorological Optical Range and Ceiling. The Meteorological Optical Range (MOR) is often used as a model output for visibility. It is often defined as the path length in the atmosphere, in which light from a source is reduced to 0.05% of the original intensity of the source (200, 2008). The ceiling represents the cloud base height. When fog is predicted in NWP models, this is translated to a high probability of low visibility occurrence in the TAF output.

Also, the use of artificial intelligence as a post-processing method can improve NWP output results. Artificial neural networks (ANN) are defined as "*Computational methodologies capable of establishing associations between the independent variables (the predictors) and the dependent variable (predictand), through the experimentation of a multitude of situations (learning data set). Information on the relationship between the predictors and the predictand is placed in a net of interacting nodes.*" (Gultepe et al., 2007). This approach has a better forecast quality compared to linear and logarithmic regression methods in forecasting (Marzban et al., 2007), however costs for using ANN are often high. Another post-processing method is the use of Random forests. The Random forest exists of a large number of single decision trees that operate like an ensemble. Every individual tree within the random forest generates a prediction and the prediction that occurs most often in all the individual trees will be chosen as the result of the model's prediction. This post-processing method is currently in development within the KNMI to improve the fog forecasts and better capture the uncertainties.

2.3.1. Schiphol Forecast System

The general TAF did not perform as needed for visibility predictions at Schiphol Airport (de Rover et al., 2008). Therefore, The Dutch Meteorological Institute (KNMI) introduced the Schiphol Kans Verwachting (SKV) in 2003, which is Schiphol's probability weather forecast. The SKV is a both deterministic as well as probabilistic forecast for the upcoming 1 to 31 hours with an update frequency of 3 hours. A low visibility procedure, which can be caused by fog, is quantified in the SKV as probability of occurrence for a combination of the levels of visibility and ceiling. The process of computing the SKV at Schiphol airport is visible in Figure 2.4.

The SKV has an improved statistical processing module that post-processes the NWP data compared to the normal TAF: the TAF Guidance (TAFG). This TAFG is produced by Meteo Service commissioned by the KNMI. The SKV system also includes an aviation meteorologist, who combines all the TAFG outputs to finite values for the SKV. The main TAFG output currently used is the HARMONIE TAFG output, but HIRLAM TAFG and ECMWF TAFG are also available for the aviation meteorologist.

The TAFG output includes deterministic weather parameters like air temperature and surface winds, but also probabilistic output for clouds and visibility: the list used for the SKV is visible in Figure 1.3. The probabilistic value of threshold exceedance for the combination of visibility and the height of cloud base is given. These thresholds can be seen in Table 1.1 and determine the LVPs. The LVP phase is a representation of the severity of visibility reduction, based on the visibility and cloud base height.

The output of the TAFG has a variable lead time for the NWP models HARMONIE, HIRLAM and ECMWF shown in Figure 2.4. For this post-processing method, local and upstream observations are used with high-resolution information on topography (Gultepe et al., 2019). Short-term predictions of visibility are highly dependent on local and upstream observational data, while long-term predictions are more dependent on NWP outputs.

The TAFG uses a different definition for visibility compared to the TAF. The Meteorological Optical Range (MOR) is often used as a measurement for visibility and also for the TAF. The Runway Visual Range (RVR) is used for visibility conditions in the TAFG. The Meteorological Service for International Air Navigation uses the definition "*The range over which the pilot of an aircraft on the centre line of a runway can see the runway surface markings or the lights delineating the runway or identifying its centre line*" (200, 2005). According to this definition, the MOR values are computed and translated to RVR values using background luminescence and the intensity of the runway light. Extra information on runway specific light occurrence is needed for accurate computation of the RVR (Blajev and Krastev, 2013). The RVR is only calculated for MOR values below 1500 meters.
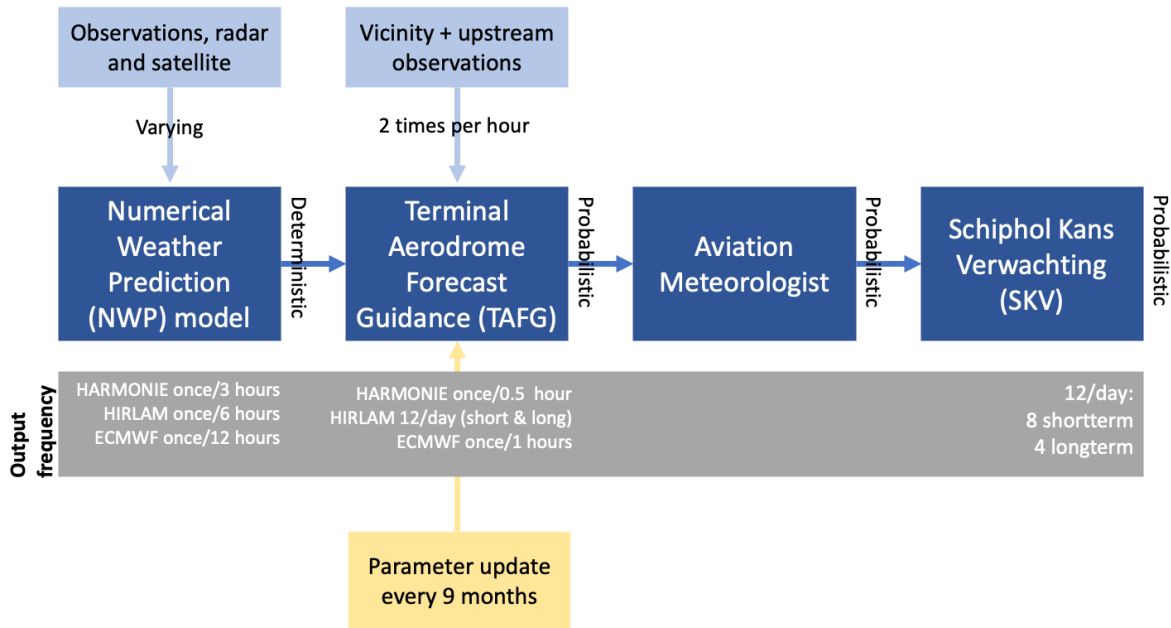
Figure 2.4: Realization of the Schiphol Kans Verwachting (SKV). The blue boxes show the general process of the NWP deterministic to the probabilistic output of the SKV on the right. The light blue boxes show the observations that are used. The grey box contains the frequency of the output corresponding to the blue boxes.

TAFG specifications

The TAFG is the post-processing model used to compute the probabilities for the LVP phases. The main processing steps are visible in Figure 1.2. The main calculation steps of the TAFG are:

1. Quality check of observations and model output

2. Improvement of model output with observations by using linear transformations

3. Deriving additional aviation relevant variables, such as the RVR and ceiling

4. Combine the probabilities of the RVR and ceiling to probabilistic forecasts of the LVP phases

The first step is a quality check of observations. After this the improvement of model output with observations using linear transformation is executed, which is the second step. Not only local data, but also other neighbouring observation stations are used by introducing observation advection. Observation advection results in the combination of both local, as well as upstream observations for the improvement of model output. For this the starting point of the trajectory is determined, after which representative up-stream observation stations are determined and their weights are calculated in order to have a predictor value. Model output is then transformed to station coordinates. This is done by interpolating the results at grid points to station coordinates.

Another process to improve the model output is the smoothing of grid points. Improved non-hydrostatic NWP models are able to produce convective cells. It is possible to predict how many and how big the cells are, but it is impossible to predict exactly where they are, due to the nature of this type of weather. Therefore multiple grid points are averaged. An algorithm has been developed in the TAFG to average these grid points, reducing resolution by roughly 8 km every step. This "smooth" procedure is now set at 8 averaging steps.

For fog, special predictors are computed in the third calculation step. Special predictors are computed to use as fog indicators, like RVR and ceiling. These predictors are non-linear transformations of wind, spread and cloud cover. These predictors are obtained by using a stepwise forward multiple regression algorithm and by minimizing the root-mean-squared error. For this, local orographic predictors are used in order to improve the consideration of non-linear and complex local effects in the regression equations. The order in which the predictands are computed is the order in which regression analysis

is performed. The most important consideration for this order is that forecasts of certain predictands serve as predictors for consecutive predictands. For example, the MOR is first calculated and then transformed to the RVR. For the visibility an extra transformation is done. Without transformation, the difference between visibility of 19 and 21 km would be as important for the statistics as the difference between 0 and 2 km. A logarithmic transformation is chosen which makes the visibility predictand much more sensitive to low visibility and insensitive for visibilities of more than 10 km.

The RVR and ceiling are probabilistic variables. The transformation of a continuous variable into a probabilistic variable is usually done by a step function which is equal to zero below the threshold value and equal one above the threshold value. This step function is smoothed to avoid values below zero.

The last step is the combining of the RVR and ceiling into LVP phases. These are based on the thresholds defined in Table 1.1.

Parameter update

Every nine months, the regression equations of the TAFG are updated. This is done using all seasonal data computed by the previous TAFG. The HIRLAM TAFG has data spanning 20 years time. HARMONIE TAFG is developed using 3 years of data from 2015 until 2018. The HARMONIE TAFG is operational since November 2020 and collecting more data since this starting point. More rare events are included in the seasonal data of the HIRLAM TAFG compared to that of the HARMONIE TAFG. Therefore, it is expected to perform better than the HARMONIE TAFG.

This chapter summarized the relevant background theory on the fog formation and forecasting process at Schiphol airport, to help understand the analyses done in this research. The difficulties in the physics of fog remain due to the unknowns in the processes shaping the micro-structure of fog and the difficulties in obtaining reliable measurements.

At Schiphol, a post-processing model TAFG is used to compute the probabilities for the LVP phases to correct for the inaccuracies in fog predictions by NWP models. Research showed that in order to have a good fog simulation, the correct simulation of temperature, humidity and wind speed is needed.

3

Methods

The research question aims to evaluate the performance of the TAFG in forecasting fog at Schiphol airport with the three NWP models as input. This chapter will first give an overview of the site that is researched and the data that is used. Afterwards, the methods used in order to answer the research question and the sub questions are described.

3.1. Site description

This research focuses on Schiphol Airport in the Netherlands. Schiphol airport is an airport with 6 runways, of which 3 North-south orientated 2 SouthWest-NorthEast orientated and 1 East-West orientated runway. Every runway has two measuring points for weather: both at the end as well as at the beginning of the runway. These locations can been seen in Figure 3.2 where the numbers represent observation location and name. The main weather station is at the cross-section of the Buitenveldertbaan, Oostbaan and the Aalsmeerbaan. Schiphol Airport's capacity planning depends on a number of factors. The runways that cause the least noise pollution have the preference, but which runways can be used also depends on the weather conditions, especially the wind direction and strength.

The Netherlands has a small area with flat surfaces, but has a highly variably use of land and a dense population. This, combined with the influences of the North Sea and mesoscale influences of urbanization, results in a strong variability of regional weather conditions (Izett et al., 2019). The locations of the water bodies and cities surrounding Schiphol Airport are visible in Figure 3.1. Izett et al. (2019) confirmed that this variability is mostly apparent under clear sky and low wind conditions, in which fog can form. This is due to the low amount of external factors, such as shortwave radiation from clouds, during stable conditions.

The North Sea and Ijsselmeer are water-bodies West and East of Schiphol that can be of influence on the fog formation at Schiphol Airport. This can be through sea breeze circulation, in which the concentration of aerosols can be increased by the salt of the sea. Another influence of the water-bodies can be through their high heat capacity, which can influence the diurnal and seasonal cycle of the temperature at Schiphol Airport (Arrillaga et al., 2018). Lastly, the Gulf Stream circulation leading to warmer coastal waters by functioning as a thermal memory for the water from Equatorial regions by transporting heat leading to a higher sea surface temperature (Izett et al., 2019).

A large city is situated at the North of Schiphol Airport, which is Amsterdam. The urban heat effect results in higher temperatures at night, under weak wind conditions. These high temperatures in the city have a negative impact, on the fog formation process in the urban area. However this leads to a circulation and higher frequency of fog occurrence at the border of the urban area, due mixing of this moist hot air with the colder rural air (Steeneveld et al., 2011). Also, anthropogenic emissions of aerosols will result in higher aerosol concentration of the urban air, which are favourable conditions for fog formation (Izett et al., 2019).
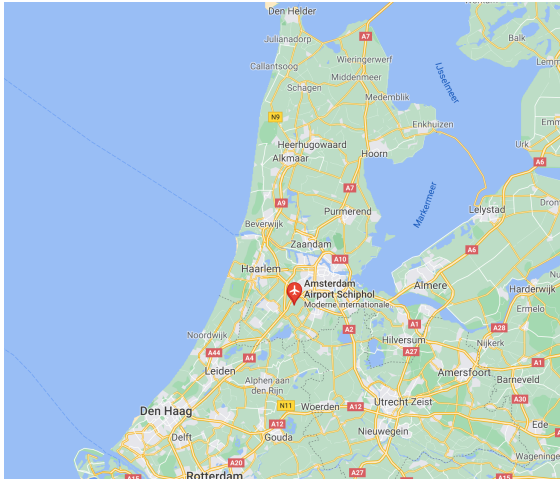
Figure 3.1: Schiphol location in the Netherlands. The surrounding cities and water bodies are visible.

Figure 3.2: Schiphol runways indicated by the blue lines. The observation locations at the beginning and end of the runways are indicated with the numbers (e.g. 18R).

Schiphol has a low fog occurrence compared to other locations in the Netherlands, which can be related to the characteristics of Schiphol and airports in general (Izett et al., 2019) and can be seen in Figure 3.3. Schiphol is situated with mostly agricultural in it's surroundings. However, airports have locally an urbanized characteristics due to the buildings and runways made of concrete. This local urbanization has a high impact on the fog formation process (Bergot et al., 2015). Also, unique for Schiphol is the large area of the Airport with almost 3000 hectares. This results in a runway at a 4 km distance of the main building, situated in a polder hence the name "Polderbaan". Research shows a high variability between the different runways of fog occurrence, with high fog occurrence at the Polderbaan (Izett et al., 2019).

Another factor influencing the fog formation at Schiphol Airport is the 500.000 aircraft movements per year, with Schiphol being a busy Airport in Europe. The influence of these aircrafts is both favorable for the dissipation as well as for the formation of fog. Aircraft motions will result in higher wind-speeds due to landing or taking off (Izett et al., 2019). This will result in higher dissipation rates of fog when aircrafts are moving. The engines of these aircraft also results in a temporary peak in the temperature. This will result in more radiative cooling needed to reach the dewpoint temperature of the lower part of the atmosphere and therefore has a negative impact on the formation of fog, since saturation will be reached later. On the other hand, aircraft motions will results in a very high concentration of aerosols in the area of the runways. This is favorable for fog formation (Izett et al., 2019).

Schiphol has a low fog occurrence compared to other locations in the Netherlands, which is favorable for airport operations. However, within Schiphol there is high variability between different runway locations. Also, the impact of fog is still of great significance due to high costs when not predicted accurately, making it a very interesting location to research.
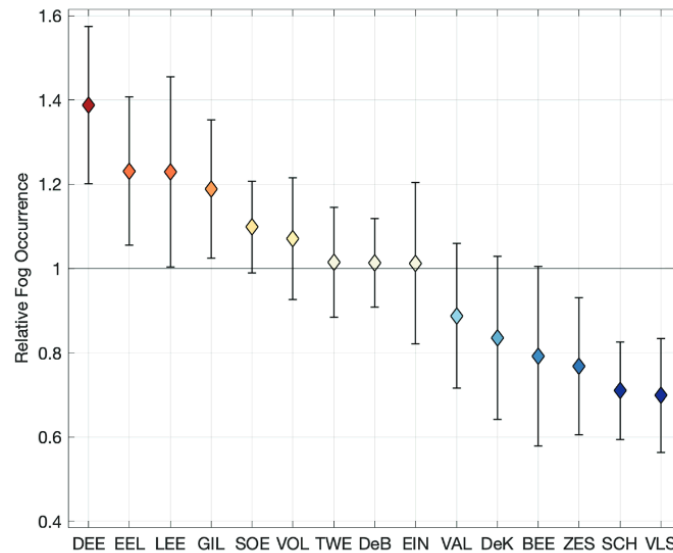
Figure 3.3: Relative occurence of fog in the Netherlands (Izett et al., 2019). The x-axis represent the observation locations for fog in the Netherlands, in which SCH is Schiphol airport. The y-axis represent the relative fog occurrence of a specific observation location, compared to the mean fog occurrence in the Netherlands. Smaller than 1 indicated a lower fog occurrence than the mean fog occurrence in the Netherlands.

3.2. Materials

Both qualitative input as well as quantitative input is used in this research. Interviews with KNMI specialists as well as literature provided qualitative knowledge considering the model specifics of the TAFG at Schiphol Airport. This will give more insights in meteorological processes relevant in obtaining the TAFG. Interviews with MeteoService (builder of the TAFG) will give insights in mathematical approaches used in the TAFG.

Three data-set types are used to obtain the quantitative results.

- **3 month TAFG output with HARMONIE as input** - The TAFG output from November 2020 until March 2021 with HARMONIE as input NWP model is used. HARMONIE TAFG is operational at Schiphol Airport since November 2020, which explains the starting date of the datasets. Four lead times are used: the 2, 4, 8 and 16 hour forecast. The output frequency of the HARMONIE TAFG is every half hour. The parameters that are used are the LVP phase probabilities, 2 meter temperature, dewpoint temperature and windspeed.

- **3 month TAFG output with HIRLAM and ECMWF as input** - The TAFG output from November 2020 until March 2021 with HIRLAM and ECMWF as input NWP is used. The parameters that are used are the LVP phase probabilities, 2 meter temperature, dewpoint temperature and windspeed. The same lead times are chosen as the HARMONIE TAFG to make a valid comparison. The output frequency of the HIRLAM TAFG is 12 times a day with 8 shortterm and 4 longterm predictions. The output frequency of the ECMWF TAFG is every hour.

- **METAR observations** - METAR observations of RVR and ceiling are used to validate the output of the LVP phases probabilities of the TAFGs. RVR values are only produced for visibility values lower than 1500 meter. This makes the dataset more limited, but still relevant for LVP phases. The RVR's are produced at four locations visible in Figure 3.12, which will be used in the spatial variability analysis. The parameters 2 meter temperature, dewpoint temperature and wind speed are used for the interpretation of the validation.

Both the qualitative and quantitative data will be relevant for answering the main research question.

3.3. Low Visibility Procedures

In this report the LVP phases are analysed. These are determined by a combination of predicted visibility (RVR) and the height of cloud base. This height of cloud base is based on a cloudiness of at least 5/8 within the eighth parts (octas) of the sky covered with clouds, this means a broken or overcast cloud cover. There are six LVP classes visible in Table 1.1 and Table 3.1: Good (G), LVP M, LVP A, LVP B, LVP C, LVP D which imply capacity restrictions from the marginal LVP and flow restriction from LVP B phase. Capacity restriction are a restrained number of air traffic operations with less arrivals and departures. Flow restrictions are the use of runways at Schiphol Airport. When flow restrictions are in force, only the parallel runways can be used. This often leads to less airplanes arriving and departing to guarantee safety in the Instrument Landing System (de Rover et al., 2008).

The probability of occurrence of LVP phases in the TAFG output used for the SKV and in this research is only upper bound. This means that when a severe LVP phase (LVP C) occurs, a LVP phase A will also occur. An adjusted definition of the upper-bound LVP phases compared to the phases defined in Table 1.1, is given in Table 3.1.

LVP phase	RVR (visibility)		Cloud base (ceiling)
< Good	RVR $\geqslant$ 5km	and	$\geqslant$ 1000ft
< Phase M	RVR $\leqslant$ 5km	or	CLB $\leqslant$ 1000ft
< Phase A	RVR $\leqslant$ 1500m	or	CLB $\leqslant$ 300ft
< Phase B	RVR $\leqslant$ 550m	or	CLB $\leqslant$ 200ft
< Phase C	RVR $\leqslant$ 350m		
< Phase D	RVR $\leqslant$ 200m		

Table 3.1: LVP phase definitions used in datasets, based on either the reduced visibility (RVR) or cloud base (ceiling).

There are three possible outcomes of the LVPs:
- Hit - Measured LVP in accordance with the TAFG forecast leading to justified measures
- Miss - Measured visibility lower than the TAFG forecast. Last minute measures like cancellations leading to high costs
- False Alarm - Measured visibility higher than TAFG forecast. Unnecessary measures were taken leading to high costs

The data of the operational TAFG for the SKV of the three models since November 2020 is first assessed. The following figures show the output of these TAFG models with different input models. The LVP phase output of HARMONIE TAFG with a lead time of 4 hours, shown in Figure 3.4, shows that the probabilities for phase A are always the highest. This is due to the upperbound LVP phases, as defined in Table 3.1. A more detailed example of the difference in forecasted LVP phases of HARMONIE with a lead time of 4 hours is visible in Figure 3.5.
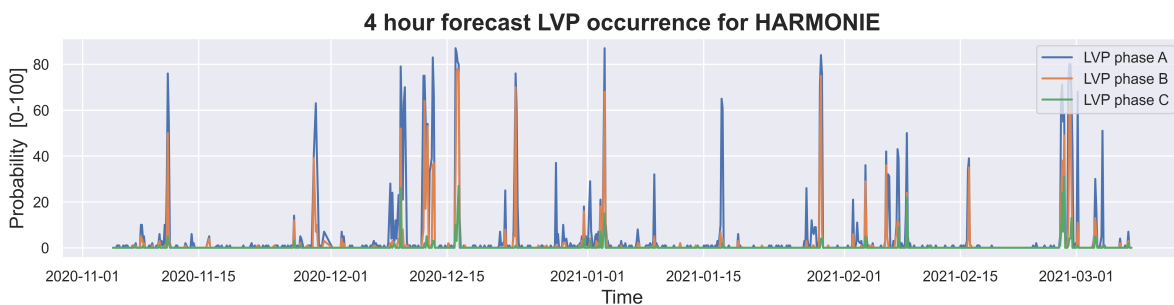
Figure 3.4: Probabilities for 4 hour forecast HARMONIE TAFG for LVP phases which are indicated by the colored lines. The y-axis represent the probability of the TAFG between 0 and 100.
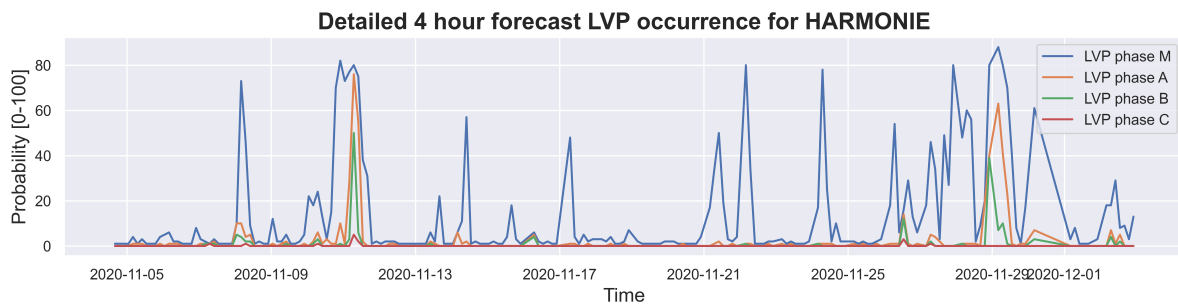
Figure 3.5: Detailed probabilities for 4 hour forecast HARMONIE TAFG for LVP phases which are indicated by the colored lines. The y-axis represent the probability of the TAFG between 0 and 100.

Figure 3.6 shows a time series of the LVP phase B forecast with a lead time of 4 hours for the three TAFG models. Better visible in Figure 3.7 is the differences in predicted probabilities for LVP phase B over time. This research will focus on quantifying these differences.
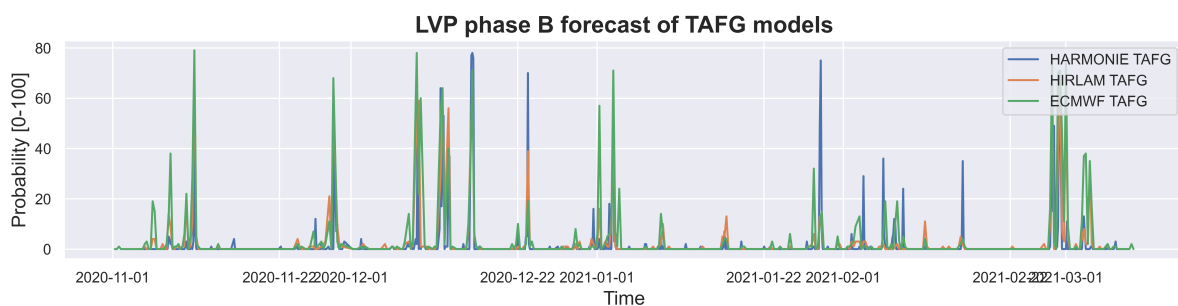
Figure 3.6: Probabilities for 4 hour forecast LVP phase B for the TAFG models which are indicated by the colored lines. The y-axis represent the probability of the TAFG between 0 and 100.
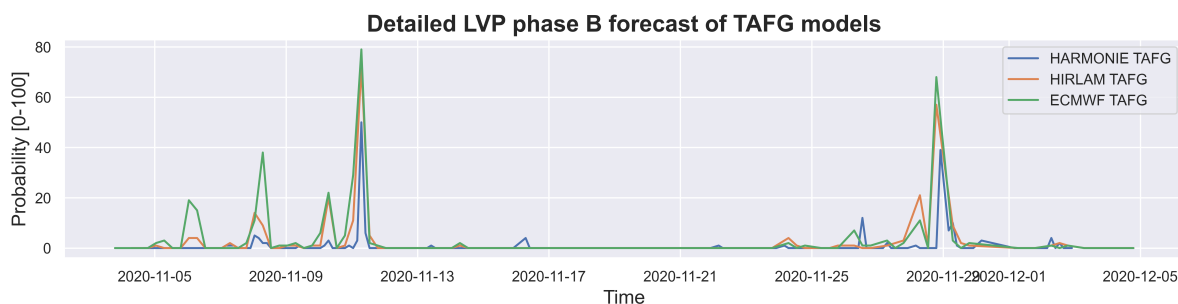
Figure 3.7: Zoom in of probabilities for 4 hour forecast LVP phase B for the TAFG models which are indicated by the colored lines. The y-axis represent the probability of the TAFG between 0 and 100.

3.4. Method
The methods used to answer the research question and the sub question are described in this section. All analysis are performed using Python and existing packages in this environment.

3.4.1. Climatology Study
The climatology gives an insight in the occurrence of weather events, in this case the visibility (MOR and RVR) and ceiling, observed over a longer period. The climatology analysis is done in order to see if the fog occurrence at Schiphol is as expected in cold, stable nights, according to the theory described in Chapter 2.

Long term trends of fog occurrence at Schiphol Airport are analysed using literature. This is done for the separate values of MOR and ceiling. The RVR at Schiphol airport is not used, since this is only available for the past 15 years. Due to the lack of RVR, estimations of the LVP phases are also not available. The data-frame used is from 1950 to 2010. The results computed by Vautard et al. (2009) are visible in Figure 3.8 and Figure 3.9.

It is clearly visible that there is a decrease in the MOR values. It is expected that this is the same for RVR, since the RVR mostly differentiates from the MOR at twilight and in evenings due to better visibility with lights at the airport. RVR is only calculated with MOR values smaller than 1500 meter. The occurrence of low ceiling has an overall decrease, but trends are much smaller compared to the MOR. There is also an increase after 2002 of the ceiling. This phenomenon can be explained by the introduction of automated cloud height measurements.

The same trends are visible in other research. Boers et al. (2015) reconstructed a decrease in fog occurrence since the 1950's. The main cause for this was the decrease in aerosol concentration and and changes in aerosol composition. This was especially seen after 1985, after which the visibility improved and the aerosol optical thickness did decrease.
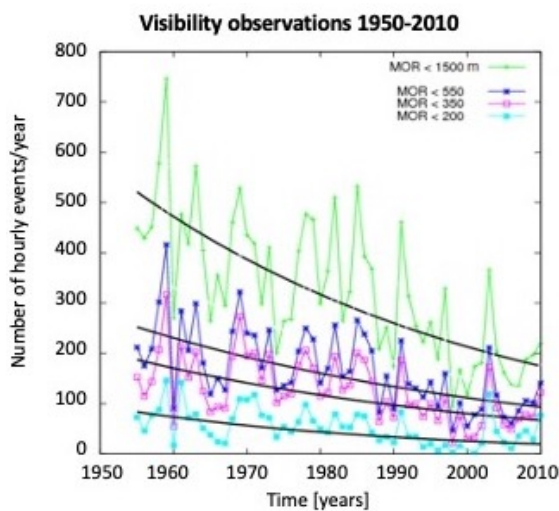
Figure 3.8: Decrease of MOR observations at Schiphol Airport from 1950 until 2010 (Vautard et al., 2009)
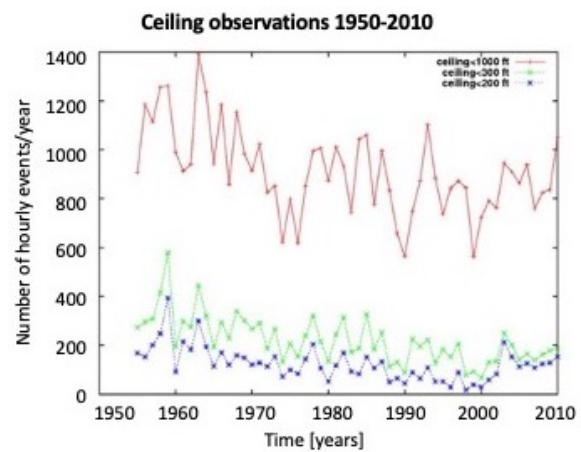
Figure 3.9: Small decrease of low ceiling observations at Schiphol Airport from 1950 until 2010 (Vautard et al., 2009)

A further climatology analysis is done, using the observations of RVR and ceiling from the KNMI at Schiphol Airport in three ways:

- The past ten years (2010-2020) of LVP occurrence are analyzed using METAR observations at Schiphol Airport. The RVR and ceiling values are obtained from the METAR data and LVP phases are manually computed. Instead of counting the number of hourly events, the number of days with at least 1 hour of LVP is computed. This technique is often used at the KNMI due to the short amount of time during a day fog is present (Tardif and Rasmussen, 2007).
- The distribution of fog occurrence throughout the day is computed using the LVP phases obtained from the METAR data from 2010 until 2020. This gives more insights in the timing of fog events in a day.
- Lastly, the seasonal trends of fog occurrence are analyzed using the LVP phases from 2010-2020. The distribution of fog occurrence per month will be computed.

This further climatology analysis will check if the occurrence trends are in line, with the ones used in literature. It will therefore function as a sanity check and can be found in section 4.1.

3.4.2. Comparing performance of probabilistic forecasts

This section shows the method used to answer the first sub question; *What are the performances of the low visibility forecasts at Schiphol Airport with HARMONIE, HIRLAM and ECMWF as an input model*

for the TAFG compared to the observations?
The TAFG output is a probabilistic forecast for LVP phases. To analyze the performance of probabilistic forecasts, special statistical methods are needed. The methods used to compare performances of probabilistic forecasts are Brier Skill Score, Reliability Diagram and Receiver Operating Characteristic. Those methods will be elaborated upon in the upcoming paragraphs. It is important that the predictand is seen as a dichotomous forecast, in which fog is occurring or not.

Brier Skill Score
The most commonly used scalar accuracy measure that is used while verifying a probabilistic forecast is the Brier Skill Score (BSS). Skill scores are computed to compare performances. The Brier Skill Score gives the performance of the probabilistic forecast compared to the observation in percentages. A good model will result in Brier Skill Scores of 100%. The Brier Skill score is therefore a relative skill of the probabilistic forecasting of a model compared to that of climatology. This is in terms of forecasting if an event occurred. For this performance indicator, the observed LVP phases obtained from the METAR observations of RVR and ceiling at the centre of Schiphol Airport are compared with the TAFG forecast for the LVP phases for all three TAFG models at all four lead times.
To obtain the BSS, first, the Brier Score is computed. This is the mean squared error of the probability forecast, shown with Equation 3.1.

$$BS = \frac{1}{N} \sum_{k=1}^{N} (y_k - o_k)^2 [0 - 1]$$

(3.1)

The observations are binary with observation $o_k = 0$ when there is no fog and $o_k = 1$ when fog occurs. Both $0 <= o_k <= 1$ and $0 <= y_k <= 1$ (bounded by zero and one) leading to the Brier Score ranging from 0 to 1. N is the total number of forecast times. The index k indicates a numbering of the forecast-event pairs N. So, (y_k, o_k) is the kth of N pairs of forecasts and observations. Perfect forecasting will occur when the Brier Score is equal to 0. This means that mean squared error of the forecast should be as small as possible.

After the Brier Score is calculated, the Brier Skill Score can be calculated as shown in Equation 3.2.

$$BSS = 1 - \frac{BS}{BS_{ref}} [\%]$$

(3.2)

The BS_{ref} is often obtained from a climatology analysis of a very large time series, at the KNMI often of 50 years. Since this research only has observations for the last 10 years, the mean values of these observations is used. This is called the sample-climatology. The sample-climatology is still a trend, but a smaller data-set than normally used when computing the Brier Skill Score.
The sample-climatology and therefore the BS_{ref} is computed using Equation 3.3:

$$BS_{ref} = \frac{1}{N} \sum_{t=1}^{N} (\overline{o} - o_t)^2 [0 - 1]$$

(3.3)

In which $\overline{o}$ is the mean observation:

$$\overline{o} = \frac{1}{N} \sum_{t=1}^{N} o_t$$

(3.4)

The index t is used for the number of observations, since the index k corresponds to the length of the forecast data-set, which differs from the length of the observation data-set.

The relative BSS difference to compare two models is defined as in Equation 3.5:

$$\text{Relative BSS difference}(x, x_{ref}) = \frac{x - x_{ref}}{x_{ref}}$$

(3.5)

This is the relative difference of the BSS of a model x, compared to the reference BSS of a model x_{ref}.

Bootstrapping

In order to get insights in the uncertainties in the Brier Skill Scores computed, bootstrapping is used. The sampling process is mimicked using random sampling with replacement. A subset of 100 random values from the different data-sets is obtained with a repetition of 1000. The BSS is calculated for these 1000 subsets for all three models (HARMONIE TAFG, HIRLAM TAFG and ECMWF TAFG), all four lead times (2, 4, 8 and 16) and their 4 LVP phases (M, A, B, C). After this, the sample mean $\overline{x}$ is calculated using Equation 3.6, in which x_n is the BSS calculated in the bootstrap set.

$$\overline{x} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n) \tag{3.6}$$

After this the standard error of the bootstrap sets can be calculated indicating the uncertainties in the calculation of the BSS seen in Equation 3.7. In which x_i is the BSS calculated in the bootstrap set and $\overline{x}$ is the mean calculated BSS of the bootstrap set. The index n is the subset size.

$$\overline{\sigma} = \sqrt{(\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}))} \tag{3.7}$$

Reliability diagram

The reliability diagram gives an overview of how well the forecasted probabilities correspond to the observed frequencies, with an example given in Figure 3.10. This can not be mistaken with the accuracy of the model.

The modelled forecasts are grouped in 10 bins, of each 10%, representing their probability, visible on the x-axis. The is done for all TAFG models (HARMONIE TAFG, HIRLAM TAFG, ECMWF TAFG) at all lead times (2, 4, 8 and 16) and their forecast of the LVP phase occurrences (M, A, B and C). The average frequency of how much these events are observed for each of the 10 groups of the forecasts is shown on the y-axis. The observed values are the computed LVP phases from the METAR observations of RVR and ceiling. In the most ideal situation, there is a linear diagonal (x=y) visible in the graph between predicted probabilities and frequencies. This means that if an event will happen with a probability of 25%, the event will happen 25% of the times at which the statement is made. The reliability is determined by the proximity of the the curve to that of the diagonal. The conditional bias is given by the deviation from the diagonal. The reliability diagram is a visual analysis, making its interpretation sometimes more difficult.
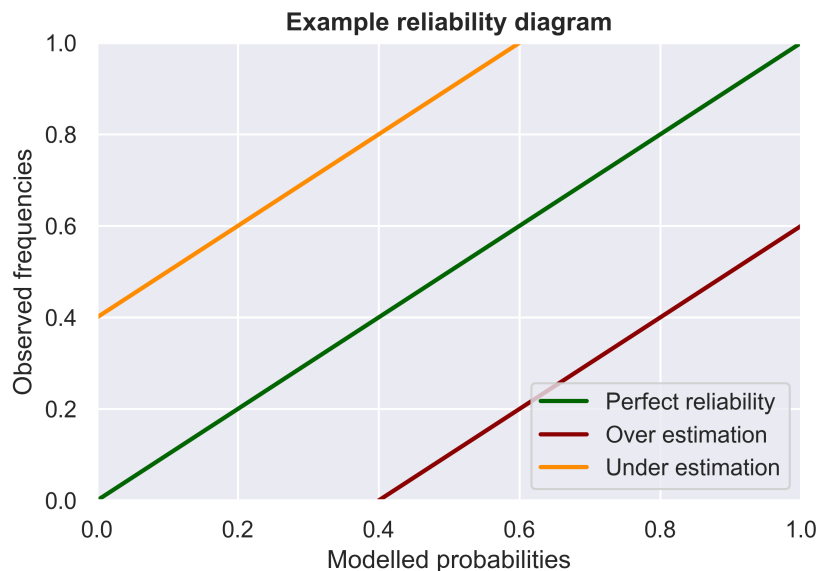
Figure 3.10: Example graph to explain how to read a reliability diagram. The x-axis show the modelled probabilities for the TAFG models from 0 to 1. The observed frequencies are on the y-axis. The colored lines show an example of the interpretation of the position of a line in the reliability diagram. The area above the green line is of an underestimation and the area beneath the green line in of an overestimation

The area beneath the green diagonal (x=y) indicates an overforecasting of the event with too high of a probability of occurrence. The area above the green diagonal (x=y) represents an underforecasting of the event. This means the event occurs more often than the probabilities indicate.

Receiver operating characteristic
The receiver operating characteristic (ROC) is the true positive rate as a function of false positive rate, while increasing the probability threshold (i):

$$\text{True positive rate}(i) = \frac{\text{hits}(i)}{\text{hits}(i) + \text{misses}(i)} \tag{3.8}$$

$$\text{False positive rate}(i) = \frac{\text{false alarms}(i)}{\text{false alarms}(i) + \text{correct negatives}(i)} \tag{3.9}$$

The threshold (i) is the probabilistic value for which the TAFG forecast is in accordance with the observed LVP occurrence. For example a threshold of 0.6 gives that a modelled forecast of 0.7, with an observed fog occurrence, is a hit. The further definition of hits and misses are defined in chapter 3.3. The equations Equation 3.8 and Equation 3.9 are repeated for thresholds from 0 to 1. Plotting the true positive rates and false positive rates for different threshold results in the ROC diagram visible in Figure 3.11. The ROC is obtained for all TAFG models (HARMONIE TAFG, HIRLAM TAFG, ECMWF TAFG) at all lead times (2, 4, 8 and 16) and their forecast of the LVP phase occurrences (M, A, B and C), compared to the computed LVP phases from the METAR observations of RVR and ceiling.

The area beneath the curve represents the score of the ROC curve calculated, with an area of 1 for a perfect score and an area of 0 for bad forecasting. The ROC value gives an insight in the ability of the forecast model to distinguish events and non-events. This functions as a representation of the resolution of the forecasted outcome. An example on how to interpret a ROC graph is given in Figure 3.11. A perfect classifier will result in a graph with a point in the upper left corner.
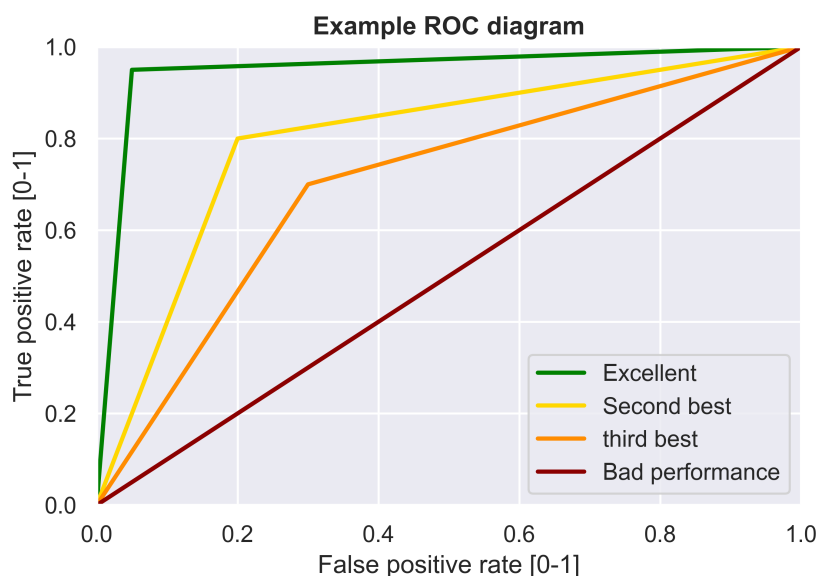
Figure 3.11: Example of how to interpret a ROC graph. The x-axis is the false positive rate and the y-axis is the true positive rate. The point lines represent the position of a line in the graph and how to interpreter the performance.

3.4.3. Spatial variability analysis
The second subquestion is the influence of the spatial variability of the surface characteristics on the performance of the TAFG at Schiphol Airport. The TAFG output includes probabilistic forecasts for one location at the center of Schiphol Airport. As mentioned before, Schiphol has a large area with runways located in different surface characteristics like the concrete surface near the Buitenveldertbaan and the Polderbaan located in the polder. These surface characteristics can have a large impact on the fog

formation. So, the spatial variability in surface characteristics can result in different fog occurrence at different locations. Therefore, LVP phase observations at varying locations at Schiphol Airport are compared to see if there is a difference in fog occurrence.

A verification has been done for the performance of the TAFG output compared to the observations at the center of Schiphol. This verification will be repeated and the TAFG output will be compared with different observation locations and the measured LVP phases at these locations. Since the fog occurrence could be different due to the spatial variability of surface characteristics, the performance in predicting the LVP phase occurrence of the TAFG models could also be different for different locations. Therefore, the spatial variability of surface characteristics and their influence on the performance of the TAFG models will be investigated using observations at different locations.

There are four locations at which the RVR is observed, these are marked in Figure 3.12. This method uses the predicted LVP occurrence of The performance indicator BSS is computed at all four locations for the LVP phases and lead times of all three TAFG models. This is done according to the BSS definition in chapter 3.4.2. The difference in the performance indicator of the four observation is investigated using the relative BSS difference in Equation 3.5. The reference score is that of location 1.

Location 2 is expected to have less fog occurrence due to the more concrete surroundings. Furthermore it is expected that location 3, the Polderbaan, has the highest fog occurrence of the four measuring locations. This is due to the Polder location of this observation point and more vegetation of the surface, with higher water availability. Since the TAFG is produced for the more central location of Schiphol Airport, it is expected that the observed fog occurrence at the Polderbaan location 3, deviates the most from the TAFG output. Therefore, it is expected that the TAFG will have a lesser Skill for the RVR location 3, compared to the RVR location 1.

Figure 3.12: Locations of the four RVR observations at Schiphol airport visible as blue boxes around the measuring locations. The numbers next to the boxes correspond to location names used in this research

3.4.4. Test cases

The performance indicators as described in chapter 3.4.2 give insights in the overall performance of the TAFG models. However, the cause behind an over- or underestimation of LVP phase occurrence is not determined by the performance indicators. In order to find causes in inaccurate LVP forecasting, the accuracy of the temperature, humidity and wind speed are tested for test cases and the biases of the variables compared to the observations can result in potential causes for inaccurate LVP forecasting. Therefore, the test cases will be used to answer the third subquestion: *Are the temperature, humidity and wind speed a cause for the inaccuracies in the low visibility forecasts of the TAFG?*

As defined in Chapter 2, the fog simulation is highly dependent on the models accuracy in forecasting the temperature, humidity and wind speed. These variables will be investigated in the test cases, in order to find a cause for the inaccuracies in the forecast of LVP phase occurrence. The dew point temperature, 2 meter temperature and wind speed are the variables that will be investigated. First the dewpoint temperature can be used as input to calculate the saturation vapor pressure, with Equation 2.1. After this, the specific humidity can be calculated:

$$q_{sat} = \epsilon \frac{e_{sat}}{p} [g/kg] \tag{3.10}$$

Timestamps from the three TAFG model outputs are divided over test case categories. The test case category that includes a timestamp is determined by the combination of the TAFG model performance of LVP phase C forecasts, the input NWP model and the lead time. This leads to 36 test case categories, which are visualized in Figure 3.13. Every test case category can therefore include multiple timestamps with corresponding meteorological variables from the forecasted TAFG output of November 2020 until March 2021.

Four LVP phases M, A, B and C are available. Only the most severe low visibility procedure, phase C, will be used in these test cases. This is because of the high impact this LVP phase has on the aviation operations and the high costs related to inaccurate forecasting of a severe fog event. The main selection is the performance of modelled TAFG output of the LVP phase C forecasts: over estimation, under estimation and good estimation of the LVP phase C. The three main categories are evaluated for the three models HARMONIE TAFG, HIRLAM TAFG and ECMWF TAFG and for their four lead times investigated in this research.

In order to find causes in inaccurate LVP phase C forecasting, the accuracy of the 2 meter temperature, dewpoint temperature and wind speed of the METAR observations compared to the TAFG output is obtained for these test cases. For example: does an overestimation of LVP phase C occurrence for HARMONIE with a lead time of 4 hours have a over- or underestimation of the dewpoint temperature?
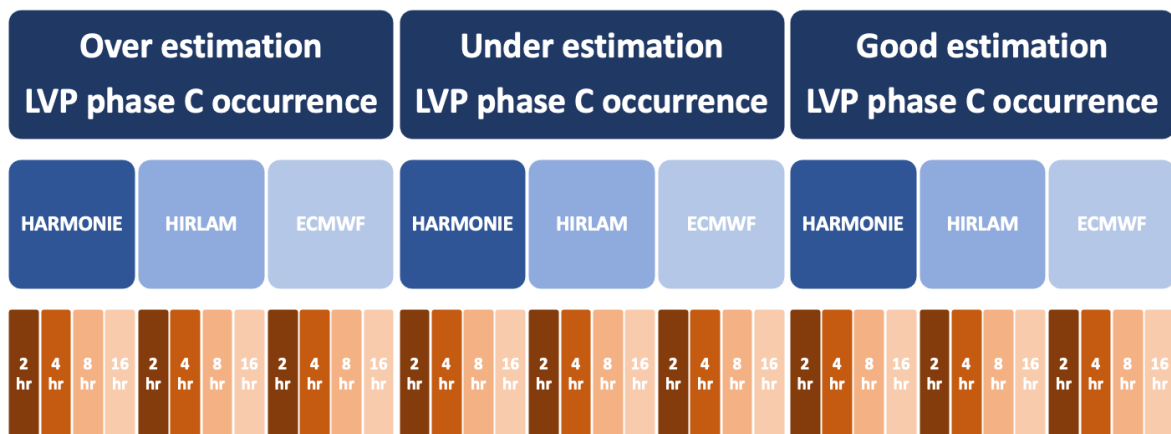
Figure 3.13: Overview of the 36 test case categories for which the performance of the temperature, dew point temperature and wind speed are tested. These cases are used to find a cause for the performance of the TAFG output.

There are a few assumptions made in the final case selection:

- Because the over- and under estimations of LVP phase C occurrence are used, the Brier Skill Scores can't be used. This is an indicator of the Skill without distinction between over- and underestimation. Therefore, the absolute error of LVP phase C forecasting compared to observations is used.
- Since the output of the TAFG models is probabilistic, a threshold should be used to select the over- and underestimation of the TAFG output. Timestamps with a 30% overestimation are selected for the overestimation phase and timestamps with a 60% underestimation are selected for the underestimation phase. A different threshold is used, due to the limited data available. These thresholds will result in at least 5 timestamps with over or underestimation in the test phase, highering the threshold will results in only a few or no timestamps. A good estimation timestamp is one that is lies in 1% over- or underestimation.
- The direction of a bias, positive or negative, is not always constant between different LVP phases for the same timestamp. For example, there can be a 86% overestimation for phase C, but a 24% underestimation for phase B for the same forecast time. This will not be taken into account.

In order to find causes of the inaccurate LVP phase C forecasting, the accuracy of the 2 meter temperature, dewpoint temperature and wind speed are tested and there biases compared to the observations can result in potential causes for inaccurate LVP forecasting.

4

Current performance

This chapter will show the results of the analysis described in Chapter 3. First a climatology analysis is done to check if the occurrence of fog at Schiphol Airport is inline with literature. Afterwards, three sub questions will be answered in the subsections of this chapter. This is done by evaluating the performance of the three TAFG models, the effect of spatial variability of surface characteristics on the TAFG model performance and lastly by finding the causes for the inaccuracies in the performance of the TAFG models using test cases.

4.1. Climatology analysis

First the LVP phase occurrence trends of the past ten years are obtained using METAR observational data from the KNMI. The results will be compared to trends from the past 50 years shown in section 3.4.1. and the theory of fog formation in Chapter 2.

Figure 4.1 shows a decrease in LVP occurrence for all the phases of at least 30% in 10 years.
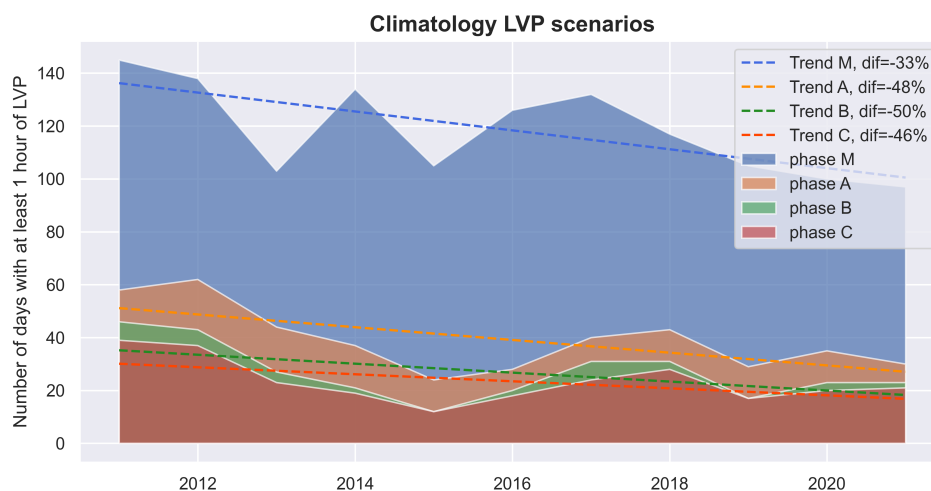
Figure 4.1: The number of days with at least 1 hour of LVP phase from 2010 to 2020. The x-axis shows the time in years. The y-axis show the number of days with at least 1 hour of LVP phase obtained from observations. The colors indicate the LVP phases and their trend lines are visible in the legend. The legend also gives the percent difference of occurrence between 2010 and 2020 ($\frac{\text{occurrence } 2010 - \text{occurrence } 2020}{\text{occurrence } 2010} * 100\%$)

It is visible that the LVP phases with operational restrictions, phase A, B and C, have a high relative reduction of LVP occurrence. This is in line with the decrease in visibility shown in the separate visibility parameters MOR and ceiling from Vautard et al. (2009) and Boers et al. (2015). This reduction in low visibility is a positive trend for aviation operations, since it's occurrence has a high impact on the safety, efficiency and productivity and economics (Gultepe et al., 2019). It is is also visible that there

is only a small difference between the occurrences of phase B and C, with some years even the same number of days with at least one hour of LVP. In Table 3.1 it is visible that there is a different threshold for visibility between phase B and C, but the threshold of 200 ft for the ceiling for both phases. When the occurrence of phase B and C is the same, it could mean that the cause for LVP phase occurrence is the exceedance of the ceiling threshold. This could indicate that the ceiling threshold exceedance is the main cause for LVP phases B and C in these years.

Figure 4.2 shows the distribution of LVP occurrence over the day for the past ten years. In literature it was found that fog usually occurs on cold clear nights. This is also visible in the graph. The fog occurrence increases at 8 PM CEST in the evening, with the peak of the fog occurrence at 4 AM CEST. After this, solar radiation can result in more mixing of the air mass and therefore more dissipation of fog could be the cause for the decrease in fog occurrence. It is also visible that there is almost no difference in the occurrence of phase B and C during the day, but there is a difference in occurrence at night. The Figure 4.3 and Figure 4.4 show that the cause for LVP occurrence is mostly due to exceedance of ceiling thresholds. Furthermore, LVP occurrence caused by RVR threshold exceedance is more common at night, with almost no LVP occurrence during the day caused by exceedance of this threshold. This can be the cause for almost no difference for the LVP B and C occurrence during the day, since the threshold for ceiling is the same for LVP phase B and C and this is main cause for occurrence during the day.
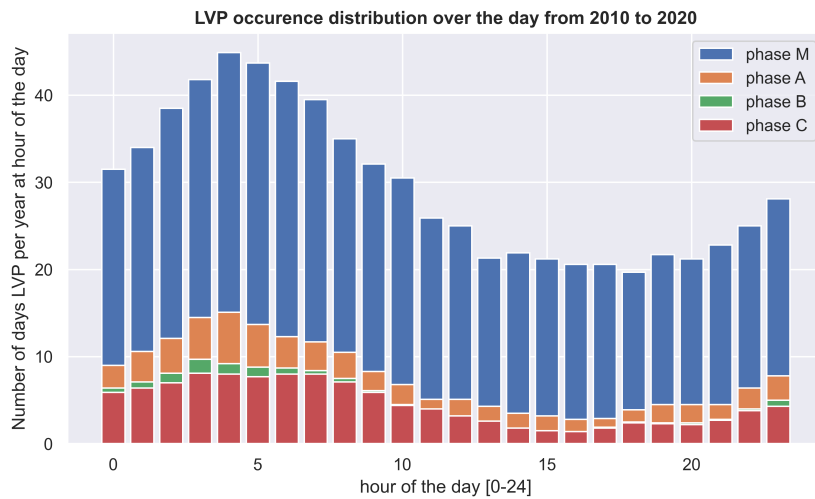
Figure 4.2: Distribution of LVP occurrence throughout the day from 2010 to 2020. The x-axis shows the CEST time of the day from 0 to 24. The y-axis shows the number of days per year with at least one hour of LVP occurrence. The colors indicate the four LVP phases defined in the legend.
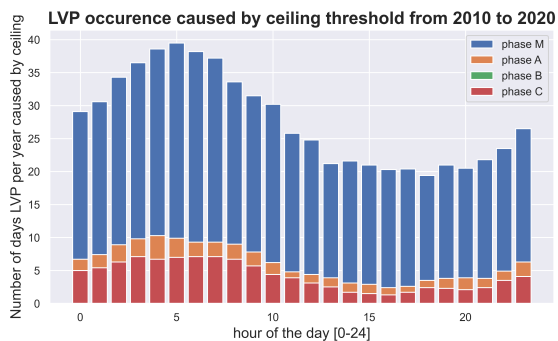
Figure 4.3: Distribution throughout the day of number of days per year with LVP occurrence caused by exceedance of ceiling threshold. The x-axis shows the CEST time of the day. The y-axis shows the number of days per year. The colors indicate the four LVP phases named in the legend.
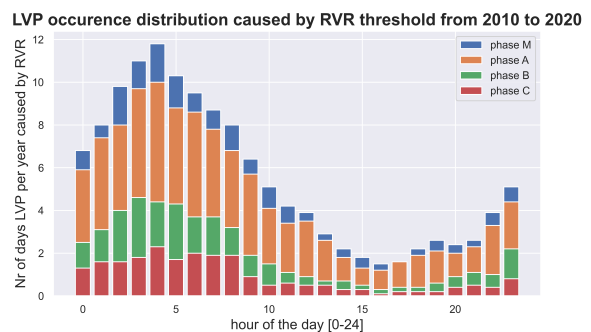
Figure 4.4: Distribution throughout the day of number of days per year with LVP occurrence caused by exceedance of RVR threshold. The x-axis shows the CEST time of the day. The y-axis shows the number of days per year. The colors indicate the four LVP phases named in the legend.

The seasonality of LVP occurrence at Schiphol is computed, visible in Figure 4.5. It is visible that the colder months have a relatively higher LVP occurrence, compared to the warmer summer months. This is as expected from the theory, which identified cold conditions for fog formation and from the Clausius-Clapeyron equation in Equation 2.1. Lower temperatures correspond to a lower saturation specific humidity. This means the air can hold less grams of water per kilogram air and saturation occurs with less available humidity compared to the summer months with higher temperatures. The four LVP phases follow the same seasonality pattern with a peak in January and December and the lowest occurrence in July. It is also visible that the difference between the 25th and 75th percentile is increasing with higher number of days with at least 1 hour of LVP occurrence. This means that there is wider range of number of days in the colder months.

Figure 4.5: Representation of the four LVP phases and their LVP occurrence throughout the year. In these graphs the number of days with at least 1 hour of LVP phase from 2010 to 2020 throughout the months in a year are presented. The four panels in the graph, represent the four LVP phases and their LVP occurrence throughout the year. The boxes correspond to the interquartile range between the 25th and 75th percentile with the median visualized by a line. The black lines are the minimum and the maximum of occurrence found between 2010 and 2020. The diamonds correspond to the outliers found in the dataset.

4.2. Performance of the TAFG model

The performance of the three TAFG model outputs are assessed using the three methods described in chapter 3. This is done in order to answer the first sub question "What are the performances of the low visibility forecasts at Schiphol Airport, with HARMONIE, HIRLAM and ECMWF as an input model for the TAFG compared to the observations.

Brier Skill Score

First the performance is evaluated using the Brier Skill Score (BSS) and relative difference as defined in Chapter 3.4.2. An example of a time-series with a good forecasting skill and the corresponding Brier Score is given in Figure 4.6. These are the results of the METAR observations and the HARMONIE TAFG output and METAR observations for LVP phase B. It is visible that when the observations and TAFG forecast correspond, there is a high Brier Score.
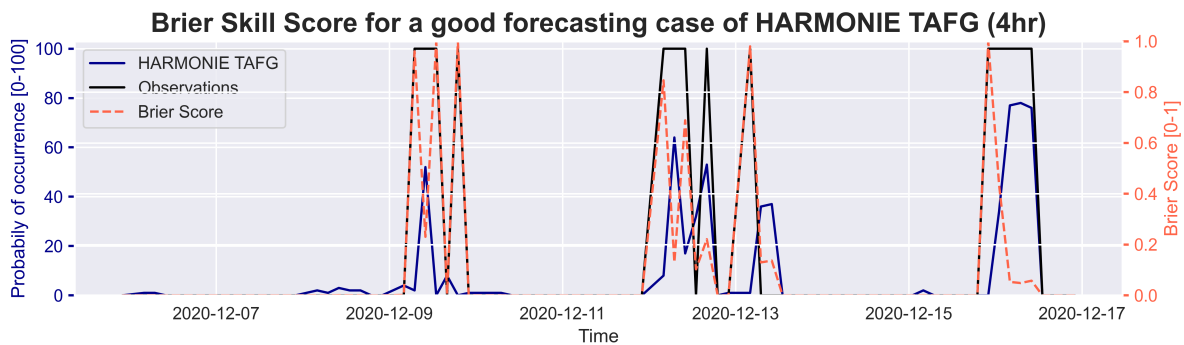
Figure 4.6: An example of the Brier Score for a good forecast of the HARMONIE TAFG with a lead time of 4 hours for LVP phase B is given. The left y-axis gives the probabilities for both TAFG output as well as the observations. The right y-axis gives the Brier Skill scores. The x-axis shows the time.

Table 4.1 shows the relative BSS difference for the three TAFG models for the four lead times of 2, 4, 8 and 16 hours. The ECMWF model does not have a BSS for phase A, since there are no probabilistic values given for this TAFG. The phase M is therefore included to have a good representation.

The BSS and their standard error, visible as the black line, are visualised in Figure 4.7 to compare them more easily.

The Table 4.1 and Figure 4.7 show a big difference in the skill of the three TAFG outputs for the LVP phases and lead times. It is visible that HARMONIE TAFG has the best skill for forecasting the phase M. Both the ECMWF TAFG as well as the HIRLAM TAFG have a worst BSS than HARMONIE for all lead times for phase M. However, this is a phase at which there are no flow restrictions or capacity adjustments yet. The phases A and B, subject to adjusted capacity and flow restriction, have a better skill for HIRLAM TAFG compared to HARMONIE TAFG for all four lead times. For LVP phase B HIRLAM TAFG has a relative forecast skill that is a least 35% better than HARMONIE TAFG, increasing for increasing lead times. Their standard error is similar. At first, one can expect a better skill of the HARMONIE TAFG, since this is a more advanced weather model. However, the data-set on which the TAFG is learning is only three years, compared to more than 20 years for HIRLAM TAFG. This can explain the better performance of HIRLAM TAFG.

The most severe LVP phase C has the best skill for all the lead times of the ECMWF TAFG. This is remarkable since this TAFG is currently available for the meteorologists, but often not used compared to the HARMONIE TAFG and HIRLAM TAFG. In this case it also has a better skill for the short-term lead times of 2 and 4 hours. However, it is visible that the standard error and therefore the uncertainty for the ECMWF TAFG is very high for all the four phases and lead times. This uncertainty is visible in Figure 4.7 as the black lines, which represent the standard deviation of the bootstrapped subsets. Due to the high uncertainty in the BSS of the ECMWF TAFG, it is not possible to draw conclusions on whether this model performs better.

Also, it is clearly shown that the skill for the HARMONIE TAFG and HIRLAM TAFG decreases with higher lead times. This makes the application for long term planning of aircraft operations difficult. However, it is visible that ECMWF TAFG BSS remains rather constant, with only a small decrease, for all the four lead times assessed. It could be further investigated if the application of the ECMWF TAFG after lead times of 4 to 8 hours would be interesting for the airport operations, after further research in the uncertainties of this TAFG output.

Lastly, it is visible that the uncertainties for all the BSS of the TAFG model outputs are very high. This is most likely due to a small data set of TAFG model output. The data set is one of three months, including the months with the highest LVP occurrence according to the climatology analysis. This can include more extreme fog events.

Brier Skill Score for different lead times and scenarios at Schiphol
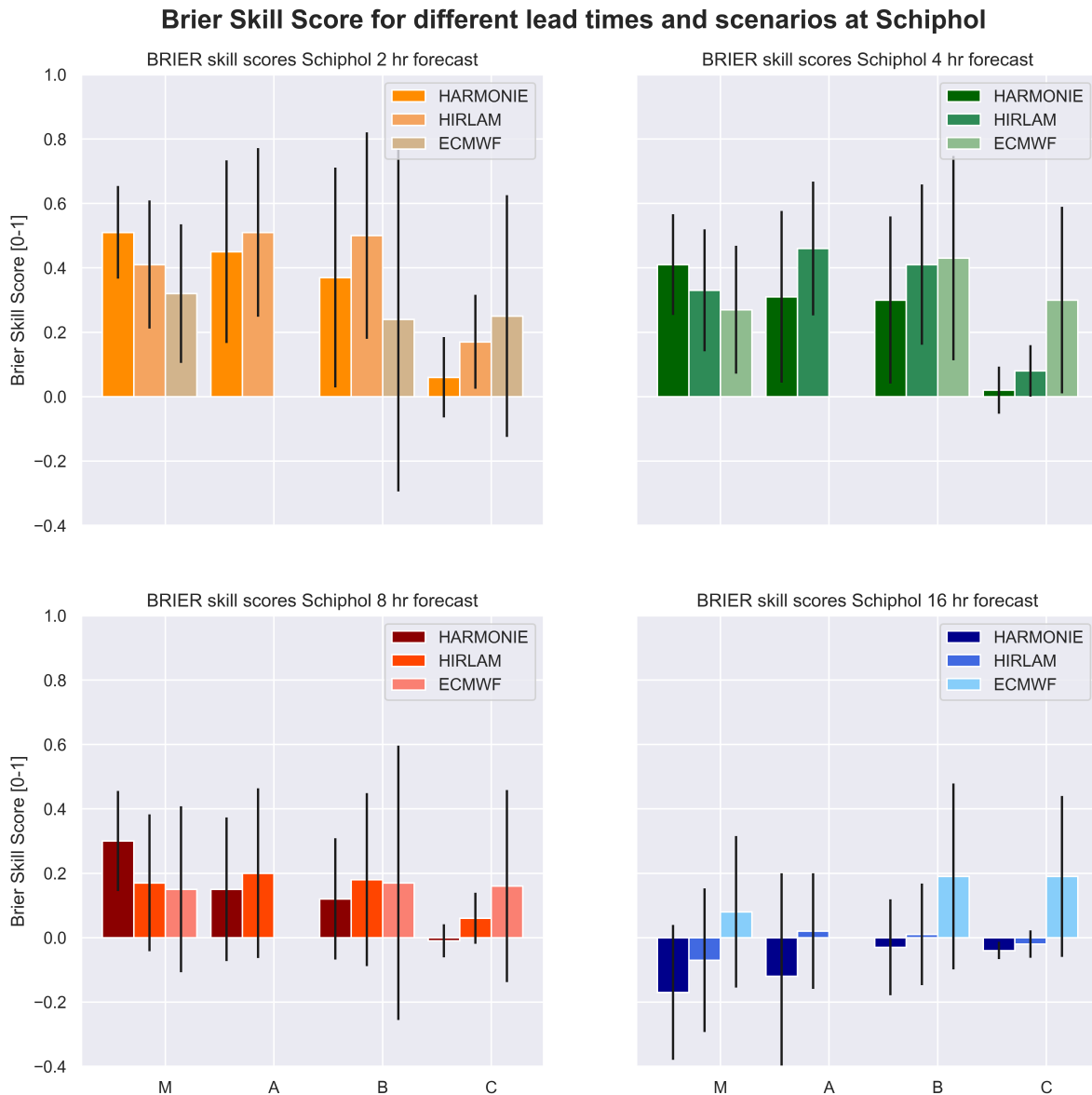
Figure 4.7: Brier Skill Scores and their uncertainties at Schiphol airport. The four panels represent the four lead times of 2, 4, 8 and 16 hours which are analysed. The x-axis of these panels show the four LVP phases. The y-axis indicate the Brier skill scores, in which 1 is good score and a negative score indicates a forecast that is less accurate than using the sample mean. The black lines are the standard errors, obtained by bootstrapping. The three models are compared for each situation with the darker color indicating HARMONIE TAFG, the medium HIRLAM TAFG and the lightest ECMWF TAFG, which is also visible in the legend.

Relative difference	HIRLAM TAFG to HARMONIE TAFG				ECMWF TAFG to HARMONIE TAFG				ECMWF TAFG to HIRLAM TAFG			
Lead time	2hr	4hr	8hr	16hr	2hr	4hr	8hr	16hr	2hr	4hr	8hr	16hr
Phase M	-0.20	-0.20	-0.43	0.59	-0.37	-0.34	-0.5	1.47	-0.22	-0.18	-0.12	2.14
Phase A	0.13	0.48	0.33	1.17	-	-	-	-	-	-	-	-
Phase B	0.35	0.37	0.50	1.33	-0.35	0.43	0.42	7.33	-0.52	0.05	-0.06	18
Phase C	1.83	3.00	7.00	0.50	3.17	14	17	5.75	0.47	2.75	1.67	10.5

Table 4.1: Relative Brier Skill Score difference of the models. The models HIRLAM TAFG, HARMONIE TAFG and ECMWF TAFG are compared for the lead times visible in the columns and the LVP phases visible in the rows. A negative value indicates a worst BSS of the first model compared to the second and a positive values indicates a better BSS of the first model compared to the second. No probabilities for ECMWF TAFG phase A are computed by the ECMWF TAFG and therefore the skill difference is also not obtained.

Reliability diagram

The second performance indicator is the reliability diagram as defined in Chapter 3.4.2. Figure 4.8 gives the reliability graph of all the three TAFG model outputs and their lead times for the four LVP phases. The graph is very extensive, but since this is a visual analysis, needed for the conclusions.

It is visible that for phase M, not subject to airport operation restrictions, there is a structural overestimation of the fog events by all the TAFG models and lead times. For all the TAFG models and lead times the perfectly calibrated line is approximately followed. It is also visible that the HARMONIE TAFG has the least deviation from the perfectly calibrated line.

For phase A, the three TAFG models and their lead times do not follow the perfectly calibrated line. There is still a general overestimation of events. However, the three TAFG models and their lead times do not show one bias. For example the 2 hour HARMONIE TAFG shows an overestimation for most probabilities, but an underestimation for the 0.50 probability.

For phase B there is both over- as well as underestimation of the events. These lines do not follow the perfectly calibrated line and no clear relationship is visible. The reliability is not very well for most of the models.

The last phase C show that there is a very low reliability for most models and lead times. The ECMWF TAFG has the best reliability for all four lead times for phase C. This is in line with results from the BSS. It is also visible that both HARMONIE TAFG as well as the HIRLAM TAFG have a large underestimation for the phase C for short lead times.

It is notable that HARMONIE TAFG shows a general underestimation fog LVP phase occurrence, since the HARMONIE model showed an overestimation of fog occurrence in research done by Román-Cascón et al. (2019).
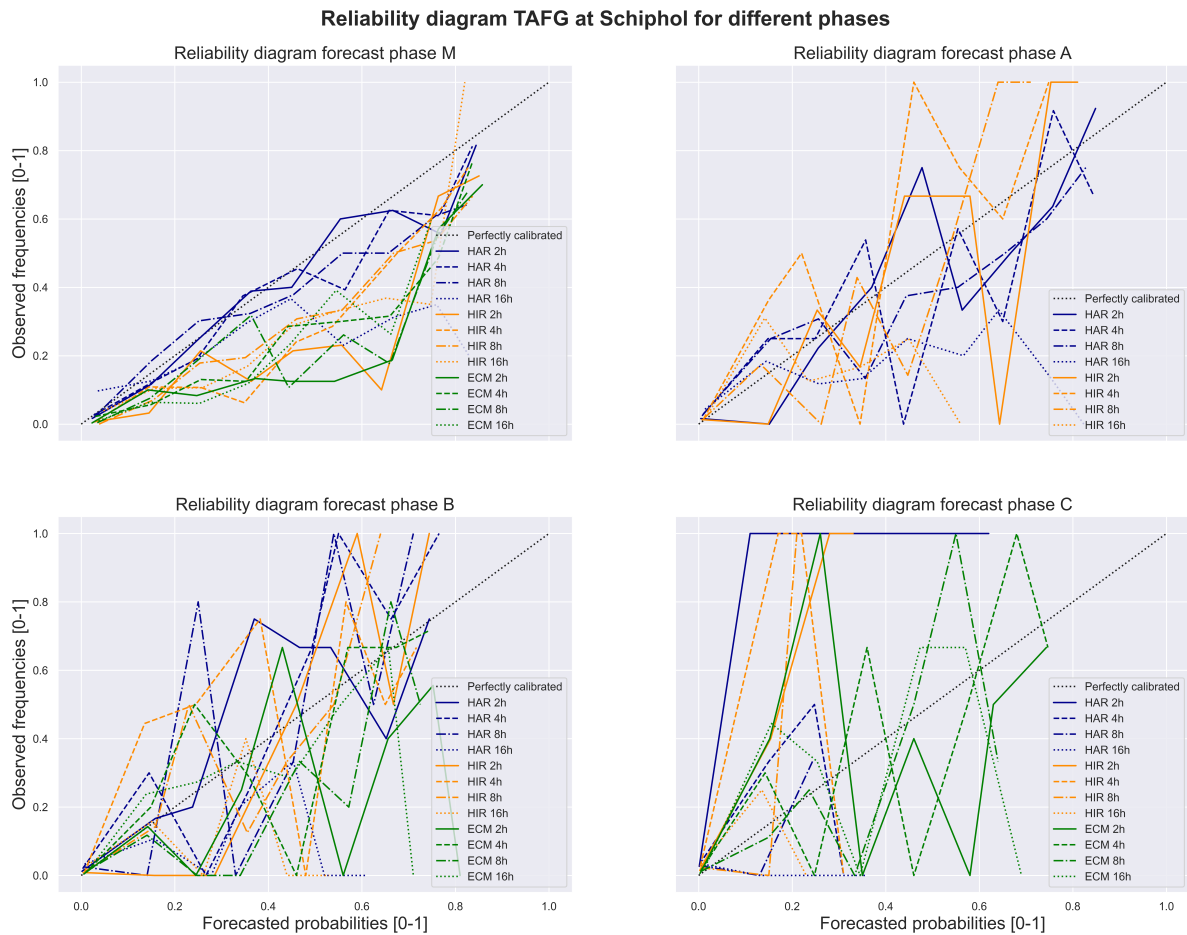
Figure 4.8: Reliability diagrams at Schiphol Airport for the TAFG output. The panels show the four LVP phases. The x-axis are forecasted probabilities of the TAFG model output and the observed frequencies on the y-axis are the METAR observations. The legend shows the datasets for which the reliability diagram is calculated. The colors indicate the model type and the linestyle the lead time of this model. The black line is the line in which the model is perfectly calibrated to the observations and therefore shows good reliability. Reliable model output will follow the pattern of this line.

It can be said that the reliability decreases with the phases becoming more severe low visibility. Furthermore, the chaotic graphs describe the unreliability of the TAFGs. Smooth graphs would be expected for reliable data with a consistent bias. These graphs also show that the limited sample size of three months is quite small, leading to noisy outcomes.

Receiver Operating Characteristic

The last performance indicator is the receiver operating characteristic (ROC) as defined in Chapter 3.4.2. The Figure 4.9 shows the true positive rate as a function of the false positive rate, so the trade-off between the sensitivity and the specificity, for different discrimination thresholds at Schiphol. The threshold is the probabilistic value for which the prediction is considered as an event to occur and is varied from 0 to 1. The values for true positive rate as function of the false positive rate for increasing thresholds, result in the graph visible in Figure 4.9. In the legend the area under the curve is also shown, indicating the performance of the TAFG and its lead time with a high area corresponding to a good performance. This means that lines closest to the top left corner of the graph have the best performance and the largest area. The color groups show the models, for example all HARMONIE TAFG output is colored blue for all the lead times. The linestyle shows the lead time of the TAFG model. For example the blue dashed line is the HARMONIE TAFG output with a 4 hour lead time.
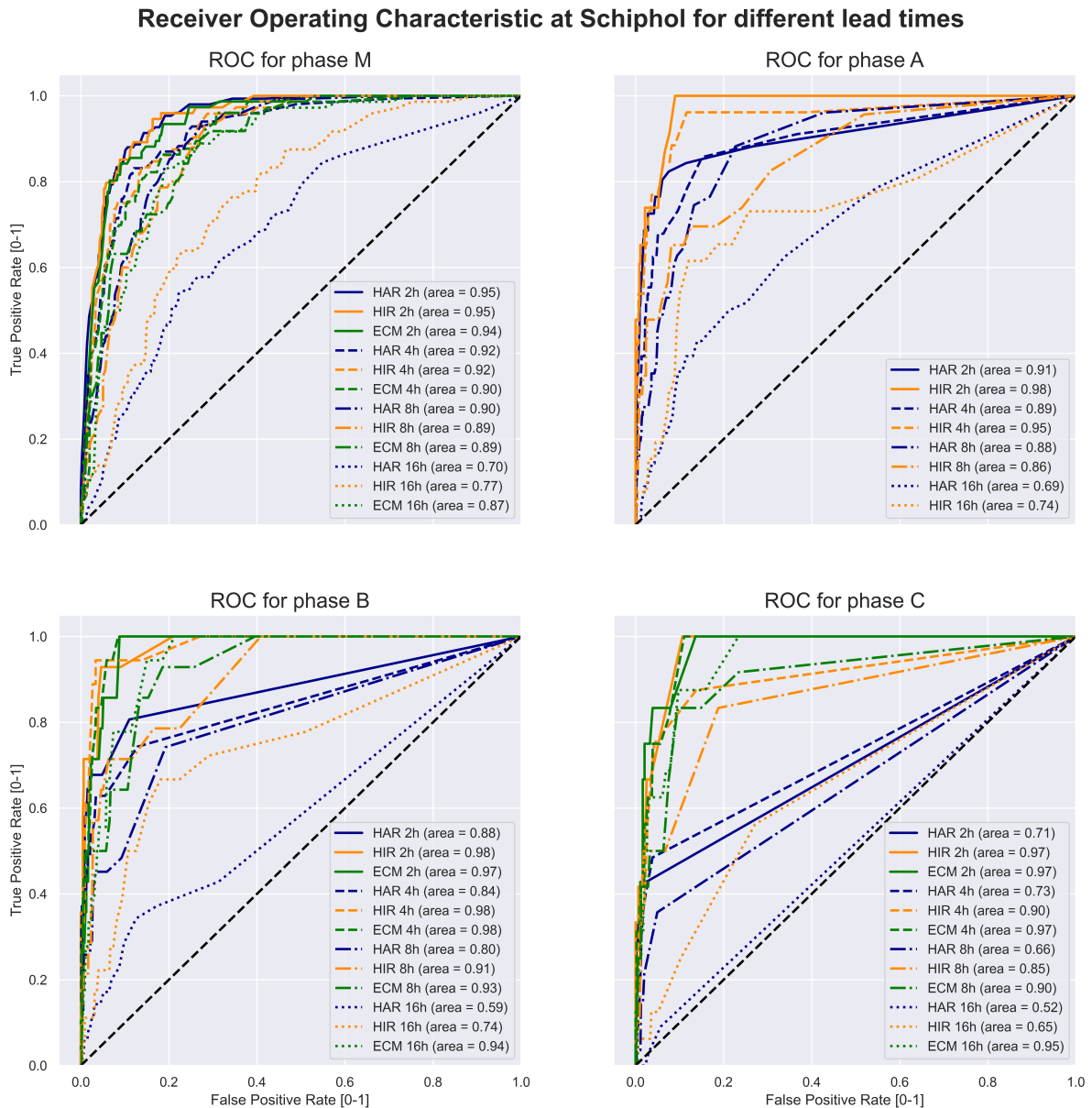
Figure 4.9: Receiver Operating Characteristic at Schiphol for the four LVP phases visible in the panels of the graph. The x-axis show the false positive rate (misses) and the y-axis the true positive rates(hits) for varying thresholds from 0 to 1. The threshold is the probabilistic value for which the prediction is considered as an event to occur. The legend shows the model outputs that are assessed and the area corresponding with the line of the graph. A good ROC score is a line that is close to the top left corner, with a high area under the graph.

For phase M there is a clear decrease in performance, when the lead time becomes larger for the HARMONIE TAFG and HIRLAM TAFG. This is visible in the dotted colored lines which are closer to the black dotted line. However, this does not count for the ECMWF TAFG, which are the green lines. This performance stays high when the lead time increases, which is also visible in the high value of 0.87 of the area beneath the curve. For the HARMONIE TAFG and HIRLAM TAFG the area decreases for longer lead times for phase M.

Phase A shows the same decrease in performance for higher lead times. It also shows the better performance of HIRLAM TAFG compared to the HARMONIE TAFG for the 2, 4 and 16 hour lead times. This can be again due to the larger training data set of the HIRLAM TAFG. For the 8 hour lead time, performances are similar for HARMONIE TAFG and HIRLAM TAFG.

Phase B shows the decrease of performance with higher lead times for both HARMONIE TAFG and HIRLAM TAFG. With HARMONIE TAFG with a lead time of 16 hours only having an area of 0.59, compared to 0.94 for ECMWF for the same lead time. This means that the HARMONIE TAFG performance, according to the ROC performance indicator, is less compared to ECMWF TAFG at this long lead time. The HARMONIE TAFG forecasting performance defined by the ROC is lower compared to both HIRLAM TAFG as well as ECMWF TAFG. The overall performance forecasting for phase B of ECMWF TAFG is high and stays high for high lead times. The performance of the HIRLAM TAFG is also high for 2, 4 and 8 hour lead times, but decreases for 16 hour lead time. These results are in line with the results of the Brier Skill Score performance indicator.

Phase C forecasted by the HARMONIE TAFG, shows a low performance compared to the other TAFG model outputs, with a 0.52 area score for the 16 hour lead time. The HIRLAM TAFG performance well on the short term with a lead time for 2 and 4 hours, but decreases to 0.65 area score for higher lead times. The ECMWF TAFG has a high performance with all scores higher than 0.9 for all lead times in forecasting phase C. Even for the short lead times, it outperforms both HIRLAM TAFG as well as HARMONIE TAFG.

It can be said that based on this performance indicator, the HARMONIE TAFG does not perform very well compared to HIRLAM TAFG and ECMWF TAFG considering the true positive rate compared to the false positive rate. For phase A HIRLAM TAFG has the best performance and for the more severe phases B and C, the ECMWF TAFG has the best rate. This is again remarkable, since this is the currently lesser used TAFG model. The ROC is also of great economic importance, since misses can result in high costs.

Taking into account all three performance indicators for this data-set, it can be said that the performance of ECMWF TAFG is better compared to HARMONIE TAFG and HIRLAM TAFG. This is remarkable, since this TAFG model output is not used as often operationally compared to the other TAFG models. However, the uncertainties for this TAFG model performance are very high and more research should be done using a larger data-set to confirm the findings.

4.3. Spatial variability

After the difference of performance of the TAFG model output is assessed, the second sub question can be adressed. The effect of the spatial variability of the surface characteristics on the performance of the TAFG is assessed according to the method described in Chapter 3.4.3.

The runway and observation locations at Schiphol have a high variability in surface characteristics, it is also expected to have a high variability in fog occurrence. This is visible in Figure 4.10, where the LVP occurrence at the RVR observation locations are plotted from data obtained by measurements of the KNMI. In this figure, location 3 has the largest deviation in the LVP phase A,B and C occurrence pattern. More days with at least one hour of LVP are measured for the past years for location 3 compared to location 1, 2 and 4. This is as expected, since this is the location that is in the Polder near Schiphol with more water available due to the grass environment compared to the concrete surrounding the other locations. The air mass will be more saturated with a large water budget, leading to more condensation of water vapor to liquid water and therefore more fog formation.
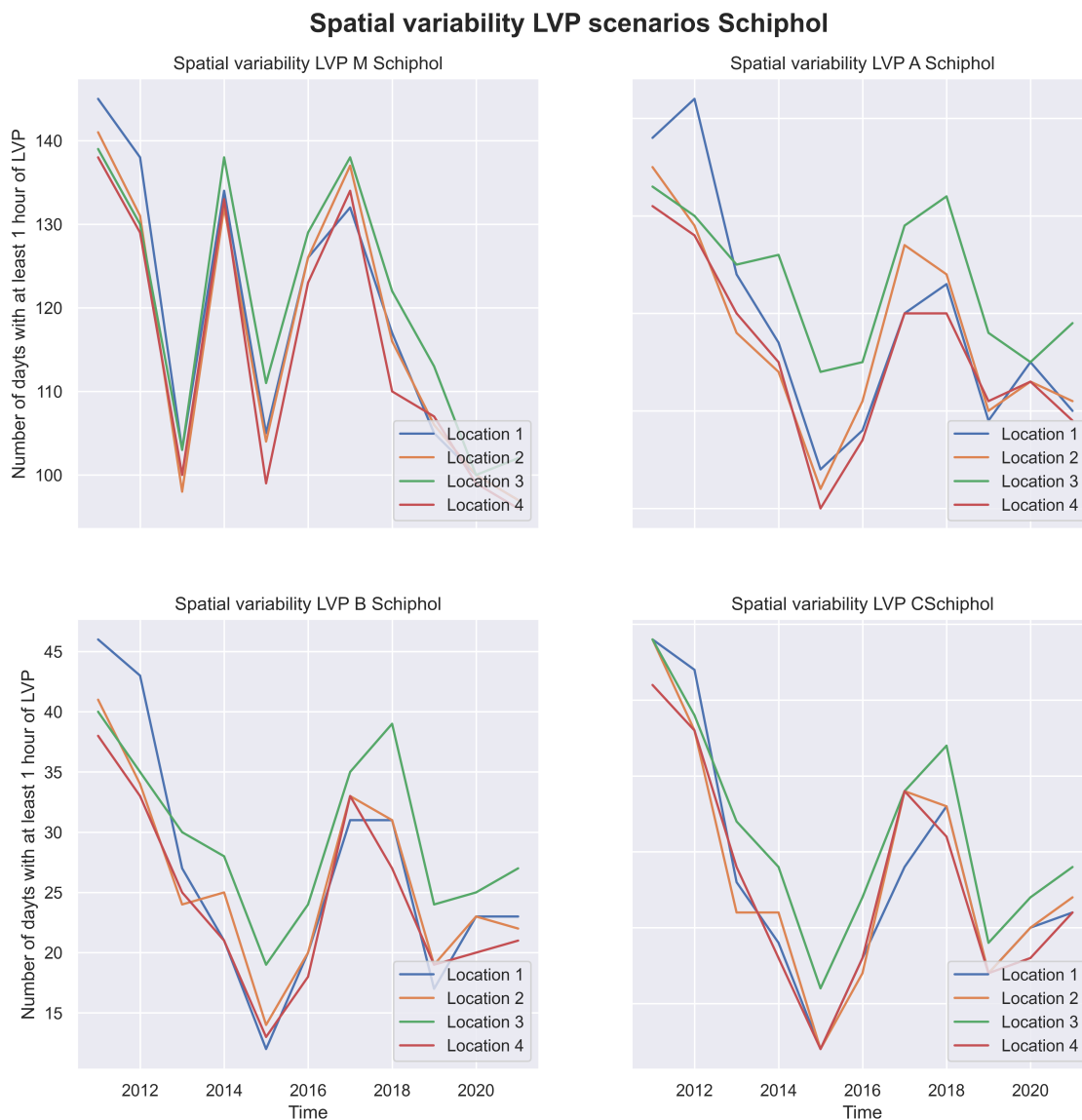
Figure 4.10: Difference of the number of days with at least 1 hour of LVP between the RVR observation locations at Schiphol airport from 2010 to 2020. The four panels represent the four LVP phases. The x-axis is the time in years and the y-axis the number of days in that year with at least one hour of LVP. The occurrence is plotted for the four observation location, which are visible in the legend. Location 3 is the Polderbaan situated in the Polder. Location 1,2 and 4 are situated around the main building with concrete runways and buildings.

The Brier Skill score is computed for the four RVR observation locations and the difference of the BSS compared to the BSS of RVR location 1 is visible in Figure 4.11 and Figure 4.12 to compare the performance. The 4 and 16 hour lead times are visible in this chapter and the 2 and 8 hour lead times are visible in the Appendix A.

There is no clear pattern visible in the BSS for the 4 observation location. However, there are a some remarks:

- HARMONIE TAFG has overall the best performance for the RVR location 1, which is in the centre of Schiphol Airport and most BSS do not differ more than 0.05 in Skill score compared to the BSS from RVR location 1
- HIRLAM TAFG does not have a clear pattern in better or worse performance compared to RVR location 1. The BSS for phase A with a 4 hour lead time has a better performance, but this is not the same for the other lead times of 4, 8 and 16 hours.
- ECMWF TAFG has a lower performance for the RVR locations 2,3 and 4, except for the lead time of 2 hours. This could be due to the mid and longterm specifics of the ECMWF weather model.
- Overall the performance of the RVR location 3 is the lowest. This is as expected, since this location has a very different surface characteristic compared to the other runways.
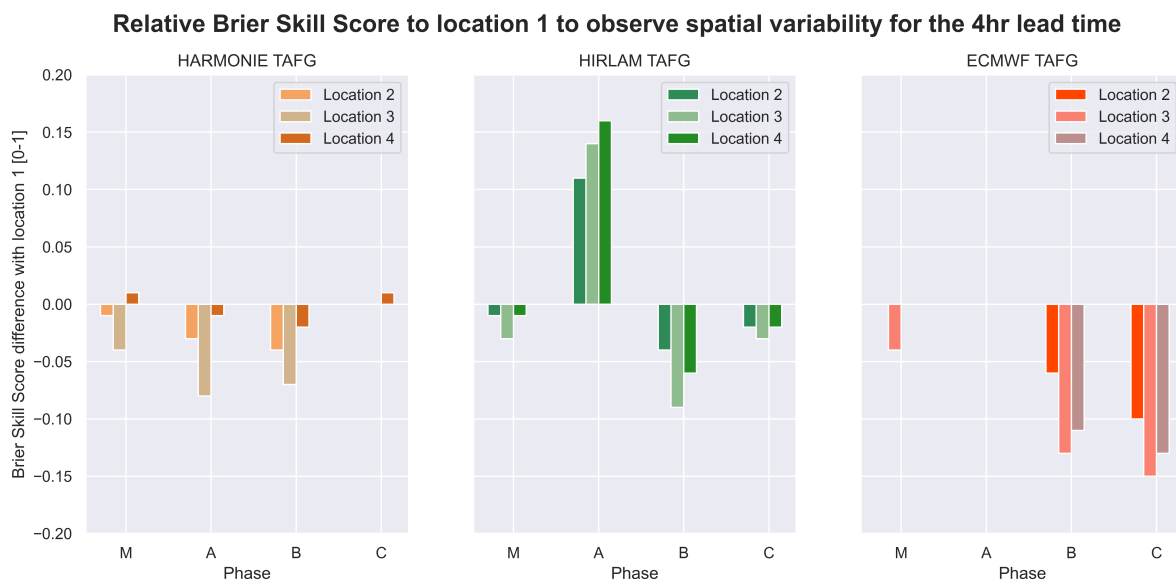
Figure 4.11: Relative Brier Skill Score of the TAFG for a lead time of 4 hours at location 2, 3 and 4 compared to RVR location 1. The three locations are visible in the legend. The panels represent the three TAFG models. The x-axis show the LVP phases. The y-axis shows the BSS difference with location 1. A negative difference, indicated a lesser performance compared to the TAFG performance at location 1.

The Polderbaan (RVR observation location 3) has a lower skill score compared to those of the other locations at Schiphol Aiport. This is the difference in performance of the TAFG. For the actual SKV, the aviation meteorologist can adjust the probabilities based on own observations or experience. Also, the performance of the TAFG at locations 2 and 4 is comparable to the performance of the TAFG at location 1 for all models and lead times. It could be that surface characteristics are very similar at locations 1, 2 and 4 and therefore the TAFG output is sufficient for all these locations. This should be further investigated before conclusions can be made.

Relative Brier Skill Score to location 1 to observe spatial variability for the 16hr lead time
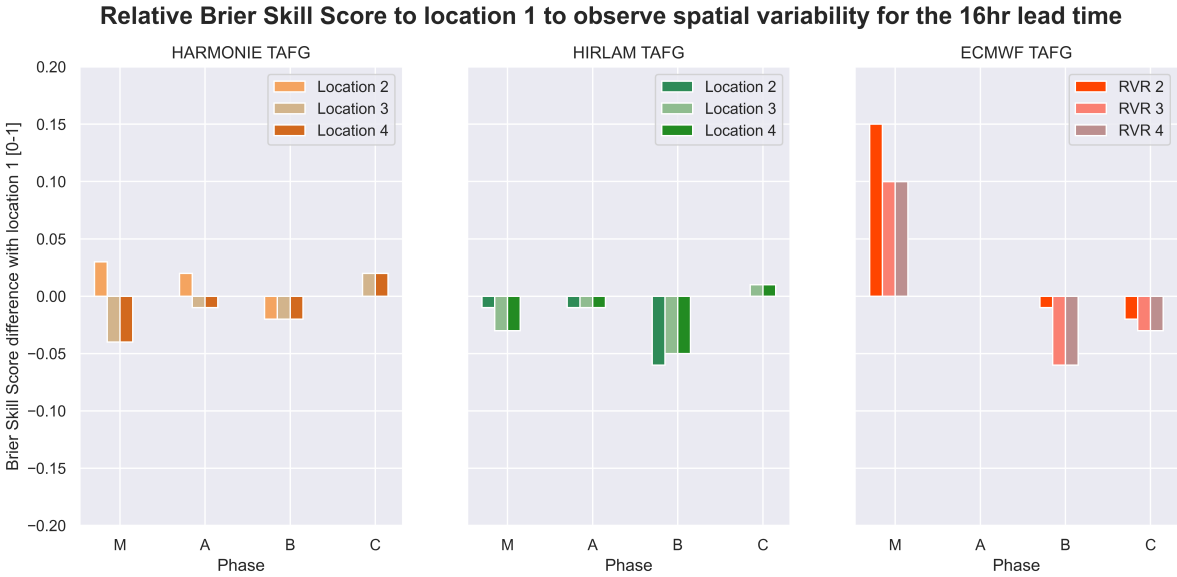
Figure 4.12: Relative Brier Skill Score of the TAFG for a lead time of 16 hours at location 2, 3 and 4 compared to RVR location 1. The three locations are visible in the legend. The panels represent the three TAFG models. The x-axis show the LVP phases. The y-axis shows the BSS difference with location 1. A negative difference, indicated a lesser performance compared to the TAFG performance at location 1.

4.4. Test cases

The previous sections 4.2 and 4.3 provide answers to sub questions one and two. In this section the causes for inaccuracies in the TAFG LVP phase C forecasts are sought, as described in subquestion three. Izett et al. (2019) describe fog as a "secondary process" for which a correct simulation depends on a correct simulation of the temperature, humidity and wind conditions. When incorrectly simulating these conditions, it can have a significant impact on the fog simulation results. This chapter will first provide an overview of the correlation between the forecasted TAFG output and observations of the variables temperature, dewpoint temperature and windspeed. After this, the timestamps of the forecasted dataset will be divided over the 36 test case categories described in section 3.4.4. For these test case categories the mean difference of the TAFG output and observations of the variables temperature, dewpoint temperature and windspeed are computed. Based on these results of the three variables, conclusions will be drawn for the cause of inaccurate LVP phase C occurrence.

First, the correlation of the observed METAR observations and modelled TAFG outputs at Schiphol Airport are compared for the variables temperature, dewpoint temperature and windspeed. Figure 4.13 shows a regression plot of the wind speed for the observed and modelled values, for the three available TAFG outputs for the dataset from November 2020 until March 2021. The x=y line shows a perfect calibration between the TAFG model output and the observations.
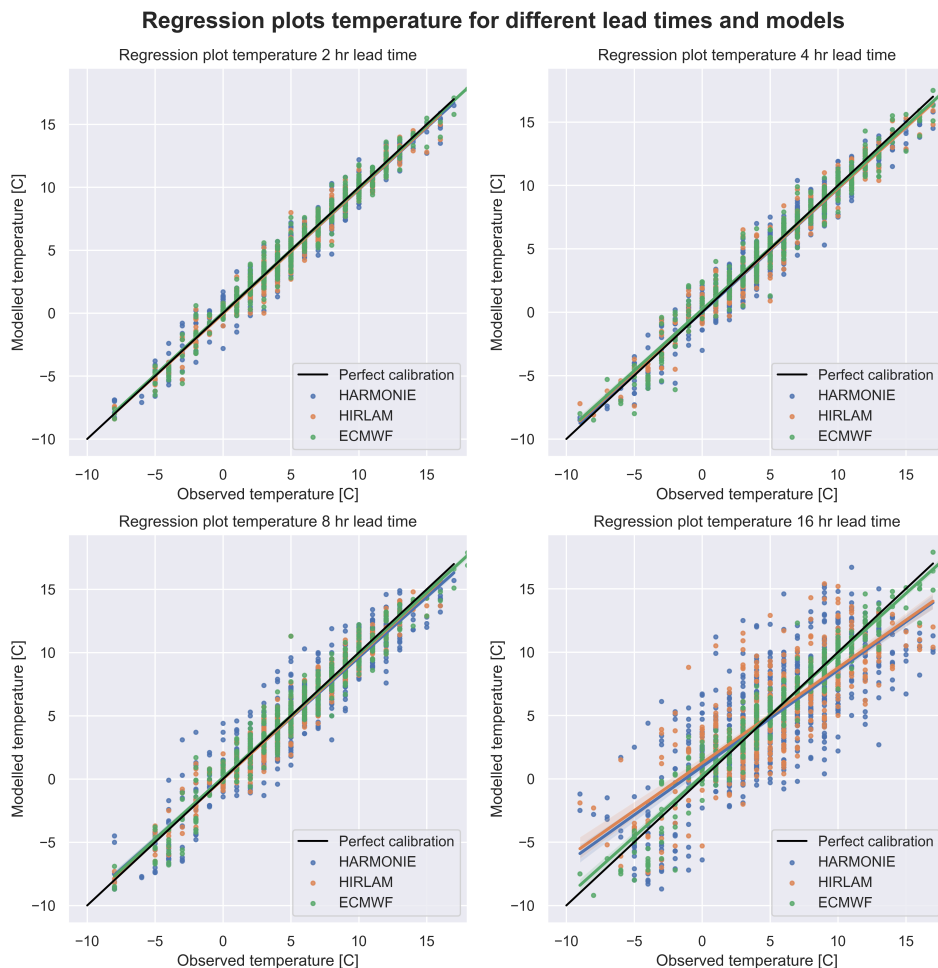
Figure 4.13: Regression plot of the observed and modelled wind speed at Schiphol Airport from November 2020 until March 2021. The panels are the four lead times that are considered in this research. The x-axis are the METAR observations and the y-axis the modelled TAFG output. Each panels plots the datapoints for the three TAFG models with colors visible in the legend. The colored lines are the linear regression lines between the observed and modelled data. The translucent bands show the confidence intervals of the plotted regressions. The x=y line is the situation in which observations and model output are the same.

It is visible in Figure 4.13 that there is not a lot of difference between the correlations of the TAFG model outputs with the observations of wind speed for the lead times of 2, 4 and 8 hours. However, in the regression plot of the 16 hour lead time it is visible that the ECMWF TAFG model output is closer to the observed wind speed. The same effect is visible for the correlations of the temperature between the TAFG model outputs and the observations in Appendix D and Figure D.1.
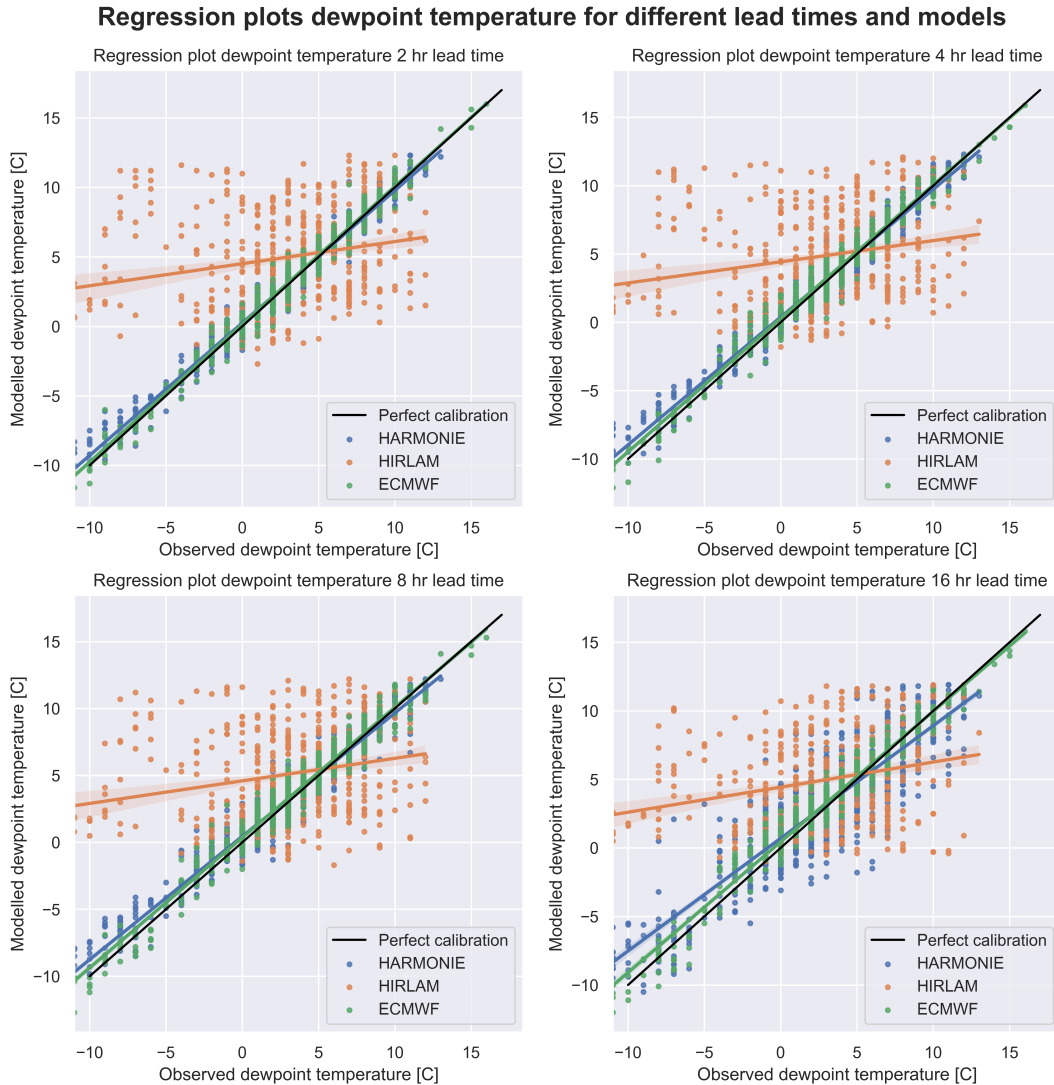
Figure 4.14: Regression plot of the observed and modelled dewpoint temperature at Schiphol Airport. The panels are the four lead times that are considered in this research. The x-axis are the observed values and the y-axis the modelled TAFG output. Each panels plots the datapoints for the three TAFG models with colors visible in the legend. The colored lines are the linear regression lines between the observed and modelled data. The translucent bands show the confidence intervals of the plotted regressions. The x=y line is the situation in which observations and model output are the same.

Figure 4.14 shows a large deviation for the HIRLAM TAFG of the observed compared to the modelled dew point temperature. To find the cause for this deviation in humidity, the time series of the dewpoint temperature of HIRLAM TAFG is obtained. It is visible in Figure 4.15 that there is an large difference in accuracy of the dewpoint temperature forecast of HIRLAM TAFG compared to HARMONIE TAFG and ECMWF TAFG.

The figure also shows a large overestimation between the observed dew point temperature compared to the HIRLAM TAFG output for the period from 5 to 15 February 2021, highlighted with the red box. This was a period of extreme cold temperatures in The Netherlands.
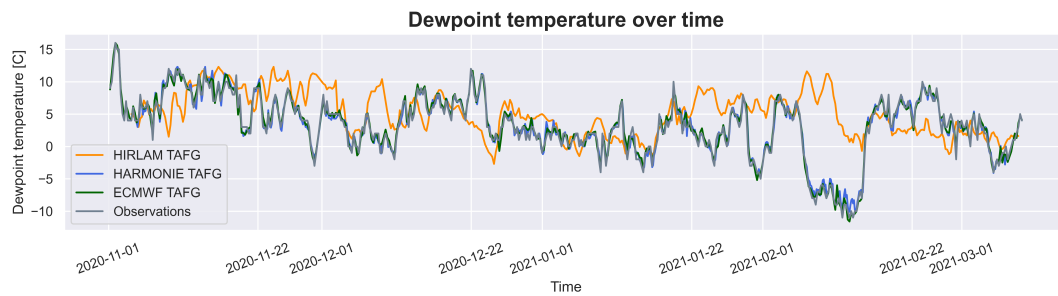
Figure 4.15: Dewpoint temperature over time for the TAFG output of HIRLAM, HARMONIE and ECMWF and observed data, indicated by the legend.

A more detailed image is visible in Figure 4.16. It is very striking that this inaccurate forecasting of the dewpoint temperature only occurs for HIRLAM TAFG and not for ECMWF TAFG and HARMONIE TAFG. Also, for the overall time series there is a larger difference of the HIRLAM TAFG dewpoint temperature compared to observation than the differences of HARMONIE TAFG and ECMWF TAFG. The data is double checked and no error was found. Raw model output should be compared to see if the cause is in the NWP model or the TAFG.
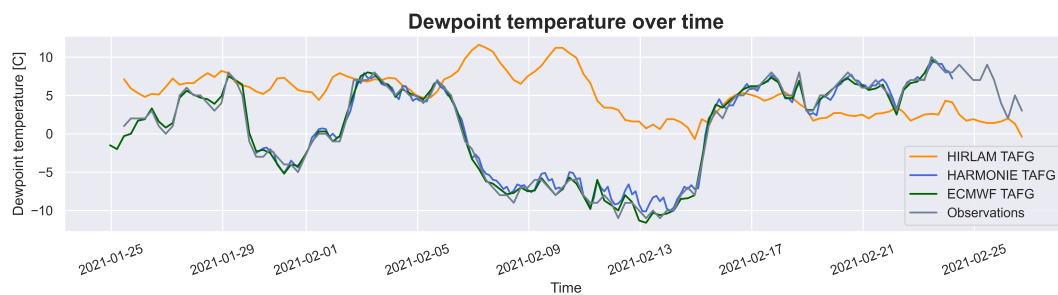
Figure 4.16: Dewpoint temperature over time for the TAFG output of HIRLAM, HARMONIE and ECMWF and observed data, indicated by the legend.

This overestimation of dewpoint temperature of the HIRLAM TAFG output is an overestimation of the humidity and can result in more saturation of the simulated air mass and therefore an overestimation of fog. However, the probability of occurrence for LVP phase A for HIRLAM TAFG visible in Figure 4.17 does not show a larger overestimation of LVP A occurrence, compared to the other TAFG outputs.
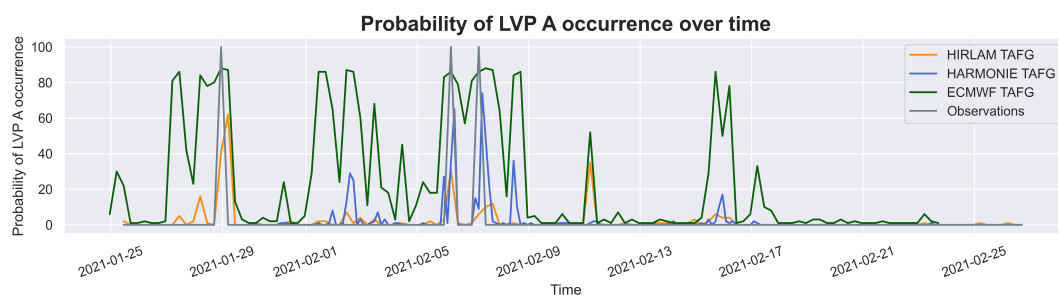
Figure 4.17: LVP phase A occurrence over time for the modelled and observed data, indicated by the legend.

The regression plots (Figure 4.14, Figure 4.13, Figure D.1) give an indication of how well the variables 2 meter temperature, dewpoint temperature and wind speed are predicted by the TAFG, but a direct relation to the performance of the TAFG LVP performance is still lacking. Test case groups are determined in section 3.4.4 based on an over estimation, under estimation or good estimation for the most severe LVP phase C, NWP model type and lead times. For these test cases with several timestamps and corresponding variables the TAFG forecasts for the variables 2 meter temperature, dewpoint temperature and wind speed are also compared to the METAR observations from November 2020 until

March 2021. In the following section the expected relationships between the 2 meter temperature, dewpoint temperature and wind speed and the LVP occurrence will be described.

The temperature is a measure for the heat budget of the air mass as seen in Equation A.15. A simulated overestimation of the temperature, can be the result of too less simulated radiative cooling. Radiative cooling is needed to reach the dew point temperature and saturation. An overestimation of the temperature can therefore lead to an underestimation of the low visibility occurrence. Therefore, it is expected to have an underestimation of LVP occurrence , and simulated fog, when there is an overestimation of the 2 meter temperature.

The dewpoint temperature at which the air is saturated with moisture. A simulated overestimation of the dewpoint temperature will indicate a higher humidity of the air mass. When there is more water available the air will be more saturated and therefore more condensation will take place. Therefore, it is expected to have an overestimation of LVP occurrence, and simulated fog, when there is an overestimation of the dewpoint temperature.

Wind is needed for advection fog, for the warm moist air to be transported over cold surfaces. However, too much wind will result in mixing of air masses and dissipation of fog. Also, for radiation fog to occur, stable conditions with low speed are required. Also, no wind can lead to no mixing of heat and moisture of the air masses resulting. In this situation the longwave radiation can dominate and therefore radiation fog can occur. So, both an underestimation as well as an overestimation of wind speed can also lead to an overestimation of fog and LVP phase occurrence.

The mean differences of the observed and modelled output of the variables are visible in Figure 4.18, Figure 4.19 and Figure 4.20. The mean difference $\overline{d}$ is defined as Equation 4.1, in which m are the modelled values and o the observed values:

$$\overline{d} = \frac{1}{N} \sum_{i=1}^{N} m_i - o_i \qquad (4.1)$$

d will have positive value for an overestimation and a negative value for an underestimation of the TAFG output compared to the METAR observations. By taking the mean of the differences of a number of timestamps, positive and negative values can compensate for one another. This should be kept in mind while evaluating the results.

Figure 4.18: Mean temperature difference between TAFG and observations for the test cases. The panels represent the TAFG output for the three input NWP models. The x-axis are the type of estimation for the LVP phase C. The y-axis represent the difference between the modelled and observed temperature. The results are obtained for all four lead times as indicated in the legend.

Mean dewpoint temperature difference between model output and observations for the test-cases
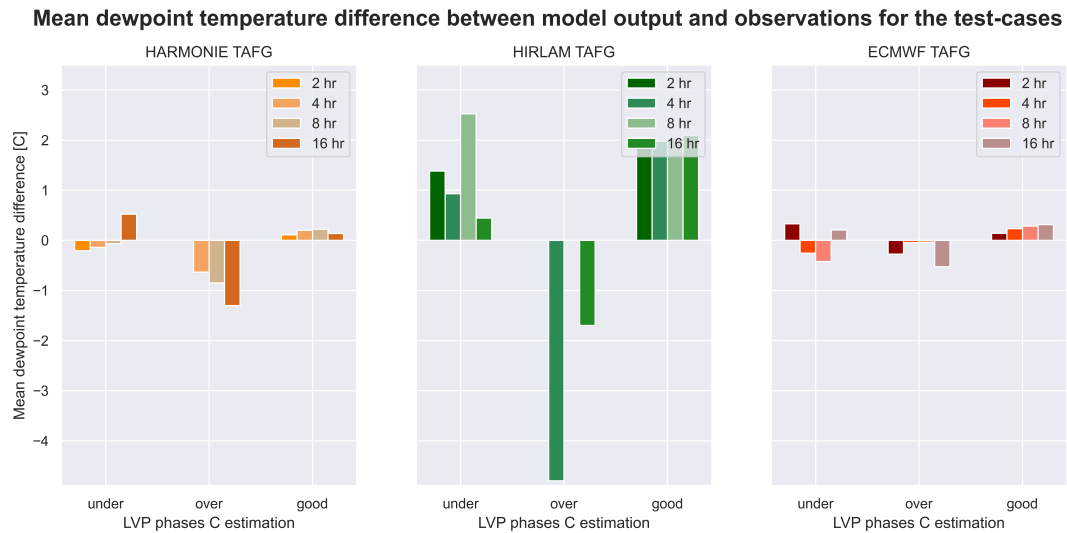
Figure 4.19: Mean dewpoint temperature difference between TAFG and observations for the test cases. The panels represent the TAFG output for the three input NWP models. The x-axis are the type of estimation for the LVP phase. The y-axis represent the difference between the modelled and observed dewpoint temperature. The results are obtained for all four lead times as indicated in the legend.

Mean windspeed difference between model output and observations for the test-cases
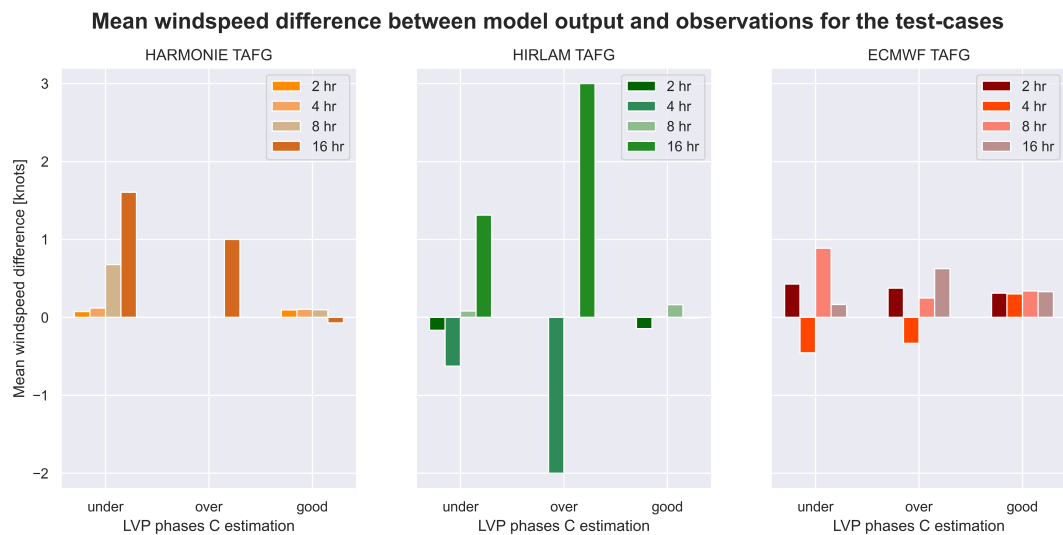
Figure 4.20: Mean wind speed difference between TAFG and observations for the test cases. The panels represent the TAFG output for the three input NWP models. The x-axis are the type of estimation for the LVP phase. The y-axis represent the difference between the modelled and observed wind speed. The results are obtained for all four lead times as indicated in the legend.

It is visible that the results of the mean difference of the TAFG model outputs and observations differ in direction (over or under) for different lead times. For example, the under estimation of LVP phase C for ECMWF TAFG, shows both an over- as well as an underestimation of the wind speed, seen in Figure 4.20. For other test cases, there is a constant direction of the mean difference between TAFG output and observations for all lead times of a model. An example of constant direction is the over estimation for LVP phase C and the under estimation of the 2 meter temperature for HARMONIE TAFG seen in Figure 4.18.

Figure 4.21 shows the summary of results. The test case categories are visible in the blue boxes (under- over- and good estimation of LVP phase C). The orange boxes show the overall pattern of the bias of the TAFG output compared to the METAR observations for the variables 2 meter temperature, dewpoint temperature and wind speed. An example is the top left orange box: this shows an under

estimation of the 2 meter temperature computed by the HARMONIE TAFG, for an over estimation of the LVP phase C.

Figure 4.21: Summary of the test case results. The upper dark blue boxes show the three test case categories: an over, under and good estimation of the LVP phase C. The orange boxes give the result for the mean bias of a variable for the TAFG output compared to the METAR observations. For example: the 2 meter temperature for HARMONIE TAFG has a mean under estimation for an overestimation of LVP phase C occurrence. The rows indicate the variables that are investigated, 2 meter temperature, dewpoint temperature and wind speed.

The three investigated variables have varying results:

2 meter temperature

For the 2 meter temperature it was expected to have an overestimation, with an underestimation of LVP phase C occurrence and the other way around. This is the case for the overestimation test cases for all three TAFG output and for the underestimation test cases for the HARMONIE TAFG and ECMWF TAFG.

The 2 meter temperature forecast of the HARMONIE TAFG shows an overestimation for underestimation of LVP phase C occurrence and underestimation for overestimation of LVP phase C occurrence. These results are in line with the hypothesis made earlier in this chapter. However, Román-Cascón et al. (2019) showed that the low surface temperatures were not captured accurately for HARMONIE. This could mean that there has been a good correction of the raw NWP model ouput to the TAFG output. However, conclusions can not be made without further research into the raw NWP model output.

Dewpoint temperature

The hypothesis for the dewpoint temperature is to have an overestimation of forecasted dewpoint temperature for an overestimation of the LVP phase C occurrence. The results show opposite results as the hypothesis for all three models for the overestimation. This should be further studied.

Also the HIRLAM TAFG shows an opposite results for the expected dewpoint temperature. This could be due to the incorrect forecasting of the dewpoint temperature during the cold days of 5 to 15 February.

Constant pressure is assumed in these test cases, but inaccuracies in the pressure will lead to very small deviations of the saturation specific humidity and therefore this effect is neglectable.

Wind speed

The windspeed is expected to have both an under- as wel as an overestimation for an underestimation

of LVP phase C occurrence and the other way around. The HIRLAM TAFG and ECMWF TAFG do not show clear patterns. Since the mean difference is calculated, it can be that both the under- and overestimation compensate for one another, resulting in no pattern for the mean difference.

The test case category of a good LVP phase C occurrence estimation is expected to have a good forecast of all three variables. This means that the 2 meter temperature, dewpoint temperature and windspeed are accurately forecasted when the LVP phase C is also accurately forecasted. This is the case for almost all variables. Only the dewpoint temperature of HIRLAM TAFG shows contradicting results, but this can be explained due to the inaccurate forecasting of the dewpoint temperature from 5 to 15 February shown in Figure 4.15.

The maximum difference of mean temperature is 2 degrees Celcius, the dewpoint temperature 4 degrees Celcius and windspeed 3 knots. This does seem as a large deviation. However, since fog and therefore LVP phases, are a local small scale process, these small deviations can have a high impact. The forecasts of the 2 meter temperature of the TAFG models is as expected for the cases of under and over estimation of LVP phase C occurrence. Since the pattern of the temperature corresponds to the hypothesis, the inaccurate forecasting of the 2 meter temperature could be the cause for the inaccuracies within the LVP predictions of the TAFGs for all three models.
Fog forms due to microphysical interactions that are complex and it's formation is also very dependant on the aerosol composition of the air (Izett et al., 2019). Therefore a correct simulation of temperature, humidity and wind conditions does not guarantee correct fog simulation.

5

Conclusions

Airport operations are highly dependent on safe weather conditions, especially on low visibility conditions. At Schiphol Airport, Low Visibility Procedures (LVP phases M, A, B and C) are defined, for which there are capacity and runway use restrictions. These Low Visibility Procedures are obtained from a post-processing model of a Numerical Weather Prediction model; the Terminal Aerodrome Forecast Guidance (TAFG). This gives a probability for a given LVP phase, computed from both the Runway Visual Range and the ceiling. Currently, the TAFG is operational for three NWP models (HARMONIE, HIRLAM and ECMWF). HIRLAM TAFG was previously used and replaced by HARMONIE TAFG since November 2020, which has a higher horizontal resolution resulting in a more detailed spatial variability and a better representation of convective clouds. ECMWF TAFG is available with more mid- to long-range forecasts, but the model is not often used by the aviation meteorologists. Inaccuracies remain in the LVP forecasts and the physics of fog. Therefore the following research question was studied:

What is the performance of the visibility and ceiling forecast of the TAFG at Schiphol Airport?

Sub questions were defined and these will be answered sequentially:

- *What are the performances of the low visibility forecasts at Schiphol Airport with HARMONIE, HIRLAM and ECMWF as an input model for the TAFG compared to the observations?*

  The HIRLAM TAFG has a better performance for LVP forecasting compared to the HARMONIE TAFG for both the Brier skill score as well as the Receiver operating characteristic, the performance indicators used in this research. This can be explained by the larger seasonality dataset of 20 years used to train the HIRLAM TAFG, compared to 3 years of seasonality data for HARMONIE TAFG.
  Also, the performance of ECMWF TAFG is better for LVP phases B and C, which indicate severe fog occurrence, compared to HARMONIE TAFG and HIRLAM TAFG for both the Brier skill score as well as the Receiver operating characteristic. This is remarkable, since this TAFG model output is not used operationally as often as the other TAFG models. Also, from lead times of 8 hours and above, the ECMWF TAFG performed better for all phases compared to the other TAFG models for the two performance characteristics. Especially a good receiver operating characteristic is of great economic importance, since this shows the false alarm rate for which the taken measures for fog occurrence can lead to high costs. However, the standard errors of the brier skill scores for the ECMWF TAFG model performance are higher compared to the other TAFG models and therefore it can not be concluded that the ECMWf TAFG is significantly better.

- *What effect does the spatial variability in surface characteristics between four observations locations at Schiphol Airport have on the performance of the TAFG?*

  Four locations at Schiphol Airport are studied. The furthest location, the Polderbaan, has a lower Brier skill score for the LVP phase occurrence of all phases for all three TAFG models compared to the Brier skill scores of the other locations. This can be explained by the differences in surface

characteristics of grass compared to concrete. The Polderbaan that has grass in it's surroundings, has a higher water availability compared to the concrete surroundings of the other observation locations. The varying Brier skill scores are based on the TAFG model outputs. For the actual SKV,the aviation meteorologist can adjust the probabilities based on own observations or experience.

The spatial variability of the surface characteristics of the other observation locations did not result in different performances of the TAFG at these locations, based on the TAFG output of all the three models for all the LVP phases.

- *Are the temperature, humidity and wind speed a cause for the inaccuracies in the low visibility forecasts of the TAFG?*

    An overestimation of temperature could suggest an underestimation of the radiative cooling. When the dewpoint temperature, and therefore saturation, is not reached an underestimation of fog and LVP occurrence can be seen. This effect was also observed in the test-cases, with an overestimation of 2 meter temperature when there was an underestimation of LVP occurrence. An inaccurate forecast of the 2 meter temperature could result in inaccurrate LVP forecasts of the TAFG.

    The dewpoint temperature is a representation of the humidity. An overestimation of the dewpoint temperature, would be an overestimation of the humidity and therefore lead to more condensation of fog and an overestimation of LVP occurrence. This effect was not visible in the test-cases. This should be further studied to draw final conclusions.

    An overestimation of wind can result in dissipation of fog and therefore an underestimation of the LVP phase. This effect was not observed in the test-cases.

    Lastly, the test cases did show a good prediction of 2 meter temperature, dewpoint temperature and wind speed for a good LVP forecast.

Overall it can be said that the performance of the ECMWF TAFG is better compared to the other TAFG models for the data-set analyzed in this research. Also, the inaccurate forecasting of the 2 meter temperature could be a cause for the inaccuracies in the TAFG forecast for the LVP phases. One should however keep in mind that, even if the temperature, humidity and wind conditions are simulated correctly, it does not guarantee correct fog simulation. Fog forms due to microphysical interactions that are complex and the formation is also very dependent on the aerosol composition of the air, which is not studied.

5.1. Discussion

This chapter describes the limitations of the research and the conclusions drawn. Recommendations are done for further research, as well as the operational use of the TAFG.

Research

- A large limitation in this research, was the use of the very small data-set of TAFG output. This was due to the small operational time of the HARMONIE TAFG, since November 2020. The data-set covers three months. These were the months with climatologically the highest fog occurrence. Therefore this data-set is still useful for a few LVP occurrences. With high seasonality in weather conditions in the Netherlands and the large return periods in weather phenomena, conclusions are hard to make based on such a small data-set. To confirm the conclusions made from this research, the same analysis should be executed using a bigger data-set. However, many conclusions were based on the ECMWF TAFG and HIRLAM TAFG performance and a large data-set length for these two TAFG outputs would be a good place to start for further research.

- In this research, the data of the raw model output is not assessed. It would be recommended to research if the TAFG imrpoves the skill of the prediction as compared to the raw model output and obtain a quantification of the optimisation of the TAFG models. The NWP models do not have the RVR or the LVP phase forecasts. However, the driving parameters of 2 meter temperature,

dewpoint temperature and windspeed could be studied.

- The finest horizontal grid size of a NWP model used in this research, was 2.5 km grid size of the HARMONIE model. This means that the Schiphol Airport is captured in three grid cells of this weather model. Therefore the level of detail of the forecast is not high. It would be interesting to research the performance of a higher resolution model, embedded in the NWP model. This is already implemented at the Charles de Gaulle airport in Paris. In the Netherlands, one could consider the implementation of the DALES model at Schiphol Airport for forecasting the small scale weather phenomena of fog.

- Another limitation of the research was the use of both deterministic as well as probabilistic parameters. The LVP phases were probabilistic, while other TAFG parameters were deterministic, e.g. temperature. Further research should be undertaken to investigate the quantitative impact of these deterministic parameters on the probabilistic forecasts.

Operational TAFG

- Based on the performance indicators used in this research, it would be advised to study the possibilities increase the importane of the ECMWF TAFG in the SKV determination. However, it is recommended to first further research the impact of the high uncertainties within the ECMWF TAFG on its overall performance. Final conclusion for implementation can be drawn, after researching a larger dataset as recommended before.

- Also, it would be advised to study the possibilities of implementing the ECMWF TAFG for long term predictions from lead times of 8 hours. However, it is recommended to first further research the impact of the high uncertainties within the ECMWF TAFG on its overall performance.

- The large difference of the TAFG performance for the Polderbaan compared to other runways was expected. In the current situation, the aviation meteorologists compensate for these differences using their own observations and experience. Computing a different TAFG for the Polderbaan would be suggested in order to be less dependent on the aviation meteorologist and to make a better forecast on the long term. This could improve the planning of aviation operations.

A
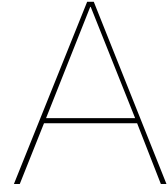
Reynolds decomposition

In order to simulate all these processes of fog formation and development, many models use governing equation for atmospheric dynamics, which also form a basis for the simulation of formation, development and dispersion of a fog layer. These will be described.

Governing equations

The gas law
This is often used in terms of the density of moist air (ρ) and with the virtual temperature (T_v):

$$p = \rho R_d T_v \tag{A.1}$$

Mass conservation
The mass conservation is also often revered to as the continuity equation and states the conservation of mass in a closed system:

$$\frac{\partial u_j}{\partial x_j} = -\frac{1}{\rho}\frac{d\rho}{dt} \tag{A.2}$$

In shallow convection the effect of compressibility is often neglected resulting in:

$$\frac{\partial u_j}{\partial x_j} = 0 \tag{A.3}$$

Momentum conservation
Splitting the momentum conservation equation in three components of the wind gives:

$$\begin{aligned}
\frac{du}{dt} &= fv - \frac{1}{\rho}\frac{\partial p}{\partial x}, \\
\frac{dv}{dt} &= -fu - \frac{1}{\rho}\frac{\partial p}{\partial y}, \\
\frac{dw}{dt} &= -\rho g - \frac{1}{\rho}\frac{\partial p}{\partial z}
\end{aligned} \tag{A.4}$$

In which the Coriolos parameter ($f = 2\omega \sin\phi$) has different signs for the acceleration to the right at northern hemisphere and to the left at the southern hemisphere.

Heat conservation
The energy conservation is often expressed without the molecular diffusion, due to it's negligibly small effect:

$$\frac{dT}{dt} = -\frac{g}{c_p}w + \frac{L_v}{c_p}\frac{dq_l}{dt} - \frac{1}{\rho c_p}\frac{\partial F_i}{\partial x_i} \tag{A.5}$$

49

In which c_p is the specific heat of dry air at constant pressure and ω is the vertical velocity of air. q_l is the liquid water specific humidity and F_i is the added heat (e.g. radiation).

The right hand side has three parts: 1. the vertical displacement 2.the phase change of water vapor to liquid water and 3. the radiation.

With the potential temperature:

$$\theta = T(\frac{p_0}{p})^{\frac{R_d}{c_p}}$$

(A.6)

and Exner:

$$\Pi = \frac{T}{\theta}$$

(A.7)

It gives:

$$\frac{d\theta}{dt} = \frac{L_v}{c_p\Pi}\frac{dq_l}{dt} - \frac{L_v}{\rho c_p}\frac{\partial F_i}{\partial x_i}$$

(A.8)

The liquid water potential temperature is:

$$\theta_l \approx \theta - \frac{L_v}{c_p\Pi}q_l$$

(A.9)

This lead to the heat conservation in terms of liquid water potential temperature as:

$$\frac{d\theta_l}{dt} = -\frac{1}{\rho c_p\Pi}\frac{\partial F_i}{\partial x_i}$$

(A.10)

Total water specific humidity

The energy conservation equation is also often expressed in total water specific humidity. The water vapor specific humidity (q_v) conservation equation is:

$$\frac{dq_v}{dt} = \kappa_{q_v}\frac{\partial^2 q_v}{\partial x_j^2} + E$$

(A.11)

Reynolds decomposition

The state of the atmosphere changes due to diabatic processes, but also due to advection which includes turbulent motions. Turbulence can cause strong fluctuations. The governing equations are not yet specified in a way usually used in models in which the turbulent fluxes can also be described. Reynolds averaging is needed to derive the turbulent flows. The variable will be split into the mean (ϕ) and the fluctuations (ϕ').

$$\phi = \overline{\phi} + \phi'$$

(A.12)

The Reynolds decomposition of the governing equations are derived to split the turbulent transport (at the right hand side of the equations) and large-scale advection (from the second part at the left hand side of the equations) from the mean properties of the air. The large scale advection also includes the large scale subsidence in the air mass.

Mass conservation

$$u'\frac{\partial \phi'}{\partial x} + v'\frac{\partial \phi'}{\partial y} + w'\frac{\partial \phi'}{\partial z} + \phi'(\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} + \frac{\partial w'}{\partial z}) = \frac{\partial u'\phi'}{\partial x} + \frac{\partial v'\phi'}{\partial y} + \frac{\partial w'\phi'}{\partial z}$$

(A.13)

Momentum budget

The Reynolds-averaged momentum equation split in the three components of the wind gives:

$$\frac{\partial \overline{u}}{\partial t} + \overline{u}\frac{\partial \overline{u}}{\partial x} + \overline{v}\frac{\partial \overline{u}}{\partial y} + \overline{w}\frac{\partial \overline{u}}{\partial z} = f\overline{u} - \frac{1}{\overline{\rho}}\frac{\partial \overline{p}}{\partial x} - \frac{\partial \overline{u'w'}}{\partial z},$$

$$\frac{\partial \overline{v}}{\partial t} + \overline{u}\frac{\partial \overline{v}}{\partial x} + \overline{v}\frac{\partial \overline{v}}{\partial y} + \overline{w}\frac{\partial \overline{v}}{\partial z} = -f\overline{v} - \frac{1}{\overline{\rho}}\frac{\partial \overline{p}}{\partial y} - \frac{\partial \overline{v'w'}}{\partial z}, \qquad \text{(A.14)}$$

$$0 = -\rho g - \frac{\partial \overline{p}}{\partial z}$$

Also, the third equation states that there is hydrostatic equilibrium in the mean state. In other words, the mean vertical velocity acceleration is so small it can be neglected.

Heat budget

For the heat budget, horizontally homogeneous turbulence conditions are also assumed:

$$\frac{\partial \overline{\theta_l}}{\partial t} + \overline{u}\frac{\partial \overline{\theta_l}}{\partial x} + \overline{v}\frac{\partial \overline{\theta_l}}{\partial y} + \overline{w}\frac{\partial \overline{\theta_l}}{\partial z} = -\frac{1}{\overline{\rho}c_p}\frac{\partial \overline{F_j}}{\partial x_j} - \frac{\partial \overline{w'\theta_l'}}{\partial z} + S_{\theta_l} \qquad \text{(A.15)}$$

In Equation A.15 the phase change of water is represented by θ. The first part of the right hand side (rhs) of the equation represents the added radiation. This is both longwave as shortwave radiation including the surface heat fluxes. The second part of the rhs represent the vertical displacement. The third term of the rhs is the sink or source term which is precipitation. The turbulent motions in the air mass are dependant on all three components.

Water budget

$$\frac{\partial \overline{q_t}}{\partial t} + \overline{u}\frac{\partial \overline{q_t}}{\partial x} + \overline{v}\frac{\partial \overline{q_t}}{\partial y} + \overline{w}\frac{\partial \overline{q_t}}{\partial z} = -\frac{\partial \overline{w'q_t'}}{\partial z} + S_{q_t} \qquad \text{(A.16)}$$

The turbulent motions ans the rhs of the equation represents the vertical motions and the sink and source term. This sink source term represents the precipitation.

B
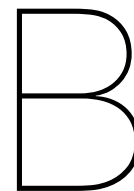
One-dimensional models

A large advantage of one-dimensional models is the low computational costs. Even though it does not always suffice for capturing the whole process of fog formation, it can be useful to get more insights in difficulties in forecasting fog (Bergot et al., 2005). To capture the small scale fog processes, single-column models are used which are externally forced by coarser NWP models with geostrophic winds and advection (Duynkerke, 1991). This gives a better understanding of these small-scale process (Müller et al., 2010) and is useful for operational use (Steeneveld and de Bode, 2018). One-dimensional models are often used for fog forecasts at a specific location, for example airports (Román-Cascón et al., 2019).

An assumption in these models is the homogeneity in the horizontal direction of the thermodynamics. This results in that the main processes within one-dimensional models for fog formation being radiative processes and turbulent fluxes of heat and moisture (Gultepe et al., 2007). Horizontal dynamic processes like horizontal advection or large scale subsidence are often not represented, or integrated from three-dimensional models. This assumption also results in large errors over time in fog formation.

    The soil conductivity and it's non-linear relation with the boundary-layer mixing is also of great importance for the radiative fog in the Netherlands (Steeneveld and de Bode, 2018). The soil conductivity has influence on the near-surface temperature, which is an important factor for time of fog formation. One-dimensional models, however treat them as homogeneous, while patterns of the soil characteristics and vegetation aren't. Patchy patterns often visible in for layers are therefore not well represented (Gultepe et al., 2007).

    Clouds have a decreasing effect on the radiative cooling of the earth's surface. In an one-dimensional models, the sky is represented with complete cloud coverage or total clear sky. However, a condition with partial cloud coverage which occurs more often, is in most one-dimensional fog forecasting models not an option for the state of clouds (Gultepe et al., 2007). It is also shown that the presence of clouds low in the atmosphere, have a sensitivity for the timing at which fog formation occurs (Bergot and Guedalia, 1994). Presence of clouds lead to less atmospheric cooling or and often no formation of fog.

The parameterizations of the turbulent boundary-layer mixing is the most important process for radiation fog researched in the Netherlands (Steeneveld and de Bode, 2018). The parameterization scheme especially had an impact on the timing of the fog formation. However, the stable atmosphere conditions in which fog often occurs, are still not accurately parameterized within models and in one-dimensional models (Gultepe et al., 2007).

    The sensitivity of the model performance of the prediction of fog is different for different one-dimensional models. Literature shows results for small sensitivity (Steeneveld and de Bode, 2018) and a high sensitivity (Haeffelin et al., 2010).

    Also important to state that different processes and parameterization do not add up linearly (Steeneveld and de Bode, 2018). When mutating two process intensities simultaneously, this does not mean that the intensities can be added up. For example the effect of simultaneous perturbations in the soil heat diffusivity and eddy diffusivity simultaneously do not add up together. This is due to the non-linear

development of fog.

The assumption of horizontal homogeneity and not well represented parameterizations lead to different effects on the one-dimensional models. The effect are both under- and overestimating of fog formation and dissipation.

C

Brier Skill Score Spatial difference

Relative Brier Skill Score to location 1 to observe spatial variability for the 2hr lead time

Figure C.1: Relative Brier Skill Score of the TAFG for a lead time of 2 hours at location 2, 3 and 4 compared to RVR location 1. The three locations are visible in the legend. The panels represent the three TAFG models. The x-axis show the LVP phases. The y-axis shows the BSS differencr with location 1. A negative difference, indicated a lesser performance compared to the TAFG performance at location 1.
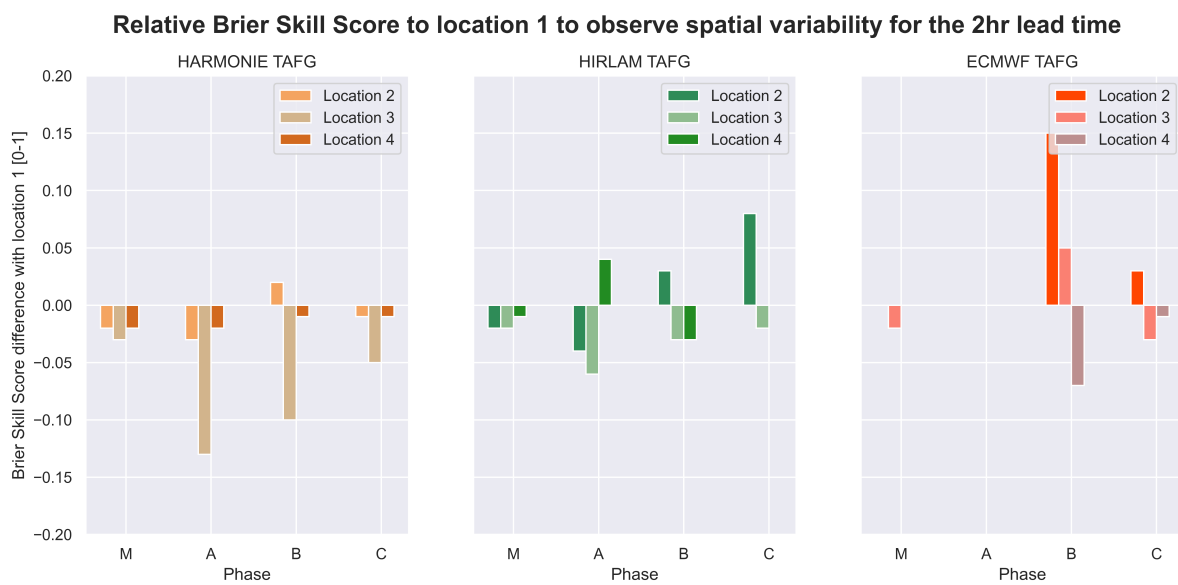
Figure C.2: Relative Brier Skill Score of the TAFG for a lead time of 8 hours at location 2, 3 and 4 compared to RVR location 1. The three locations are visible in the legend. The panels represent the three TAFG models. The x-axis show the LVP phases. The y-axis shows the BSS differencr with location 1. A negative difference, indicated a lesser performance compared to the TAFG performance at location 1.

D

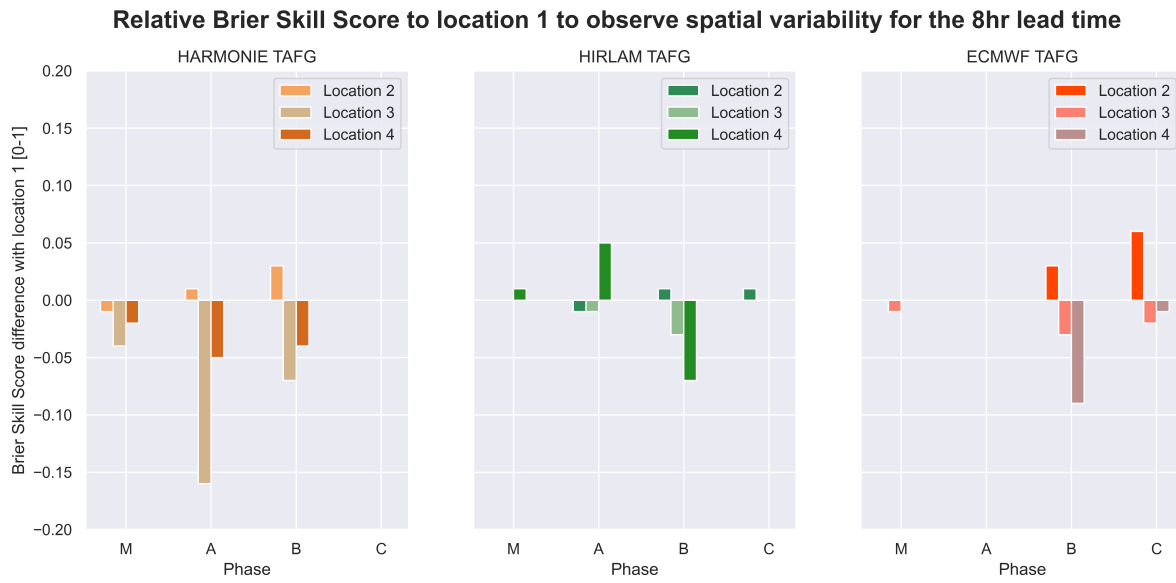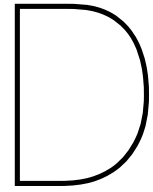# Regression plots

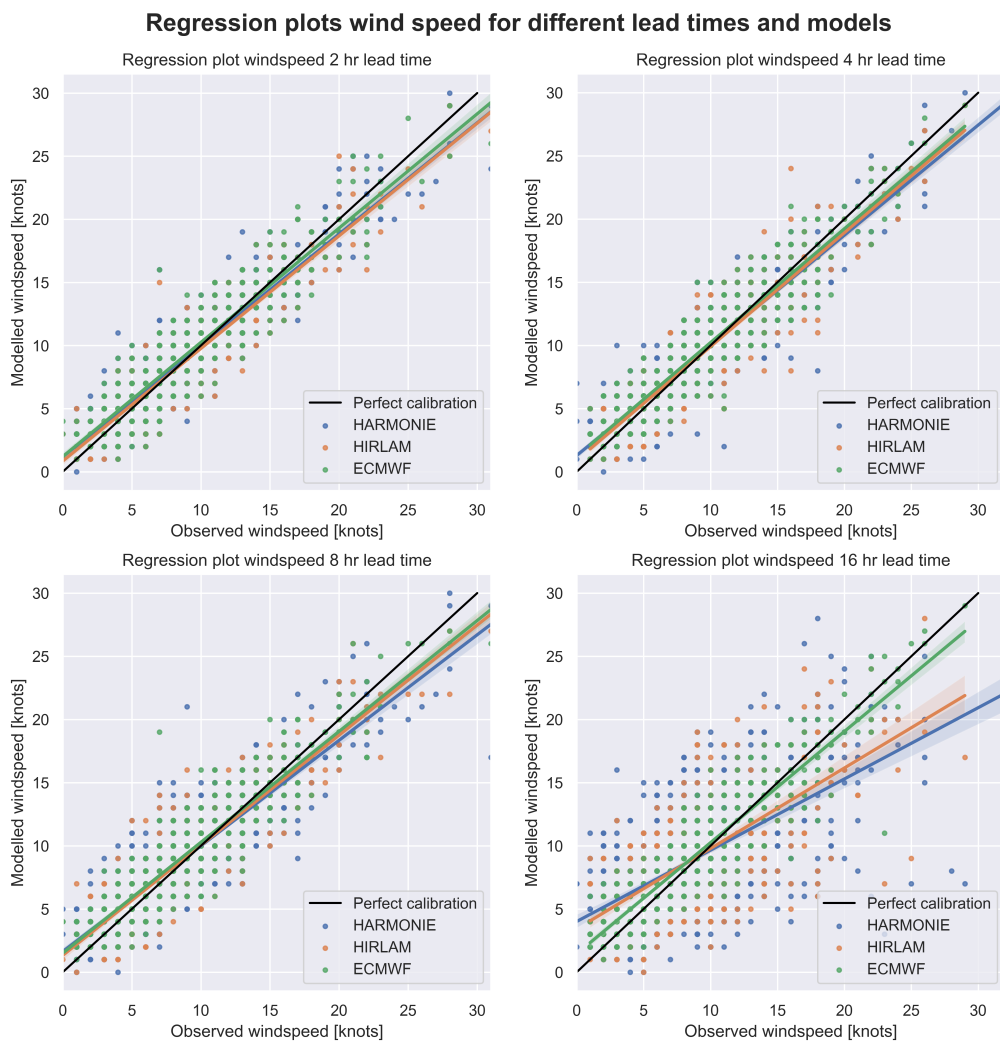**Regression plots wind speed for different lead times and models**

Figure D.1: Regression plot of the observed and modelled temperature at Schiphol for the four lead times and the three TAFG model outputs HARMONIE, HIRLAM and ECMWF.

Bibliography

(2005). Manual of Runway Visual Range Observing and Reporting Practices. Technical report, International Civil Aviation Organization.

(2008). *A Dictionary of Weather*. Oxford University Press, New York, 2nd edition.

Alduchov, O. and Eskridge, R. (1997). Improved Magnus' form approximation of saturation vapor pressure. Technical report, Oak Ridge Operations, Oak Ridge, TN.

Arrillaga, J. A., de Arellano, J. V., Bosveld, F., Baltink, H. K., Yagüe, C., Sastre, M., and Román☐Cascón, C. (2018). Impacts of afternoon and evening sea☐breeze fronts on local turbulence, and on CO_2 and radon☐222 transport. *Quarterly Journal of the Royal Meteorological Society*, 144(713).

Belo-Pereira, M. and Santos, J. A. (2016). A persistent wintertime fog episode at Lisbon airport (Portugal): performance of ECMWF and AROME models. *Meteorological Applications*, 23(3).

Bergot, T., Carrer, D., Noilhan, J., and Bougeault, P. (2005). Improved Site-Specific Numerical Prediction of Fog and Low Clouds: A Feasibility Study. *Weather and Forecasting*, 20(4).

Bergot, T., Escobar, J., and Masson, V. (2015). Effect of small-scale surface heterogeneities and buildings on radiation fog: Large-eddy simulation study at Paris-Charles de Gaulle airport. *Quarterly Journal of the Royal Meteorological Society*, 141(686).

Bergot, T. and Guedalia, D. (1994). Numerical Forecasting of Radiation Fog. Part I: Numerical Model and Sensitivity Tests. *Monthly Weather Review*, 122(6).

Blajev, T. and Krastev, A. (2013). Severe weather risk management survey final. Technical report, Eurocontrol.

Boers, R., Weele, M. v., Meijgaard, E. v., Savenije, M., Siebesma, A. P., Bosveld, F., and Stammes, P. (2015). Observations and projections of visibility and aerosol optical thickness (1956–2100) in the Netherlands: impacts of time-varying aerosol composition and hygroscopicity. *Environmental Research Letters*, 10(1).

Boutle, I. A., Finnenkoetter, A., Lock, A. P., and Wells, H. (2016). The London Model: forecasting fog at 333 m resolution. *Quarterly Journal of the Royal Meteorological Society*, 142(694).

de Rooy, W. (2014). The Fog above Sea Problem in Harmonie: Part 1 Analysis.

de Rover, C., Vogelezang, D., van Bruggen, H., Maat, N., Smit, L., Heijstek, J., Keet, M., Wijngaard, J., and ten Hove, R. (2008). Improved Low visibility and Ceiling Forecasts at Schiphol Airport. Technical report, KNMI.

Duynkerke, P. G. (1991). Radiation Fog: A Comparison of Model Simulation with Detailed Observations. *Monthly Weather Review*, 119(2).

Gultepe, I. and Isaac, G. A. (2007). Cloud fraction parameterization as a function of mean cloud water content and its variance using in-situ observations. *Geophysical Research Letters*, 34(7).

Gultepe, I., Minnis, P., Milbrandt, J., Cober, S. G., Nguyen, L., Flynn, C., and Hansen, B. (2008). The Fog Remote Sensing and Modeling (FRAM) field project: visibility analysis and remote sensing of fog.

Gultepe, I., Sharman, R., Williams, P. D., Zhou, B., Ellrod, G., Minnis, P., Trier, S., Griffin, S., Yum, S. S., Gharabaghi, B., Feltz, W., Temimi, M., Pu, Z., Storer, L. N., Kneringer, P., Weston, M. J., Chuang, H.-y., Thobois, L., Dimri, A. P., Dietz, S. J., França, G. B., Almeida, M. V., and Neto, F. L. A. (2019). A Review of High Impact Weather for Aviation Meteorology. *Pure and Applied Geophysics*, 176(5).

Gultepe, I., Tardif, R., Michaelides, S. C., Cermak, J., Bott, A., Bendix, J., Müller, M. D., Pagowski, M., Hansen, B., Ellrod, G., Jacobs, W., Toth, G., and Cober, S. G. (2007). Fog Research: A Review of Past Achievements and Future Perspectives. *Pure and Applied Geophysics*, 164(6-7).

Haeffelin, M., Bergot, T., Elias, T., Tardif, R., Carrer, D., Chazette, P., Colomb, M., Drobinski, P., Dupont, E., Dupont, J.-C., Gomes, L., Musson-Genon, L., Pietras, C., Plana-Fattori, A., Protat, A., Rangognio, J., Raut, J.-C., Rémy, S., Richard, D., Sciare, J., and Zhang, X. (2010). Parisfog: Shedding new Light on Fog Physical Processes. *Bulletin of the American Meteorological Society*, 91(6).

Herman, G. R. and Schumacher, R. S. (2016). Using Reforecasts to Improve Forecasting of Fog and Visibility for Aviation. *Weather and Forecasting*, 31(2).

Hodges, D. and Pu, Z. (2016). The Climatology, Frequency, and Distribution of Cold Season Fog Events in Northern Utah. *Pure and Applied Geophysics*, 173(9).

Izett, J. G., Wiel, B. J. H., Baas, P., Hooft, J. A., and Schulte, R. B. (2019). Dutch fog: On the observed spatio□temporal variability of fog in the Netherlands. *Quarterly Journal of the Royal Meteorological Society*, 145(723).

Jacobs, A. J. M. and Maat, N. (2005). Numerical Guidance Methods for Decision Support in Aviation Meteorological Forecasting. *Weather and Forecasting*, 20(1).

Marzban, C., Leyton, S., and Colman, B. (2007). Ceiling and Visibility Forecasts via Neural Networks. *Weather and Forecasting*, 22(3).

McFiggans, G., Artaxo, P., Baltensperger, U., Coe, H., Facchini, M. C., Feingold, G., Fuzzi, S., Gysel, M., Laaksonen, A., Lohmann, U., Mentel, T. F., Murphy, D. M., O'Dowd, C. D., Snider, J. R., and Weingartner, E. (2006). The effect of physical and chemical aerosol properties on warm cloud droplet activation. *Atmospheric Chemistry and Physics*, 6(9).

Müller, M. D., Masbou, M., and Bott, A. (2010). Three-dimensional fog forecasting in complex terrain. *Quarterly Journal of the Royal Meteorological Society*, 136(653).

OFCM (2017). Surface weather observations and reports,Federal Meteorological Handbook No. 1. Technical report, National Oceanic and Atmospheric Administration - U.S. Department of Commerce, Washington D.C.

Rémy, S. and Bergot, T. (2009). Assessing the impact of observations on a local numerical fog prediction system. *Quarterly Journal of the Royal Meteorological Society*, 135(642).

Román-Cascón, C., Yagüe, C., Steeneveld, G.-J., Morales, G., Arrillaga, J. A., Sastre, M., and Maqueda, G. (2019). Radiation and cloud-base lowering fog events: Observational analysis and evaluation of WRF and HARMONIE. *Atmospheric Research*, 229.

Román-Cascón, C., Yagüe, C., Steeneveld, G.-J., Sastre, M., Arrillaga, J. A., and Maqueda, G. (2016). Estimating fog-top height through near-surface micrometeorological measurements. *Atmospheric Research*, 170.

Roquelaure, S. and Bergot, T. (2007). Seasonal Sensitivity on COBEL-ISBA Local Forecast System for Fog and Low Clouds. *Pure and Applied Geophysics*, 164(6-7).

Roquelaure, S., Tardif, R., Remy, S., and Bergot, T. (2009). Skill of a ceiling and visibility Local ensemble Prediction System (LEPS) according to fog-type prediction at Paris-Charles de Gaulle Airport. *Weather and Forecasting*, 24(6):1511–1523.

Steeneveld, G. J. and de Bode, M. (2018). Unravelling the relative roles of physical processes in modelling the life cycle of a warm radiation fog. *Quarterly Journal of the Royal Meteorological Society*, 144(714):1539–1554.

Steeneveld, G. J., Koopmans, S., Heusinkveld, B. G., van Hove, L. W. A., and Holtslag, A. A. M. (2011). Quantifying urban heat island effects and human comfort for cities of variable size and urban morphology in the Netherlands. *Journal of Geophysical Research*, 116(D20).

Szintai, B., Szűcs, M., Randriamampianina, R., and Kullmann, L. (2015). Application of the AROME non-hydrostatic model at the Hungarian Meteorological Service. *Quarterly Journal of the Hungarian Meteorological Service*, 119(2):241–265.

Tardif, R. and Rasmussen, R. M. (2007). Event-based climatology and typology of fog in the New York City region. *Journal of Applied Meteorology and Climatology*, 46(8):1141–1168.

Van Der Velde, I. R., Steeneveld, G. J., Wichers Schreur, B. G., and Holtslag, A. A. (2010). Modeling and forecasting the onset and duration of severe radiation fog under frost conditions. *Monthly Weather Review*, 138(11):4237–4253.

van Dinther, D., Hartogensis, O. K., and Holtslag, A. A. (2015). Runway Wake Vortex, Crosswind, and Visibility Detection with a Scintillometer at Schiphol Airport. *Boundary-Layer Meteorology*, 157(3):481–499.

Vautard, R., Yiou, P., and van Oldenborgh, G. J. (2009). Decline of fog, mist and haze in Europe over the past 30 years. *Nature Geoscience*, 2(2).

Westerhuis, S., Fuhrer, O., Cermak, J., and Eugster, W. (2020). Identifying the key challenges for fog and low stratus forecasting in complex terrain. *Quarterly Journal of the Royal Meteorological Society*, 146(732).