ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

Comparison of two physical- and statistical-based post-processing methods for high-resolution NWP visibility forecasts over 13 North European airports

Nikkilä, M., Ylhäisi, J.S., Neiglick, S., Koivunen, T.
Finnish Meteorological Institute
jussi.ylhaisi@fmi.fi

Short-term visibility forecasting for Terminal Aerodrome Forecasts (TAFs) is one of the remaining strongholds where human forecasters are still able to use their expertise to outperform direct model output (DMO) or many of the solely model-based post-processing products. In this study, we apply two distinct post-processing methods for 24 hour TAF visibility forecasts over the Northern European domain covering 13 airports and compare their forecasts with the Direct Model Output (DMO) of a high-resolution HARMONIE-AROME model.

- The first post-processing method is a physical-based diagnostic method, which has been developed and empirically refined by the duty forecasters at the Finnish Meteorological Institute over the course of several years. The method takes into account the worsening of the visibility from the default value, caused either by precipitation of various forms or by low stratus clouds.
- The second method is a statistical-based analogue method, where the prevailing weather conditions from the recent past 6-hour METAR observations are compared with the longer observational time series of several years in order to distinguish the most similar past situation to the current one. After identifying the most similar past situation, that situation is correspondingly used to forecast the present situation.

Our verification of visibility forecasts is mostly based on the ICAO visibility classification defined in Annex 3 (2016). The forecasts are verified from Sep2017 to Dec2017. Our results show that the post-processing has mixed effects for the forecasts: Excluding the visibility class 0...150m, the performance of the diagnostic post-processing is better for the low visibility classes as compared with DMO, but the differences might not be statistically significant. Unfortunately, our analogue forecasts do not perform as hoped. The performance is obviously lower as compared to DMO, in terms of both error variance and bias. In addition, the analogue forecasts do not show a similar degradation as a function of forecast length and the additional benefit of having access to observational information is not anyhow evident from the forecasts: Even the 3-hour forecasts perform very poorly as compared with DMO. Analogue method could be extensively refined though using different weighting schemes, but unfortunately this kind of optimization was computationally too demanding due to long time series used in the calculation of analogies. Physical-based post-processing method clearly shows more promise to it, which has also been observed in operational setting. There remains plenty of room for the forecaster to make subjective evaluation of the available products.

Figure 1. Airports analysed in the study

Methodology

Physical-based method

The development of the physical-based diagnostic method is mainly inspired by the characteristics of the HIRLAM model (Bengtsson et al., 2017) and the accumulated experience of the aviation forecasters in using it. HIRLAM model is still widespreadly used in operative aviation weather forecasting at FMI, even though its successor HARMONIE-AROME has largely replaced it. However, the typical model-simulated visibility characteristics can be seen in HARMONIE-AROME as well: Mean visibility bias is positive throughout the axis, pronouncedly for the poor visibility classes (see Figure 2).

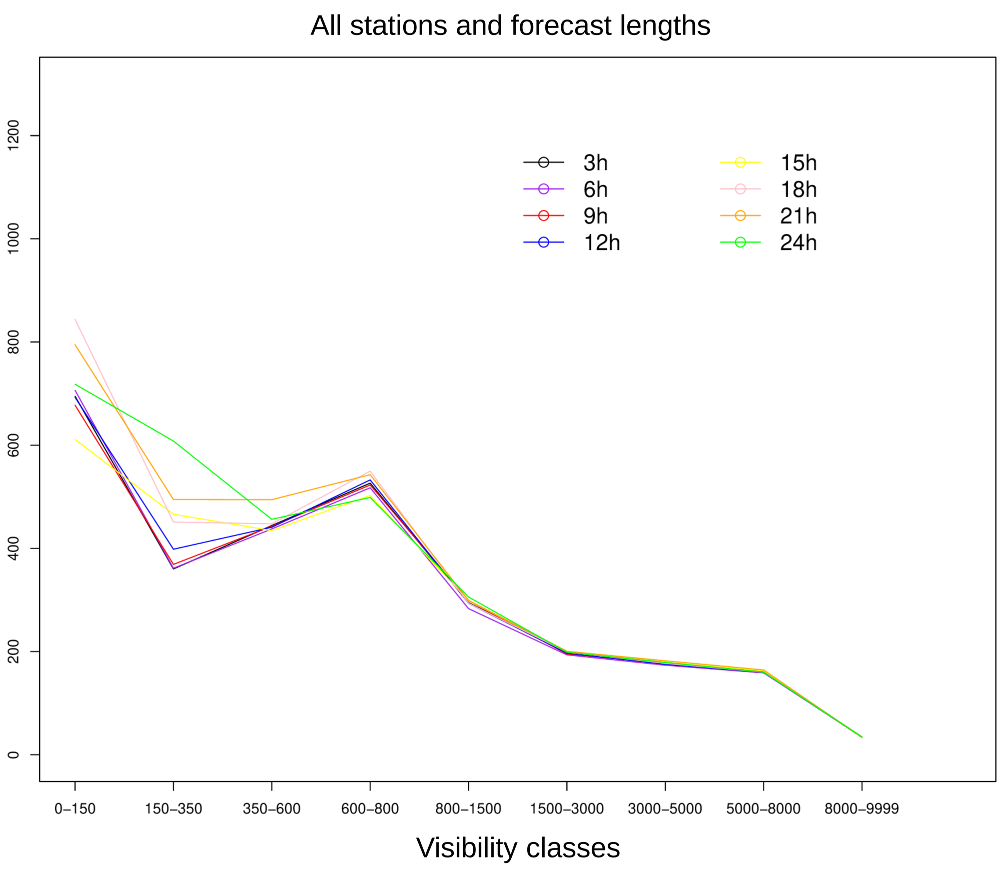

Figure 2. Mean error of HARMONIE-AROME visibility forecasts over the whole domain, for individual forecast hours.

The method first calculates visibility in rain, using form-dependent empirical formulas which only use precipitation intensity as an input. Then, another empirical visibility formula is applied for model-simulated relative humidity (when RH>80%). In order to being able to better capture the worst weather situations, visibility forecasts are made worse at poor conditions. Precipitation visibility is further worsened when (RH_surf>85%, N_max(<1000ft)>50%, CLB<500ft) and RH visibility when (N_max(<1000ft)>65%, CLB<1000ft). Finally, the minimum from these two visibility values is taken.

The physical-based postprocessing scheme can be applied to DMO of any model that provides the output of the required parameters (precipitation form, intensity, humidity, cloud coverage) for the algorithm and this is also being done at FMI for several models. The method is essentially a diagnostic visibility scheme, which does not require model-simulated visibility as an input.

One of the main motivations for the development of physical visibility scheme has been the observation that models cannot properly simulate local phenomenas which are observerd in reality. This is due to model biases, representation error due to model resolution or due to physical model deficiencies. An open-source C++/Lua –repository for the visibility parameterization and ~70 other diagnostic model variables is available at https://github.com/fmidev/himan/blob/master/himan-plugins/source/visibility.cpp

Analogue method

The analogue method does not need any model forecasts as it is a purely observation-based product, but the observational time series preferably needs to be as long as possible. The analogue method gives each past situation a scaled similarity index, based on fuzzy logic. In addition to assigning a degree of similarity to each observation, the method applies varying weighting for data points, depending on the METAR variable and the time difference as compared to present time (the most recent observation is given the highest weight). Finally, the most similar weather situation from the past is used for forecasting. The used method is comprehensively described in Tuba and Bottyán, 2017.

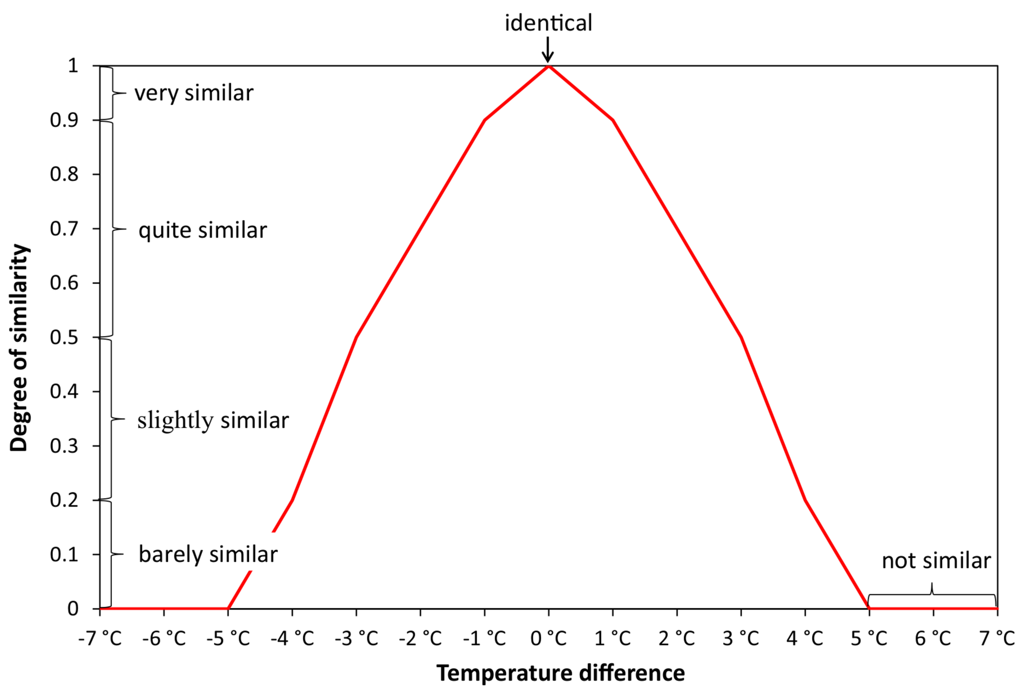

Figure 3. Illustration of fuzzy logic principle (adopted from Tuba and Bottyán, 2017)

variable	weight
Visibility	0.321
Hour of the day	0.107
Day of the year	0.071
Surface pressure	0.071
Temperature	0.107
Dew point	0.107
Wind speed	0.143
Wind direction	0.071
Sum	1.00

Table 1. Subjective weights of the METAR variables used in analogue method

Data used

The 13 airports used are shown in Figure 1. From these stations, METAR observations were used as the verification baseline. Most of the observations are done by automatic weather stations, but over some stations manual observations are still done to a varying degree. A composite of manual and automatic observations was used. The METAR used to verify forecasts is done 20 past and is supposed to represent weather during the previous half hour. The fraction of the available observation time series for the stations during the time period 2000-2017 varies from 25% to 90%, Finnish stations having a better extent. Sufficient length of the observation time series is crucial for analogy forecasts, which search for the most similar weather situation in the station observation history.

The verification period for the forecasts was the period Sep2017-Dec2017. This was due to calculation of the analogue forecasts being computationally intensive.

In the analysis, forecasts from the forecast hours 3,6,9,12,15,18,21,24 are grouped together in order to increase sample size, as no significant obvious among them were seen.

The model data used was the control member of the 2.5km HARMONIE-AROME –based MEPS ensemble from September 2017 onwards, whereas the model data prior to this was based on an earlier version from the same model. However, the visibility scheme in the model is similar between these two model versions.

Results

The verification results in Tables 2-4 are calculated for the Sep2017-Dec2017 period, based on dichotomous verification scores and using ICAO Annex 3 visibility classes (values in meters). The verification scores are Probability of detection (POD), False alarm rate (FAR) and Peirce Skill score (PSS).

From the results it can be seen that both of the post-processed forecasts are improving over the DMO output, but the physical method performs much more consistently over the whole visibility distribution. The physical method is not really able to produce values <150m. Even though these events only constitute a minor fraction from the observations, they have a significant impact. The verification scores of the physical algorithm are better compared to DMO for almost all classes and scores. If the definition used in POD values are extended to exact or either of the nearest two classes, improvement of physical algorithm is further reinforced (not shown). Also, it has to be noted that the physical-based algorithm has so far only been verified at the forecasters' desk by using eye-ball method! The results could likely be refined considerably by tuning the constant values in it.

CLASS	DMO	PHYSICAL	ANALOGY
0-150	0.22	0	0.09
150-350	0.11	0.17	0.19
350-600	0.05	0.15	0.08
600-800	0.03	0.06	0.04
800-1500	0.18	0.22	0.08
1500-3000	0.39	0.29	0.14
3000-5000	0.16	0.16	0.16
5000-8000	0.14	0.14	0.15
8000-9999	0.53	0.79	0.84

Table 2. POD values for different producers, Sep2017-Dec2017

CLASS	DMO	PHYSICAL	ANALOGY
0-150	0.97	-	0.92
150-350	0.92	0.89	0.82
350-600	0.95	0.91	0.91
600-800	0.97	0.96	0.96
800-1500	0.92	0.90	0.91
1500-3000	0.92	0.87	0.88
3000-5000	0.95	0.89	0.88
5000-8000	0.94	0.89	0.88
8000-9999	0.06	0.08	0.13

Table 3. FAR values for different producers, Sep2017-Dec2017

CLASS	DMO	PHYSICAL	ANALOGY
0-150	0.20	0	0.09
150-350	0.10	0.15	0.18
350-600	0.04	0.13	0.07
600-800	0.03	0.05	0.03
800-1500	0.15	0.19	0.07
1500-3000	0.22	0.22	0.10
3000-5000	-0.02	0.09	0.07
5000-8000	-0.07	0.05	0.06
8000-9999	0.38	0.49	0.33

Table 4. PSS values for different producers, Sep2017-Dec2017

It has to be noted, however, that the FAR values are still very high and POD values relatively low for all producers. High FAR values indicate that most forecast cases are not observed in reality. This is affirmed by the total number of poor visibility events in observations and simulations, as producers (DMO/physical) are heavily overforecasting the number of poor visibility events: The fraction of total number of events are 180% and 176%, respectively. For comparison, this fraction is 95% for the analogy forecasts (not shown). This gross overforecasting in part indicates that verification scores based on continuous values are not very encouraging: The forecast error standard deviation is >2000meters for all producers, all forecast lengths and all visibility classes.

Statistical approaches are undoubtely better at downscaling grid-box scale model values to local station points, but they are very much dependent on the long observation time series from the station points. In this study, observation time series are likely too short for the analogy method to work properly. For a few stations, the observation time series only comprises less than 5 years which is not enough so that enough events of poor visibility would have time to accumulate. The method could further be refined through better tuning of the METAR variable weights, past time step weights or the fuzzy logic functions. Also, model data could be used in parallel with observations in order to form hybrid forecasts.

Further errors in the results are caused by the short verification time period, heterogeneous observation practices used and the interpretation differences between model and observational variables (model values are from 10m whereas observations are from 2m). Our results confirm that forecasting visibility is still a very challenging problem both for physical and statistical-based postprocessing schemes.

References

Annex 3, 19th edition, 2016: Meteorological Service for International Air Navigation. International Civil Aviation Organization, Chapter 4: Meteorological observations and reports & Chapter 6: Forecasts & Appendix 5: Technical specifcations related to forecasts.

Bengtsson, L., U. Andrae, T. Aspelien, Y. Batrak, J. Calvo, W. De Rooy, E. Gleeson, B. Hansen-Sass, M. Homleid, M. Hortal, K-I. Ivarsson, G. Lenderink, S. Niemela, K. Pagh Nielsen, J. Onvlee, L. Rontu, P. Samuelsson, D. Santos Munoz, A. Subias, S. Tijm, V. Toll, X. Yang, M. Odegaard Koltzow, 2017: The HARMONIE-AROME Model Configuration in the ALADIN-HIRLAM NWP System. American Meteorological Society, 145, s. 1919-1935.

HIMAN: weather parameter post processing tool. https://github.com/fmidev/himan/

Nikkilä, M., 2018: Näkyvyysennusteiden tuottaminen erilaisin menetelmin ja niiden keskinäisen hyvyyden vertaaminen lentosäässä. Masters' thesis (in Finnish).

Tuba, Z., Bottyán, Z., 2017: Fuzzy logic-based analogue forecasting and hybrid modelling of horizontal visibility. Meteorol. Atmos Phys. DOI 10.1007/s00703-017-0513-1