VISIBILITY AND FOG FORECASTING BASED ON DECISION TREE METHOD.

Ferenc Wantuch

Hungarian Meteorological Service,

Kitaibel Pál u. 1 H-1525 Budapest, Hungary; E-mail: wantuch.f@met.hu

Abstract-The paper describes a visibility and fog forecasting model developed and used at the Hungarian Meteorological Service (HMS) for last 3 years. The investigated model is a perfect prognostical model (PP). Characteristics of the model, such as input data, statistical approach, decision trees and threshold numbers, are described in this paper. The model was tested for both measured sounding and predicted data. The verification of the model led to very good results, so it was applied to aeronautical forecasting as well as to nowcasting. Information and short review about different type of other visibility models are also given.

Key-words: NWP parameters, perfect prognosis(PP), model output statistics (MOS), FOGSI-index, decision tree.

1. Introduction

Visibility forecast is very important for transportation, especially for air traffic where its accuracy is prominent. The WMO/ICAO requirements are very rigorous in aviation meteorology (ICAO-1998). Verifications regarding to the aeronautical forecasts show, that the reason of poor Terminal Weather Forecasts – in about 70 per cent of all cases - are the weak or not suitable visibility predictions. At HMS earlier it was not available any special numerical method, which could be an aid for forecasters in the prediction of the visibility, so they could use only traditional tools.

The European forecasters use different methods in the practice. One possibility is the diagnosis of fog from satellite images (*Kerenyi et al.,* 1995). Some organisations e.g. EUMETSAT, Central Institute for Meteorology and Geodynamics Austria (ZAMG), Swedish Meteorological Institute (SMHI), and Météo France use NOAA AVHRR, and Meteosat images in order to analyse fog and low cloud from satellite data. Another possibility is the improvement of 1-D-models applied in UK, Sweden, Portugal, Belgium (*Stessel* and *Ottoy,* 1999) and also in France. Some case studies have been validated with promising results.

The third way is the use of statistical methods and decision support systems for fog and low cloud forecasting. In the frame of it different methods, like decision trees, linear regression, Kalman-filter (*Kilpinen* and *Juha,* 1992) and neural network (*Pasini at al.,* 1999) were considered for probability forecasts. In general, the results of all these methods were promising, so we considered the problem from statistical point of view.

Let us denote by y the estimated parameter, that is the predictand and let $x_{1,}, x_2, ...x_p$ detected meteorological elements which are the predictors. In this case we have to construct a function:

$$y = f(x_{1,}, x_2, ...x_p) + \varepsilon \tag{1}$$

where ε is the error of the method. One can use this function in the following estimated form:

$$\tilde{y} = f(\tilde{x}_{1,}, \tilde{x}_2, ...\tilde{x}_p) \tag{2}$$

where $\tilde{x}_{1,}, \tilde{x}_2, ...\tilde{x}_p$ are known from NWP. This is the basic concept of the perfect prognosis method. Suppose that (2) is constructed directly from $\tilde{x}_{1,}, \tilde{x}_2, ...\tilde{x}_p$.

$$y = f(\tilde{x}_{1,}, \tilde{x}_2, ...\tilde{x}_p) + \varepsilon \tag{3}$$

In this way we get a model output statistical method. Based on this idea an automatic visibility forecast method can be constructed. The input data of the visibility prediction is in the given case the ALADIN mesoscale model output. This is a hydrostatic, spectral limited area numerical weather prediction which was developed by collaboration between Météo-France and some Central-and Eastern-European hydrometeorological services including HMS. The main dynamical characteristic of the model, like the preparation of initial, lateral boundary conditions, the physics and post-processing was discussed by *Horányi at al.* (1996).

In order to find a connection with visibility at first we made a comprehensive statistical research of direct measurements and derived physical quantities. The best correlation was received by the fog stability index. The index was calculated according to the following formula:

$$\text{FOGSI} = 2 \mid T_{sfc} - T_{850} \mid + 2\ (T_{sfc} - T_{d\,sfc}) + 2\ W_{850} \tag{4}$$

where

T_{sfc}	temperature near the surface,
$T_{d\,sfc}$	dew point near the surface,
T_{850}	temperature on 850 hPa level,
W_{850}	wind speed on 850 hPa level.

FOGSI index takes into account the temperature gradient, (that is the measure of the stability), the impact of moisture near the surface and the mixing by wind.

2.The result of statistical research

The FOGSI index is highly correlated with the observed visibility especially in autumn-winter time when fog and mist frequently occurs. Because of the strong relationship, we could use a regression connection based on two years long dataset as follows:

$$Visibility = -1.33 + 0.45 \bullet FOGSI \tag{5}$$

Fig. 1. shows connection between the FOGSI index and the observed visibility (measured in kms) in October 1996. Based on this figure we can make the following considerations. There is a critical interval of FOGSI, above the upper limit which the calculation of visibility by regression is adequate to use. On the other hand if the FOGSI number is smaller than the lower limit of this domain, one can predict fog in all cases. The variance of the visibility values inside the critical interval is very high, consequently the statistical method is uncertain. It means that in this interval one can not decide about the visibility based on FOGSI, e.g. if FOGSI is equal to 25 it might represent whether fog, mist or good visibility in the same time. If data are taken into account only from the critical interval, then a statistical connection between FOGSI and visibility becomes very poor (*Fig. 2.*). Its physical reason is, that several other effects, which play a great role in development of visibility, were neglected in FOGSI definition. Therefore one has to introduce some new weather predictors and methods. Such kind of parameters is reasonable to select, which can be computed from TEMP data and NWP model output as well. After thorough investigation the mean relative humidity of lower air layers (925 hPa -surface) and upper layers (850-700 hPa), the near surface wind speed and relative humidity were chosen to be included into the decision process.

Having examined a large number of cases, it was found that in winter period the cold air can be accumulated near the surface, mainly in the valleys and basins. Sometimes the surface temperature is colder 2-5 degrees with, in extreme cases this difference reaches 10 ^{0}C as compared with teh temperature of the 850 hPa level. This inversion stratification called "the cold air pad" (*Tóth,*1984; *Bóna,* 1986), which represents a very stable state of the atmosphere. We came to the conclusion, that it is necessary to treat cold air pad situations separately and for these days other threshold numbers have to be determined. In order to specify different visibility categories inside the critical interval, one had to work out a new procedure. The steps of the process are summarised in *Fig. 3*.

3. Decision tree

In this chapter the principles and steps of the decision-making procedure will be discussed and demonstrated in *Fig. 4*. The main characteristic of this tree is that each physical condition of the air column represents only one category. The threshold numbers based on two years of surface and radiosounding measurements of Budapest-Lőrinc.

Suppose that the lowest layer of the air (between surface and 925 hPa) is dry or moderately wet and windy, then the visibility depends on the water content near the surface. In this case the process uses simply the regression line for determining the visibility. If the lowest layer is dry and the wind is weak, then the visibility depends on the water content of the air near the surface. If this layer is wet, then we choose mist otherwise a good visibility category is selected. If the lowest layer is medium wet, then four subclasses are constructed. In these subclasses the radiative cooling effects of the atmosphere above a ground level point are represented. This influence was modelled as the difference of mean relative humidity between the upper and the lower layers as follows:

1) if the air near the surface is dry and we do not include radiative cooling effect near the surface, then can be calculated the visibility by the help of the regression line.
2.) if the air near the surface is wet and we have not radiative cooling influence in the air column, then the decision is the misty weather.
3.) if the air near the ground level is wet and we have radiative cooling effect, then is fog formation we expected.
4.) if the air near the surface is medium wet and we include radiative cooling, in this case is misty weather predicted.
5.) Finally, supposing that the air near the ground level is very wet, then in case of radiative cooling we expect foggy otherwise misty category.

For cold air pad situations similar decision tree was constructed. The main differences are in the values of the threshold numbers. If very high relative humidity and weak winds occur together, it will be foggy weather.

4.Test results

The test of diagnostic method presented in this paper led to the following results. *Fig. 5* illustrates the visibility at 00 UTC for each day of January in 1997, where JANAKTL means the measured, JANMODSZ the computed values. With regard to reliable visibility forecast good estimation of the small values is especially important. As it is shown, under 5 km both lines give similar range of sight, even if the dotted line sometimes a little bit underestimates the real data, so it makes the prediction e.g. for. aviation more safety. For larger visibility values the difference is not so important.

The next two figures illustrate the correlations between measured and computed visibilities with the use of decision tree procedure (JANMODSZ) (*Fig. 6.*) and without (JANSSI) (*Fig. 7.).* High correlation (0.83) was reached with the more developed method, as opposed to the low correlation received 0.40 apply only the simple FOGSI index (4). According to our experiences the described decision tree procedure improves the results in all cases.

An even more illustrative picture is presented in *Fig. 8.*, where one can follow the hour by hour (continuos line) changes of real visibility compared to the 48 hour forecasts (columns). The run of observed and predicted visibility values is comparable to each other, although some hours long shift might be detected. Regression coefficients were calculated based on the radiosounding data of Budapest and used for Szeged. It can be concluded, that equation (5) adequate for most of the places of Hungary. A possible explanation is, that it is due to the relatively smooth surface of the country.

5. Conclusion

The described method is mainly used in aeronautical meteorology. After the test period this method was installed at HMS Weather Forecasting and Aviation Meteorology Department. According to 3 years long experience, efficiency of the method strongly depends on the quality of the ALADIN model output near the ground level.

Another application area is nowcasting, where the application of the above outlined procedure for fog formation and dissipation, as well as the horizontal visibility are led to significant improvements.

References

Horányi,A.., *Ihász,* I. and *Radnóti,* G., 1996: ARPEGE/ALADIN: A numerical weather prediction model for Central-Europe with the participation of the Hungarian Meteorological Service. *Időjárás* 100, 277-301.

ANNEX 3, International Standards and Recommended Practices, Meteorological Service for International Air Navigation To the Convention on International Civil Aviation, Thirteenth edition – July 1998 (ICAO).

Bóna, M., 1986: Hideg-légpárnák aeroszinoptikai vizsgálata. Meteorológiai Tanulmányok. 54. szám .

Stessel, J.-P. and *Ottoy, H.,* 1999: Dense Fog Forecasting with an Interactive Expert. *Cost-78 Project II.3*, *final report*, Fog and Low Clouds: Statistical Methods and Decision Support Systems for Fog and Low Cloud Forecasting.

Kerényi, J.,.G. Szenyán, I., Putsay, M.. and.*Wantuch, F.,* 1995.: Cloud detection on treshold technique for NOAA-AVHRR images for the Carpathian Basin, *Proceedings of 1995 Meteorological Satellite Data Users` conference*, Winchester, United Kingdom, 4-8 Sept 1995, pp: 565-569.

Kilpinen and *Juha*, 1992. The Application of Kalman Filter in Statistical Interpretation of Numerical Weather Forecasts. *12th Conference on Probability and Statistics in the Atmospheric Sciences,* June 22-26, 1992, Toronto, Ont, Canada, 11-16.

Pasini, A., Pelino, V. and *Potestà, S.,*1999. A Neural Network model for visibility nowcasting from surface observations: results and sensitivity to physical input variables, submitted to *Journal of Geophysical Research, D.*

Tóth, P., 1984.: Parametrizáció bevezetése a hideg légpárnák keletkezése és feloszlásának analizise céljából. *Meteorológiai Tanulmányok*. 51. szám

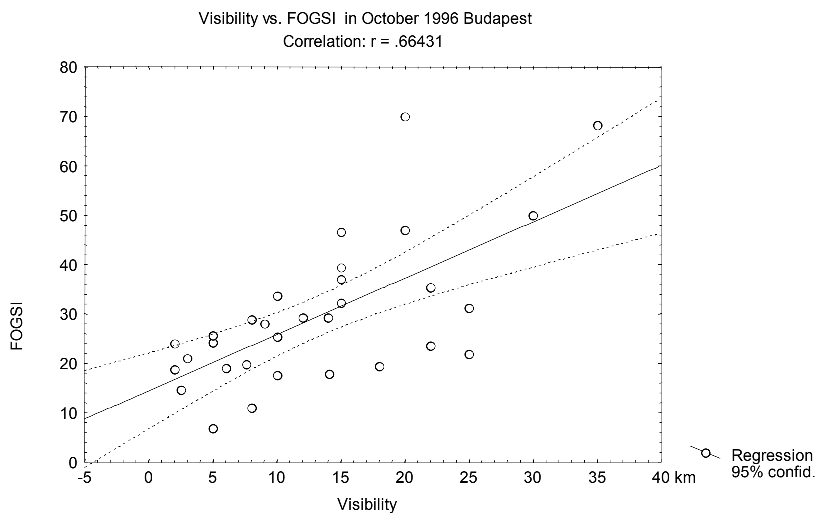
Visibility vs. FOGSI in October 1996 Budapest
Correlation: r = .66431
80
70
60
50
40
30
20
10
0
FOGSI
-5
0
5
10
15
20
25
30
35
40 km
Visibility
Regression
95% confid.

Fig. 1.

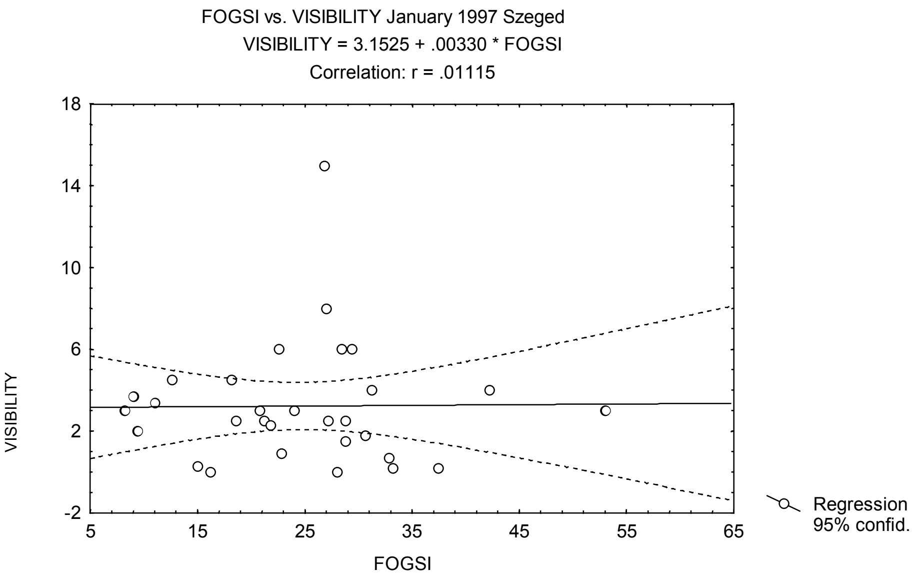
FOGSI vs. VISIBILITY January 1997 Szeged
VISIBILITY = 3.1525 + .00330 * FOGSI
Correlation: r = .01115
VISIBILITY
FOGSI
18
14
10
6
2
-2
5
15
25
35
45
55
65
Regression
95% confid.

Fig. 2.

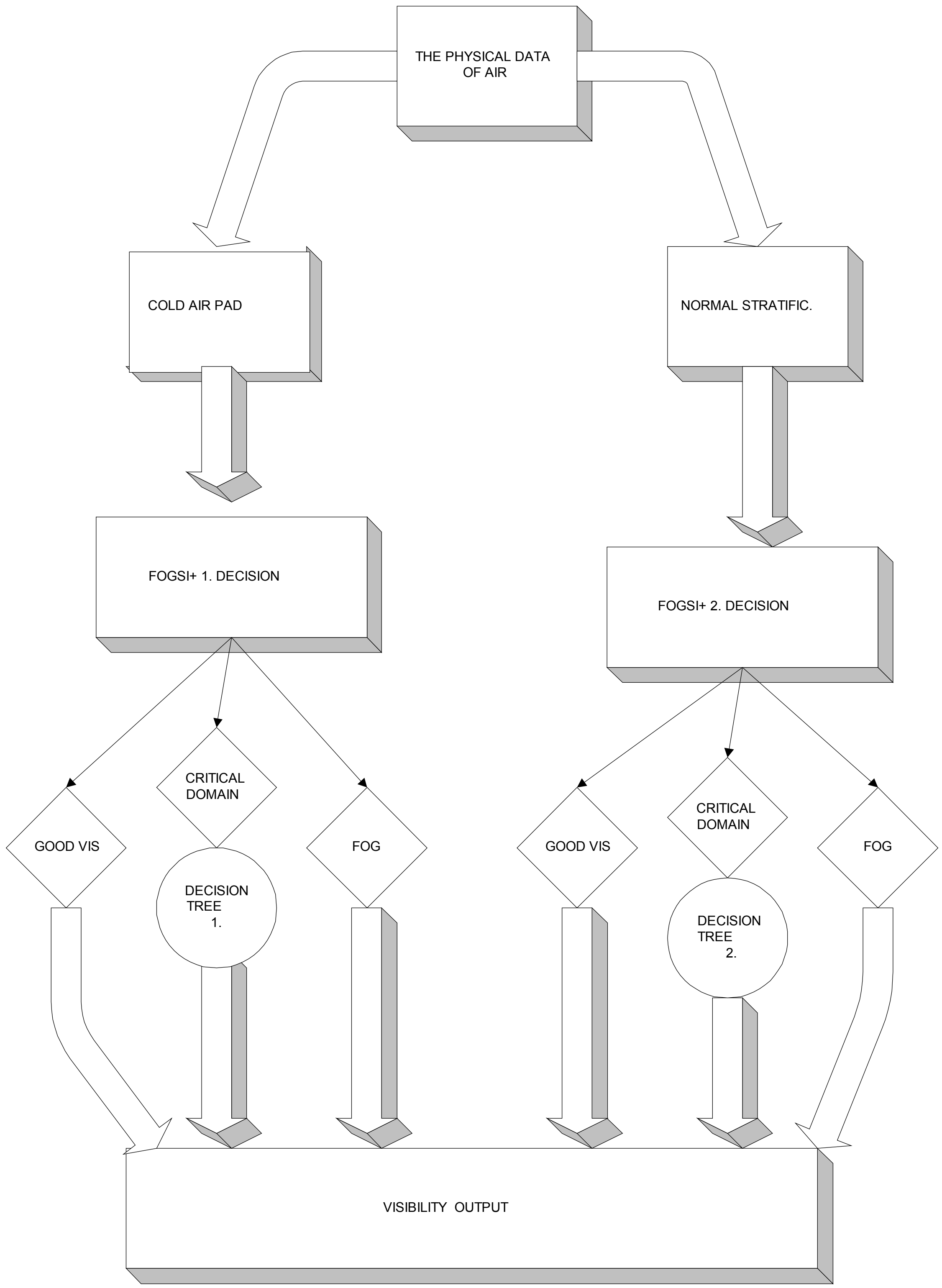
THE PHYSICAL DATA OF AIR
COLD AIR PAD
NORMAL STRATIFIC.
FOGSI+ 1. DECISION
FOGSI+ 2. DECISION
GOOD VIS
CRITICAL DOMAIN
FOG
DECISION TREE 1.
GOOD VIS
CRITICAL DOMAIN
FOG
DECISION TREE 2.
VISIBILITY OUTPUT

Fig. 3.

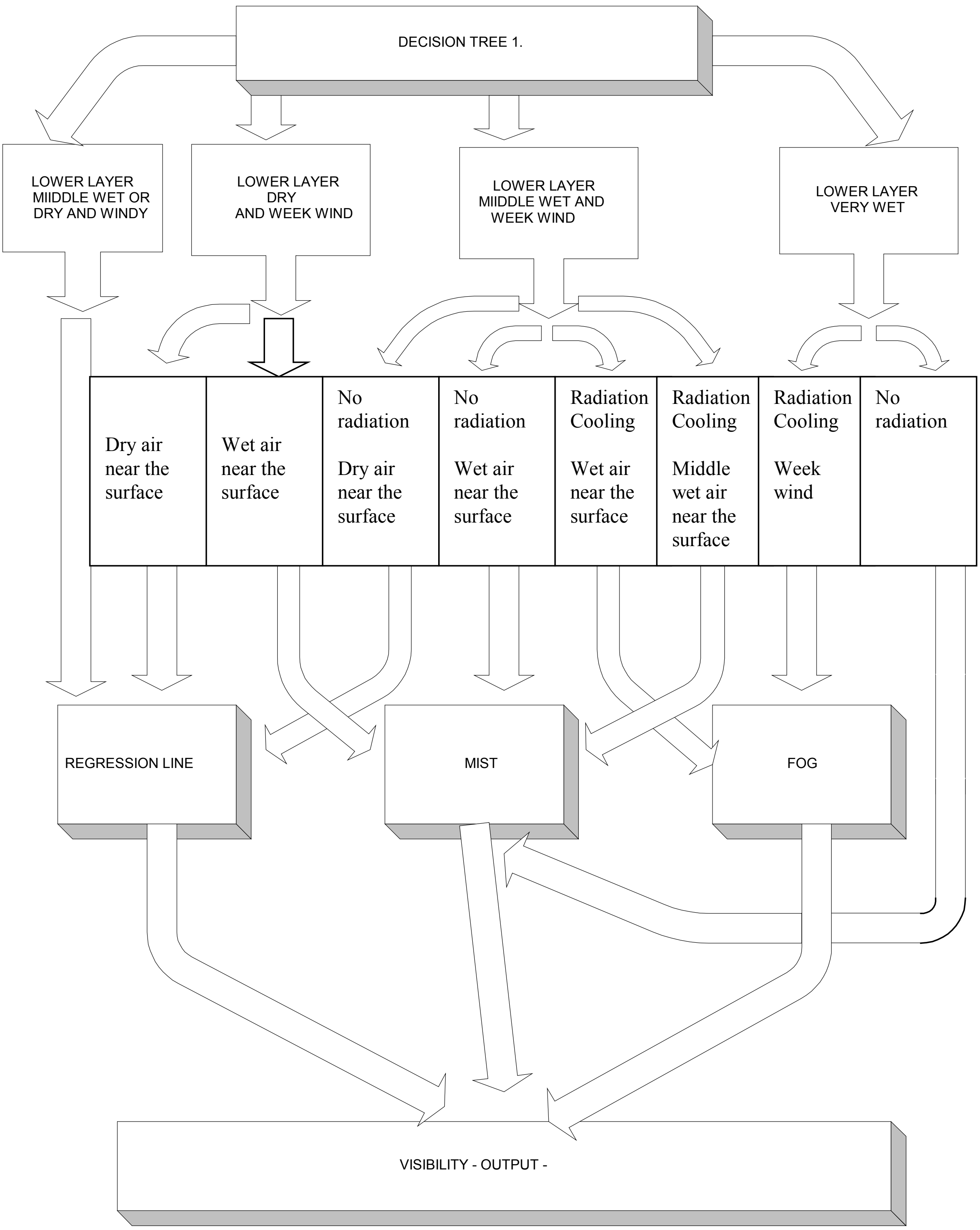
DECISION TREE 1.
LOWER LAYER MIIDDLE WET OR DRY AND WINDY
LOWER LAYER DRY AND WEEK WIND
LOWER LAYER MIIDDLE WET AND WEEK WIND
LOWER LAYER VERY WET
Dry air near the surface
Wet air near the surface
No radiation Dry air near the surface
No radiation Wet air near the surface
Radiation Cooling Wet air near the surface
Radiation Cooling Middle wet air near the surface
Radiation Cooling Week wind
No radiation
REGRESSION LINE
MIST
FOG
VISIBILITY - OUTPUT -

Fig. 4

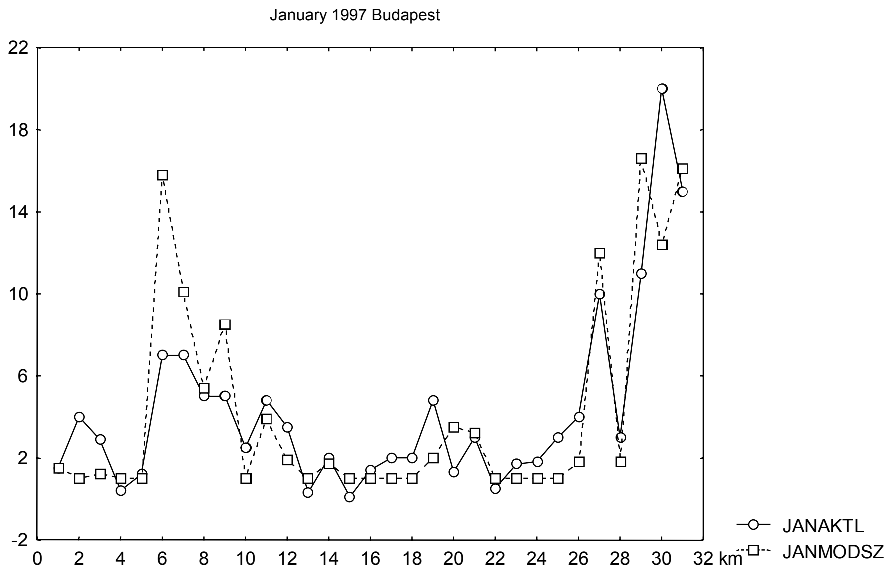
January 1997 Budapest
22
18
14
10
6
2
-2
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32 km
JANAKTL
JANMODSZ

Fig. 5.

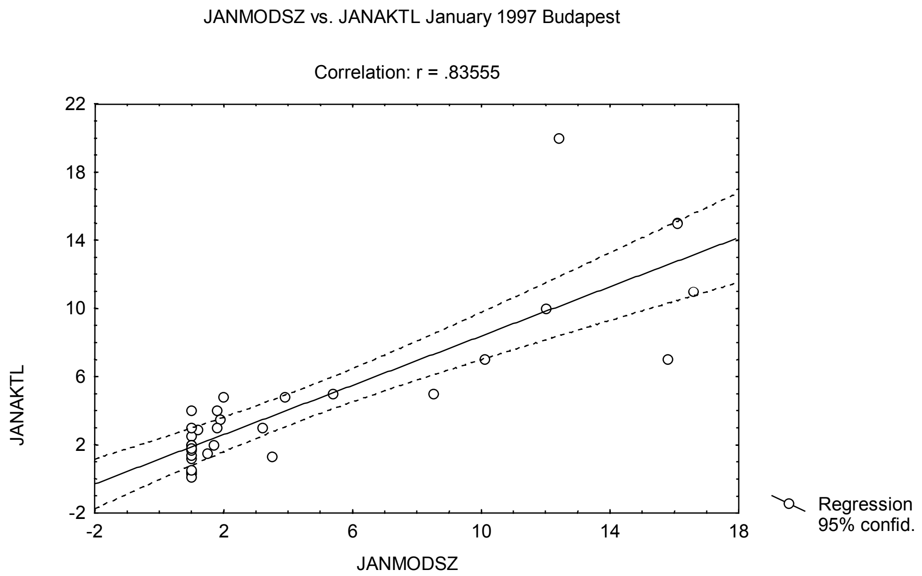
JANMODSZ vs. JANAKTL January 1997 Budapest
Correlation: r = .83555
22
18
14
10
6
2
-2
-2
2
6
10
14
18
JANAKTL
JANMODSZ
Regression
95% confid.

Fig. 6.

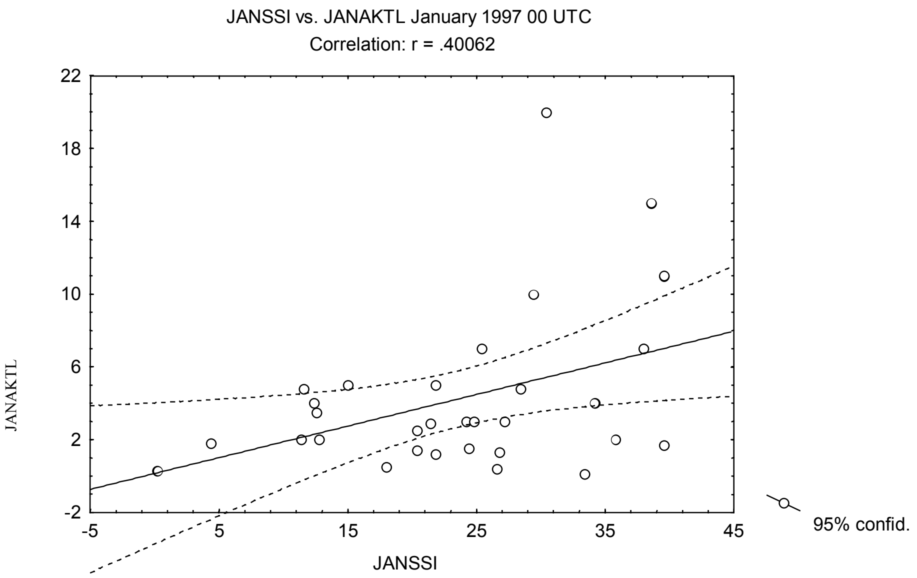
JANSSI vs. JANAKTL January 1997 00 UTC
Correlation: r = .40062
22
18
14
10
6
2
-2
-5
5
15
25
35
45
JANAKTL
JANSSI
95% confid.

Fig. 7.

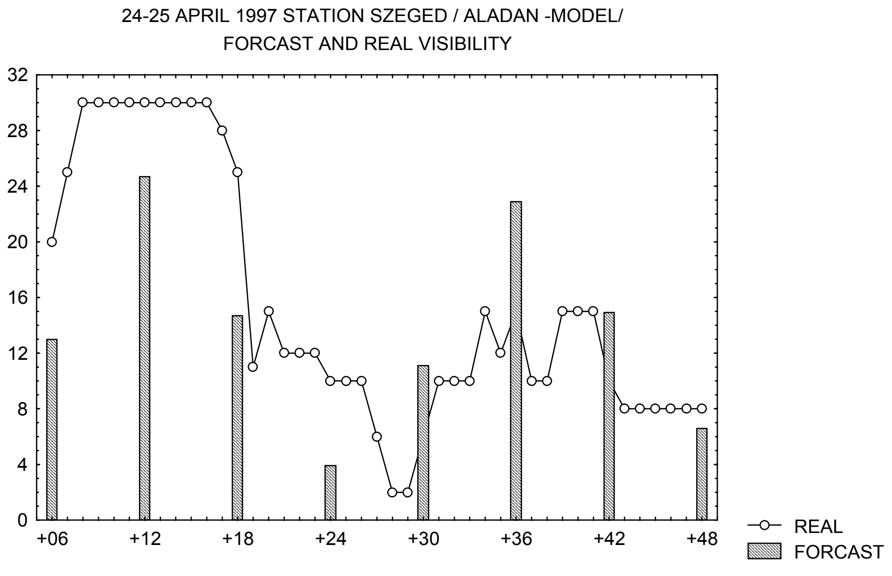
24-25 APRIL 1997 STATION SZEGED / ALADAN -MODEL/
FORCAST AND REAL VISIBILITY
32
28
24
20
16
12
8
4
0
+06
+12
+18
+24
+30
+36
+42
+48
REAL
FORCAST

Fig. 8.

Fig. 1.
Correlation between FOGSI and observed visibility.

Fig. 2.
Correlation between FOGSI and observed visibility (all data derived from critical domain).

Fig. 3.
The first steps of the process.

Fig. 4.
Decision tree in case of normal stratification.

Fig. 5.
Observed and diagnosted visibility.
Fig. 6.
Correlation between observed visibility and calculated with decision tree
Fig. 7.
Correlation between observed and diagnosed visibility.

Fig. 8.
Predicted and observed visibility in Szeged.