D3.4b Road weather model development – Friction model and Fog model

Project:	ROADIDEA 215455
Document Number and Title:	D3.4b Road weather model development
Work-Package:	WP3
Deliverable Type:	Report
Contractual Date of Delivery:	14.05.2010
Actual Date of Delivery:	03.07.2010
Author/Responsible(s):	FMI, ARPAV
Contributors:	Marjo Hippi (FMI), Ilkka Juga (FMI), Andrea Rossa (ARPAV), Franco Zardini (ARPAV), Francesco Domenichini (ARPAV)
Approval of this report:	TC
Summary of this report:	The Pulp Friction pilot combines road weather modelling with background road weather information. New friction model has been run operationally during the winter 2009/2010. The calculation points correspond to road weather stations including road condition sensor, and thus verification data is available. Data from two previous winters were used in the statistical analysis and the friction equation development work. The amount of calculation points in the pilot reached almost 100. The outputs through Roadidea platform are viewed via web browser or mobile phone application. The Veneto Fog Pilot system takes direct and indirect visibility information and merges it into a fog probability and a fog alert product. This system is based on a dedicated visibilimeter network, a meteorological surface observation network and satellite observations. It can work on various levels of complexity, based on the Roadidea visibilimeter network.
Keyword List:	Road weather modeling, friction measurements by optical devices, statistical friction modeling, friction and road conditions Road weather modeling, fog pilot, visibility model
Dissemination level:	Public (PU)

Co-funded by the European Commission under the 7th Community Framework Programme for Research and Technological Development

Change History

Version	Date	Status	Author (Partner)	Description
0.1	12.05.2010	Draft	Marjo Hippi (FMI) Ilkka Juga (FMI)	1st draft
0.2	25.05.2010	Draft	Marjo Hippi (FMI) Ilkka Juga (FMI) Andrea Rossa (ARPAV) Franco Zardini (ARPAV) Francesco Domenichini (ARPAV) Jörgen Bogren (KLIM)	2nd draft
0.3	01.06.2010	Draft	Jörgen Bogren (KLIM)	Final Draft
1.0	30.06.2010	Final	Jörgen Bogren (KLIM)	Final version
2.0	01.07.2010	Final	Pirkko Saarikivi (FORC)	Editorial Check

Distribution list

European Commission	Emilio Davila Gonzalez Wolfgang Höfs
ROADIDEA partners	E-mail list

www.roadidea.eu

Table of contents

1 Introduction – Friction model

Road weather monitoring networks are a very important component of road safety in countries where slippery conditions occur frequently (e.g. during winter time). Road weather models and products have been developed in many countries to help timely implementation of road maintenance operations as well as making better road weather forecasts. The Pulp Friction pilot aims at combining road weather modelling with road weather information as a background. Information on the FMI's road weather model, including friction modelling, and the Pulp Friction pilot were presented in the Roadidea report D3.4a [ROADIDEA, 2009].

The friction model has been run operationally during the winter 2009/10 with the outputs of the runs being stored. The calculation points are located at sites where a road weather station (including the Vaisala DSC111 sensor) is installed, so that verification data can be directly collected from the station measurements. Data from two previous winters were used in the statistical analysis and the friction equation development work. The amount of calculation points in the pilot reached almost 100. The outputs were delivered to the Roadidea platform hosted by Destia. There are two channels for the use of the model, either via Internet browser or mobile phone application.

The winter 2009-2010 was cold in Finland and there was quite much snow on the ground compared to previous winters. Winters 2007/08 and 2008/09 were quite different with higher temperatures and less snow. The latest version of the statistical friction model was developed using observations from these two winters, so it is possible that the formulas are not working so well in colder temperatures.

Information about the quality of friction forecasts and the results of the end user survey are presented in D8.3b.

2 Model development and used data

The FMI road weather model has been under further development during the winter 2009-2010 by improving the friction equations as well as the general structure of the road weather model.

2.1 First version of the friction model

The development of the first version of the statistical model is described in [ROADIDEA, 2009] and the work was also introduced to the R&D community at the EMS meeting in Toulouse, France, 28 September – 02 October 2009 [Hippi et al., 2009] and at the SIRWEC conference in Quebec City, Canada, 5-7 February 2010 [Hippi et al., 2010] and [Nurmi et al., 2010].

The first versions of the statistical friction equations were based on the optical friction measurements data (with the Vaisala's DSC111 instrument) during the winter 2007/08. In those equations, the dependent parameters with substantial correlations in the regression were the road surface temperature and the sum of snow and ice thicknesses. On wet surfaces, the water thickness was the only dependent parameter. These parameters had the highest correlations with the observed friction. FMI also tested the use of

other parameters, for example the temperature difference between road surface temperature and dew point temperature (indicating the formation of hoar frost), but it did not display much correlation with the observed friction in the whole dataset (although in some individual cases it might be important). The main goal for the development work was that the friction equations would later be part of the operational road weather forecasting system, using background (input) data from a road weather model.

The resulting statistical equations for friction were:

For snow and/or ice covered roads:

Friction = A x f(snow_mm + ice_mm) + B x T_road + C, where the thicknesses of snow and ice are in equivalent mm, f is a mathematical function/operation, by which the nonlinear and scattered distribution is made more compact for regression

For water covered roads:

Friction = D x water_mm + E

For dry roads:

Friction = 0.82

The coefficients A-E are statistical fitting parameters depending on local conditions. These equations were calculated for a couple of points in southern and eastern Finland based on observations from winter 2007/08 and they were tested with independent data from winter 2008/09. The results were promising with relatively high correlations between the modelled and observed friction (0.85 for snow and ice covered roads and 0.93 for water covered roads). However, when the observed friction was low, the statistical equation gave a bit too high friction values in average. This might result from the fact that winter 2007/08 was very mild and during winter 2008/2009 there were more wintry conditions. Another problem that was noticed later was the too high thicknesses of snow and ice in FMI's road weather model. The friction equations use the model data as background information and any systematic error in that data could result in too low friction values in operational forecasts.

2.2 Further development of the statistical equations

The friction model equations were further developed by extending the dependent data and examining the utilization of different mathematical functions in making the observed friction distribution more linear and compact for regression application. We also investigated more thoroughly the connection between friction and the thicknesses of snow, ice and water. A special situation appears when all those three forms exist simultaneously. This can result either from prevailing temperature around 0 ^{o}C or from the use of salting in colder conditions (then the water is supercooled).

As the method was further developed at FMI, the dependent data covering two winters, 2007/08 and 2008/09 was validated against a new independent data-set from winter 2009/10 (November - March). All these data were delivered by the Finnish Transport Agency. The new updated statistical equations were calculated using the dependent data from the two earliest winters at four locations: Utti, Anjala, Orivesi and Kuopio (see loca-

tions in Fig. 1). The data from those four stations covered in total 207943 individual observations, of which 40503 (19.5%) were connected to snow and/or ice covered roads, 64602 (31.1 %) to water covered roads and 102838 (49.5 %) were linked to dry road surface. Scatter plots of observed friction vs. the thickness of snow and/or ice as well as vs. the water layer thickness are depicted in Figure 2. The regression equations were calculated with these data, and for comparison, separate local equations were also calculated for those four locations using only the data from the specific station. Then, both the "general" and local site equations were tested with the independent data (winter 2009/10) at the four locations: Utti, Anjala, Orivesi and Kuopio.

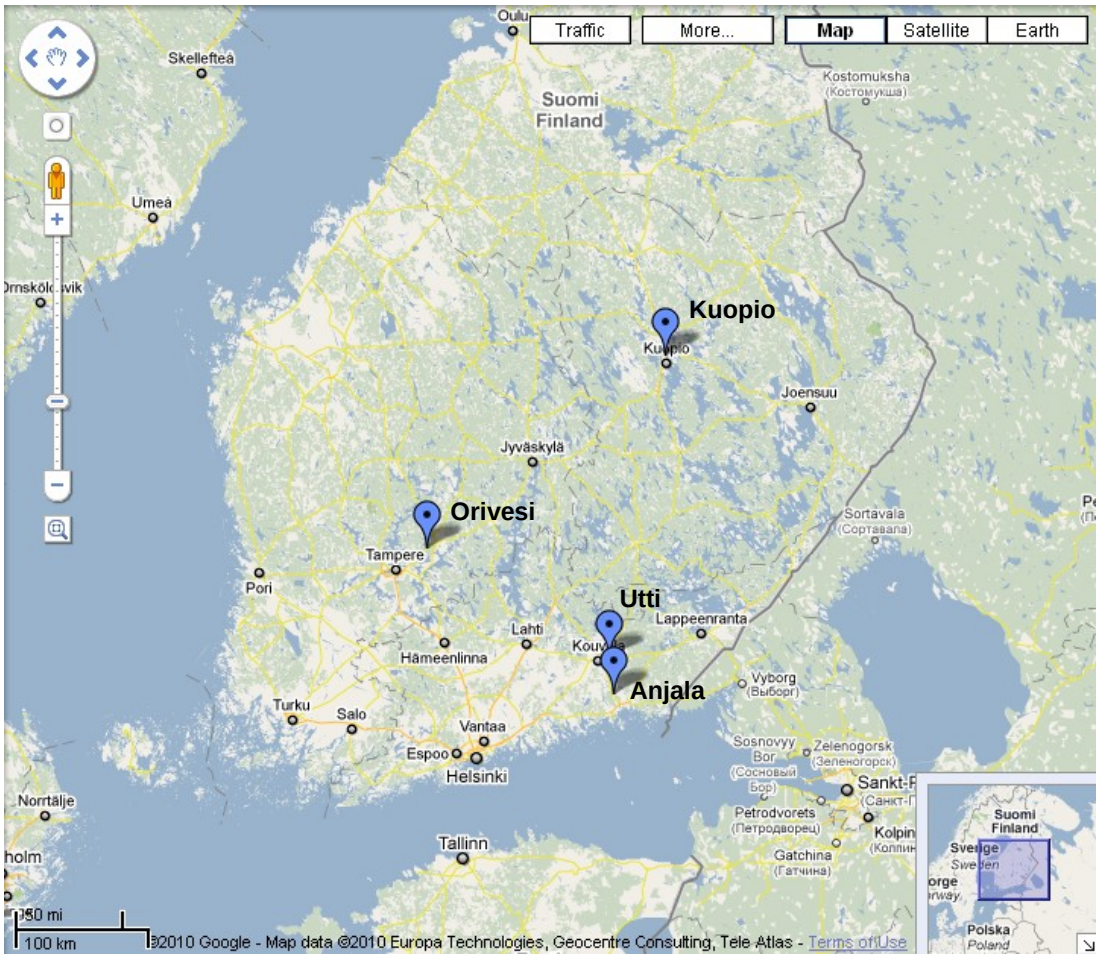


Figure 1. Locations of Utti, Anjala, Orivesi and Kuopio road weather stations.

2.3 General road weather model development

The road weather model has been improved during this last winter. One of the main problems of the model has been the too big storage of water/snow/ice/frost. The amount of ice and frost has been the most troublesome values. The problem is that too big amount of modelled ice/snow/frost on the surface produces too low values of the

modelled friction. The main improvement work has been focused on the storage process and the influence of traffic has been made more efficient so that the wearing of water/snow/ice/frost occurs faster now in the model. Some changes dealing with road surface temperature were also introduced. After these improvements, the model seems to perform better, but the storage terms are still too big compared to observations.

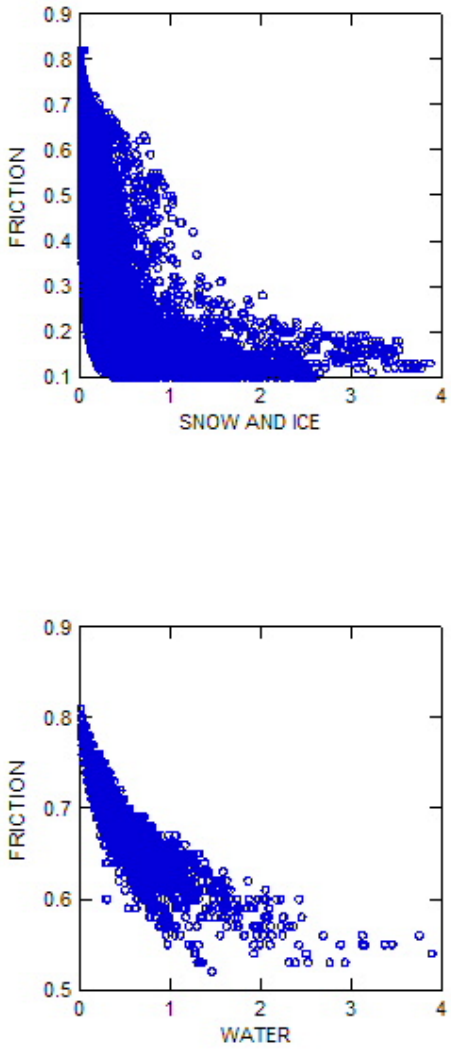


Figure 2. Scatter plot of observed friction vs. the layer thickness (water content) of snow and /or ice (above) and vs. the thickness of water (below) on the road. Data based on DSC111 measurements during winters 2007/08 and 2008/09 in Utti, Anjala, Orivesi and Kuopio (data source: Finnish Transport Agency).

3 Implementations of the friction model

One of ROADIDEA project's goals was to develop new services for drivers. In the Pulp Friction pilot two different kinds of services were developed for end users. One of the implementations is a Google maps-based product presenting short time road weather forecasts, including friction forecasts (see Fig. 3). The product can bee seen on the web page http://pilot.roadidea.eu/friction/ and the site is available for everyone. This product is mainly aimed for professional end users, like meteorologists and road maintenance personnel. There has been about 90 calculation points where the friction model has been running operationally.

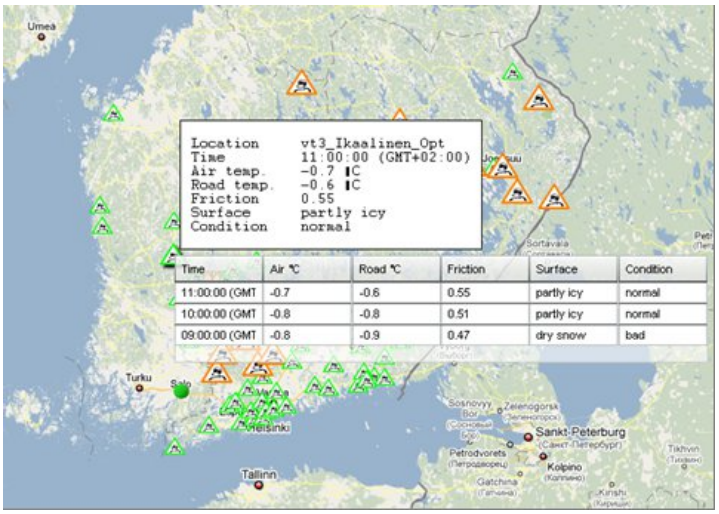


Figure 3: Forecasted friction and other road weather parameters presented in the Google maps ROADIDEA application. It is based on close collaboration between ROADIDEA partners FMI, Destia and Demis.

Another product for end users is the mobile phone application presenting information about road weather forecasts and accident information. Additionally, pictures from the nearest road weather cameras can be seen via this service. The application locates the user and the presented information is from the area where the user is situated. This service is aimed for all drivers. A screen shot of the application is presented on Fig. 4.

Figure 4: Screen shot of ROADIDEA mobile phone application presenting e.g information of (observed and) forecasted friction.

4 Results: the statistical friction equations and their validation

The dependence of friction on snow, ice and water thickness was investigated at FMI. Pearson correlation coefficients were calculated by using the dependent data from winters 2007/08 and 2008/09. The distributions were quite wide and nonlinear (Figure 2), so that we had to first make them more compact with the use of a mathematical function/operation, after which the linear regression was applied. Both logarithmic and square root functions were tested here, and the square root proved to be the most workable. Applying the square root function to the thicknesses of water, ice and snow substantially improved the Pearson correlations in the dependent data. For example, while the correlation between snow and friction was -0.50, the correlation between the square root of snow and friction rose to -0.71. Eventually, the resulting statistical method for friction prediction is:

Friction F1 for icy and snowy road surfaces:

$$F1 = A_1*SQRT(Snow\_mm)+B_1*SQRT(Ice\_mm)+C_1*T\_road+D_1$$

Friction FW for water covered road surfaces:

$$FW = A_w*SQRT(Water\_mm)+B_w$$

In case of simultaneous existence of snow, ice and water on the road, the equation is:

$$F2 = A_2*SQRT(Snow\_mm)+B_2*SQRT(Ice\_mm)+C_2*SQRT(Water\_mm)+D_2*T\_road+E_2$$

As mentioned earlier, this situation (equation F2) is the result of temperature being around 0 ^{o}C or of salting actions. When salt is used in sub-zero temperatures, it melts

snow and ice and the resulting water layer is super cooled. The effect of salting is not included in the present version of FMI's road weather model and super cooled water cannot exist in the model, so the operational use of equation F2 is not reasonable and F1 is used instead in the friction pilot application. Anyway, the equation F2 was also tested here with the independent data to see if it had better scores than F1.

The coefficients in the equations of F1, FW and F2 were calculated from the whole dependent data: Winters 2007/08 and 2008/09, observations from Utti, Anjala, Orivesi and Kuopio. In addition, we calculated local point models for these four locations (using the two winter data from the specific location only). These equations were validated with the independent data (winter 2009/10) from these four points. The observed thicknesses of snow, ice and water were used in calculation of the modelled friction F1, F2 and FW, and those friction parameters were then validated with the observed friction. The results are shown in Table 1.

Table 1. Pearson correlation coefficients for modelled friction (F1, F2, FW) when tested with observations from winter 2009/10 (independent data). The verifying friction observations are from four locations: Utti, Anjala, Orivesi and Kuopio. The local friction models for these points were also evaluated.

	Utti	Anjala	Orivesi	Kuopio
F1	0.86	0.83	0.78	0.48
F2	0.88	0.88	0.81	0.61
FW	0.98	0.98	0.97	0.96
F1, local model	0.85	0.87	0.82	0.53
F2, local model	0.86	0.88	0.85	0.62
FW, local model	0.98	0.98	0.97	0.96

The results depicted in Table 1 show that the Pearson correlations for the equation (FW) of water covered roads in all four locations are very high (both for the "general" and local model), so we have a straight forward, relatively simple relationship between the water layer thickness and friction. For the equations (F1 and F2) of snow and ice covered roads (and in some cases also simultaneous existence of water) the correlations are lower. The results for Utti and Anjala (in southeastern Finland) are relatively good (see also Fig. 5), but the values for Kuopio are much weaker. This might result from the fact that the winter 2009/10 data included on average much smaller thicknesses of ice than the data from the previous two winters at the Kuopio location. However, winter 2009/10 was cold especially in eastern and northern Finland and there were probably more days with dry cold winter weather and rather loose/drifting snow on the roads than ice.

If we compare the values of F1 and F2 we notice that the inclusion of water in the equations for snow and ice covered roads (F2) improves the correlations a bit, so the slushy conditions are then modelled in a more proper way. Also the local models for F1 and F2 have a bit higher correlations than the "general model" except for Utti, where the "general model" had slightly higher correlations than the local point model.

Figures 5 and 6 visualize the validity of the friction modelling equations. From Fig. 5 we can see that although the distribution in the scatter plot of modelled (F1) vs. observed

friction is quite wide and a bit curved, the result is relatively good, with great majority of the values near the diagonal. For the equation of water covered roads (FW) the results are very good (as the correlations in Table 1 show also). Fig 6 shows the modelled (F1) and observed dependence of friction on the thickness of snow and ice. We can see that the modelled friction simulates the observed one quite well. Although the absolute values of the modelled friction at the lower end with thicker snow and ice layer are not exactly correct, the accuracy is sufficient for example in such applications where we use categorization (thresholds, for example 0.6/0.4/0.2).

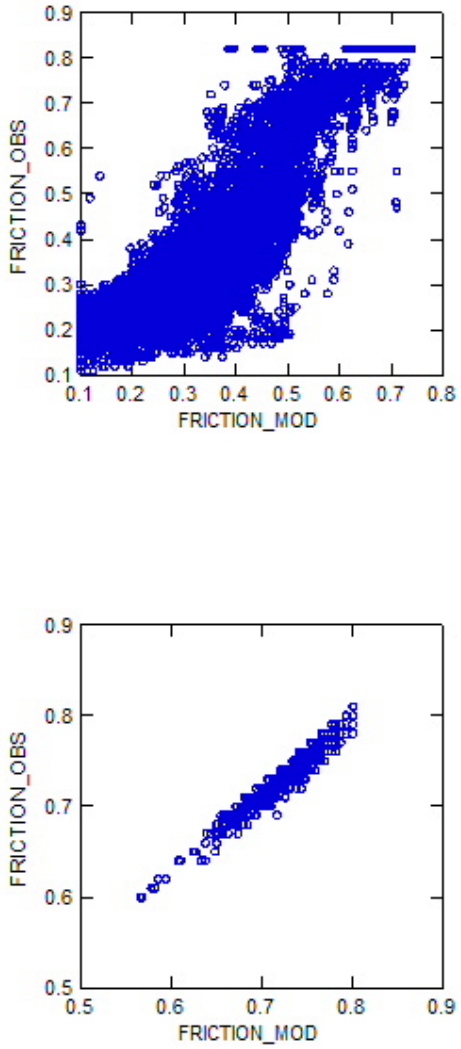


Figure 5. Above: Scatter plot of observed (FRICTION_OBS) and modelled (FRICTION_MOD) friction (equation F1) in case of snowy and/or icy road, Pearson correlation r = 0.86. Below: The same for water covered roads (equation FW), r = 0.98. Based on data from Utti during winter 2009/10 (the independent data).

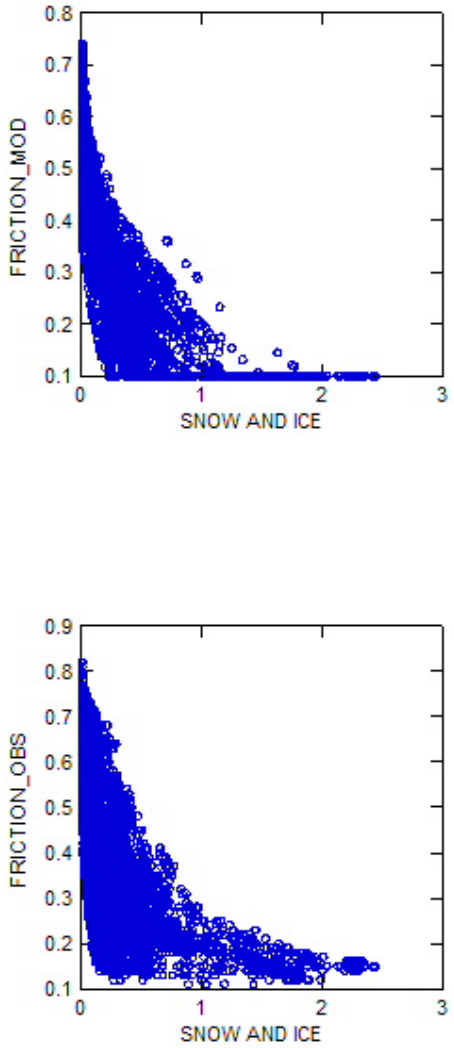


Figure 6. Modelled friction F1 (Friction_MOD, above) and observed friction (Friction_OBS, below) in function of the observed thickness of snow and ice (water content). Based on observations in Utti during winter 2009/10 (the independent data).

One of the problems in the friction modelling is the lack of information of road maintenance actions. Hopefully the information of salting and snow removal would come public and available so the information could be used to initialize the state of the roads in the road weather model (the background information of the friction equations).

5 Discussion

Modelling of road surface friction seems to be a new way to predict slipperiness and this study is a pioneer in its field globally. The friction model has raised interests in many places where it has been presented. As the work presented here is based on the observations made by Vaisala's optical DSC111 instrument, with this statistical modelling we are in a way simulating the measuring instrument itself. The friction values measured by

the device have been validated (Pilli-Sihvola, 2008) and a relatively good agreement with mechanical friction measuring devices has been observed.

The results for friction modelling are promising although verification results point there is clear room for further improvements. The statistical model could be improved to some extent with more data covering different kinds of weather situations and point locations. Also, we could improve the mathematical and statistical methods used in the equations. The meteorological input to the friction model comes mainly from the background data from the road weather model and it is very important to validate and improve that model also. The lack of the effect of road maintenance actions in the road weather model is an insufficiency that is challenging to improve. When evaluating the friction forecast we should keep in mind that the road maintenance actions worsen the verification results. The friction forecasts should rather be regarded as risk forecasts, indicating the worst road conditions that can exist without maintenance actions.

As an alternative, instead of deterministic forecasts it would be suitable to calculate probabilistic friction or slipperiness forecasts using the developed friction model.

6 References

[Hippi et al., 2009] Hippi, M., I. Juga and P. Nurmi: A statistical model for road surface friction forecasting applying optical road weather measurements. EMS Annual Meeting Abstracts, Vol. 6, EMS2009-389, 2009. 9th EMS/9th ECAM. Available from http://meetingorganizer.copernicus.org/EMS2009/EMS2009-378.pdf.

[Hippi et al., 2010] Hippi, M., I. Juga and P. Nurmi: A statistical forecast model for road surface friction. SIRWEC 15th International Road Weather Conference, Quebec City, Canada, 5-7 February 2010. Available from http://www.sirwec.org/conferences/Quebec/full_paper/15_sirwec_2010_paper_hippi.pdf.

[Nurmi et al., 2010] Nurmi, P., M. Hippi and I. Juga: Evaluation of FMI's new forecast model of road surface friction. SIRWEC 15th International Road Weather Conference, Quebec City, Canada, 5-7 February 2010. Available from http://www.sirwec.org/conferences/Quebec/full_paper/21_sirwec_2010_paper_nurmi.pdf.

[Pilli-Sihvola, 2008] Pilli-Sihvola, Y.: Optical remote road surface state and temperature sensors: Survey of functionality and usability. ITS World Congress New York, November 16-20, 2008.

[ROADIDEA, 2009]: ROADIDEA D3.4a Road Weather Model Development. ROADIDEA document.

7 Introduction – Fog model

During the Roadidea project, partner ARPAV have developed the Veneto Fog Pilot system, which takes direct and indirect visibility information and merges it into a fog probability and a fog alert product. This system is based on a dedicated visibilimeter network, a meteorological surface observation network and satellite observations. It can work, in principle, on various levels of complexity, based however on the Roadidea visibilimeter network. Complexity here mainly means the way in which meteorological information is related to the presence of fog, extended spatially, and the usage of PBL specific variables which are believed to be important for fog diagnostics.

In the perspective of judging the portability of the Veneto Fog Pilot system it is necessary to evaluate the degree of added value from added complexity. Again, this is especially important for the effort needed to pre-process the data from the meteorological surface observation network, and the description of the state of the planetary boundary layer (PBL). The simpler the system, the more portable the system! Even in its simplest form, there are a number of tuning parameters (e.g. relative weights of visibility measurements to satellite observation) whose value may not be known a priori.

The overall work needed to systematically evaluate the quality of the various components of the fog alert system can be subdivided into:

- performing sensitivity analysis w.r.t. PBL variables and different methods of spatial interpolation;
- optimizing the relative weights of visibility observations, meteorological information, and satellite data;
- illustrating the impact/benefit of additional data sources (visibility reports of drivers).

This verification and optimization benefitted significantly from the Roadidea extension, in that valuable data became available during the Winter season 2009-10. As a matter of fact, the Winter season 2008-09, in which the data collection of visibility measurements of the ARPAV Roadidea network and the satellite information just started, featured an unusually low frequency of fog events.

As the Fog Pilot is inherently a probabilistic tool verification needs to be done accordingly. The methodology is not straightforward and, therefore the optimization of a number of model parameters is presented, based on quality measures defined. The potential impact of parameters describing the state of the PBL is presented and finally also the value of additional visibility information is illustrated and the potential of the open architecture of the Fog Pilot for becoming a community tool is underlined.

8 Evaluation and verification of the fog pilot model

8.1 Verification methods and measures for probabilistic estimates

The Fog Pilot output consists of areal products which are constructed on a limited set of direct in situ visibility measurements (visibilimeter network) and indirect observations (meteorological station network and satellite cloud classification) to interpolate the point information on to the area. Verification is, therefore, of principal importance for establishing the "goodness" of the fog probability maps and warnings that are output from the Fog Pilot.

The verification process was chosen to act on the products derived from the individual data sources, as well as on the final merged product. This allows judging the added value of the individual data sources over a simple spatial interpolation of direct visibility measurements in a systematic way. Also it provides measures of the performance of the product, which in principle allows a feedback to correct the product by tuning internal parameters of the Fog Pilot model. Such a measure also yields the best probability threshold for fog warning by balancing the False Alarm Rate and Probability of Detection of the output. Two kinds of reference data were used in the verification process:

- Products derived from indirect observations can be compared directly to the visibility measurements as they are independent;
- Products derived from the direct visibility observations are not independent of the visibility measurements, so that a cross verification needs to be applied; this means that for any visibility measurement the product was calculated by excluding this observation and then compared to it, as it now is independent; this method was applied for the visibilimeters alone and the merged products;
- All products can be compared with direct human observations (performed for a test period in winter 2009-2010).

As the Fog Pilot is inherently a probabilistic instrument, methods used for the verification need to be selected accordingly. Of the many statistical indices and methods available for verification purposes, the following, which can be considered standard, were applied:

- Reliability diagram, which compared the probability estimates with the effective observed fog frequency given the probability estimate from the fog model; when these correspond, and therefore these points lie close to the 1:1 line, the model is said to be reliable;

- Probability of Detection (POD) and Probability of False Detection (POFD, i.e. false alarms) for various event thresholds, and the corresponding relative operating curve (ROC), which is obtained by connecting the resulting POD/POFD values;.if the ROC curve is close to the 1:1 line, there is no skill in the model, while the more the curve approaches the upper-left corner the more skilful the model, as large values of POD are obtained with low values of POFD; conversely, if the ROC curve lies below the 1:1 line, the model had less skill than a random guess, as the POD is smaller than the POFD. The area between the 1:1 line and the ROC curve is called ROC area and is positive/negative if the curve is above/below the 1:1

line. The ROC area is a frequently used summary measure for probabilistic forecast models.

Another index of goodness of the Fog Pilot estimates is proposed here that measures the difference between the average of the probability density distribution of the "events" and the "non-events". This measure expresses, to some extent, the "event non-event separation" and is large for a good model which is able to distinguish between fog and no-fog easily, while it is small for a poor model.
Examples of evaluation indexes are given in the results section.

8.2 The economic value of a warning system

It is instructive to express the impact of a warning system in terms of its economic value. This is done by assuming a loss L which is incurred in case of an event without having taken mitigating action, and the cost c of the mitigating action. This cost can be compared with a number of scenarios, such as never or always taking mitigating action, taking action with a perfect or a random forecast system (see equations 2.1). It is clear that the benefit of a warning system depends on the ratio of c / L, the cost-loss ratio. If the costs for mitigating actions is very small compared to the potential loss, it will be convenient to minimize the misses at the price of having many false alarms (which will not cost much in this case). If conversely c is large mitigation is not convenient and one may have to bear the losses (which may not be very high in this case).

To calculate the relative economic value of a realistic warning system it can be compared to having no warning system at all, i.e. to the scenario of never taking mitigating action. The headroom for the benefit, on the other hand, it given by the costs incurred having an hypothetical perfect warning system.

Table 2. Contingency table for verification and assessment of the relative economic value of the Fog Pilot warning system. 'hits' denote the correct warnings, 'false alarms' the incorrect warnings, 'misses' the missed warnings, and 'correct negatives' the correctly non-warnings. Note that hits + misses = number of events, and false alarms + correct negatives = number of non-events.

	EVENT	***NOT EVENT***
WARNING	Hits	false alarms
NO WARNING	Misses	correct negatives

Using the notation defined in the contingency table shown in Table 2 the total costs of various scenarios are expressed in equations 2.1. For example, no warning system will cost L every time an event happens, while a perfect warning system reduces the costs to c every time an event happens, assuming $c < L$. The relative economic value can then be expressed as the relative difference of costs between any too approaches. In the case of a realistic (non-perfect) warning system is given by $1 - c / L - misses / n_{events}$. So the difference to a perfect forecast system is equal to $misses / n_{events}$ which highlights the importance of keeping the misses low in case of large losses L.

$$COST(never\_protect) = L \cdot n_{events}$$

$$COST(alway\_protect) = c \cdot (n_{events} + n_{non\_events})$$

$$COST(perfect\_warning) = c \cdot hits = c \cdot n_{events}$$

$$COST(protect\_when\_warned) = c \cdot (hits + false\_alarms) + L \cdot misses$$

$$Economic\_Value = \frac{L \cdot n_{events} - c \cdot (hits + false\_alarms) - L \cdot misses}{L \cdot n_{events}} = 1 - \frac{c}{L} - \frac{misses}{n_{events}}$$

$$Economic\_Value(never\_protect) = 0$$

$$Economic\_Value(perfect\_warning) = 1 - \frac{c}{L}$$

(2.1)

The same concepts can be expressed in terms of the POD and POFD, which can be seen as characterizing a warning system in relative terms. From equation 2.2 it is clearly seen that a high POD boosts the economic value, while a high POFD is reducing it. For example, for a cost-loss ratio of 0.5, POD of 70%, POFD of 20%, and an event-non_event ratio of 1:5 the economic value is –0.15, while for a cost-loss ratio of 0.2 it becomes +0.36. It emerges that if the cost for mitigating action is relatively high, even a good warning system is not able to reduce costs compared to having no warning system. The most impacting factor here is that rare events are inherently linked to many false alarms, given the POFD.

$$COST\ (protect\ \_when\ \_warned\ ) = c \cdot (n_{npn\_events} \cdot POFD\ + n_{events} \cdot POD\ ) + L \cdot (1 - POD\ )$$

$$Economic\ \_Value\ = (1 - \frac{c}{L}) \cdot POD\ - \frac{n_{non\_events}}{n_{events}} \frac{c}{L} \cdot POFD$$

(2.2)

8.3 Data set

The data set that was used for this verification exercise spans the 175 day period from 4 September 2009 to 26 February 2010 (taking advantage of the Roadidea extension!). More specifically, a subset of around 760 hours, including some 20 days with fog occurrence, was selected from this period in order to evaluate the pilot 'around' fog situations. This was done on the one hand to reduce the size of the data set and thus improve the performance of the verification (and optimization) procedure. On the other hand, it reflects the effective operational deployment during meteorological conditions which can produce fog. Indeed, during summer or during the passage of a perturbed weather system the (climatological) fog frequency is close to zero. For the chosen verification data set the base rate, or frequency of occurrence, of fog is about 1:5, i.e. five times more hours with good visibility than hours with reduced visibility. The single record is composed by the following elements:

- The visibility coming from a visibilimeter of the network
- The correspondent information on "event" or "non-event", which depends on the chosen warning threshold (500 meters in our case)

- an estimated probability for “event”
- a date label
- a describing label (e.g. location, instrument name).

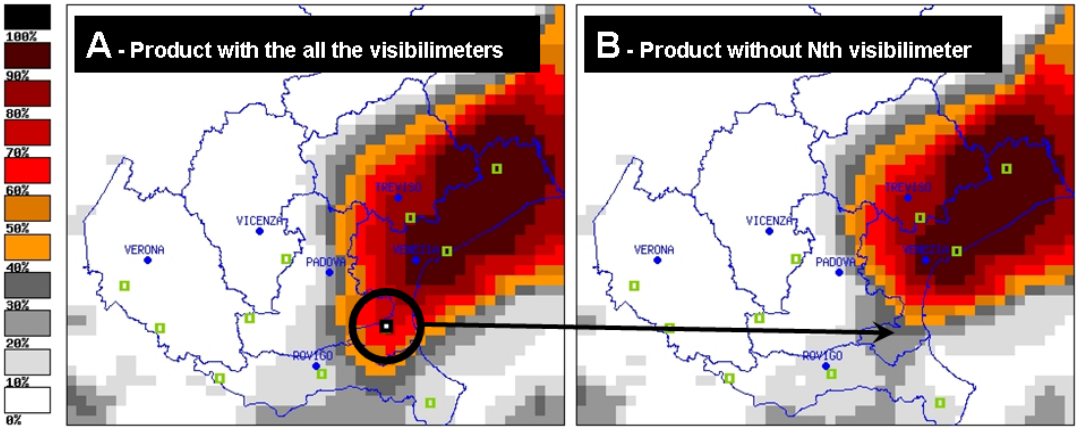


Figure 7. Fog probability product obtained using the complete visibility network (A) and without a visibilimeter (B). The production in modality B use the Nth visibilimeter for the verification.

The estimated probability is the value calculated on the grid mesh containing the visibilimeter station.

Probabilistic verification of products

In case of probability fields generated independently of the visibility information, we apply the verification to the normal dataset (satellite estimation, and CART estimation from meteorological parameters at the ground), compared with the visibility records at the ground.

Probabilistic cross verification of products

In case of probability fields generated with the contribution of the visibility records, we create a special dataset. Every record referred to a specific visibilimeter contains an estimated probability value generated excluding the information coming from the visibilimeter itself, but including the other visibilimeters in the network. This method is particularly useful when is impossible to find an external measurement of the phenomenon. The so produced estimate is independent of the visibility measurements, and provides the excluded measurements for independent verification.

The case in the **Error! Reference source not found.** shows the difference in the product obtained “crossing” a visibilimeter in the network. The numerical result in the B map, in the point correspondent to the Nth visibilimeter, will be compared with the value measured by the visibilimeter. In that case, a value of 42% of fog probability will be compared with a 450 meters of measured visibility (i.e. event).

Verification output

Standard validation for Satellite and Stations, and crossed validation for Visibilimeters and final merged output are reported.

Human observations and webcam:

The human observations on the field, recorded until now, confirm in general the reports obtained from visibilimeters. A systematic analysis of the images from the archive of a webcam in the town of Venice was used to complement the Fog Pilot verification. The results of product are compared with this direct estimation of visibility, performed along the period 1 November 2009 – 1 April 2010.

The results are available in the following Paragraphs.

a b c

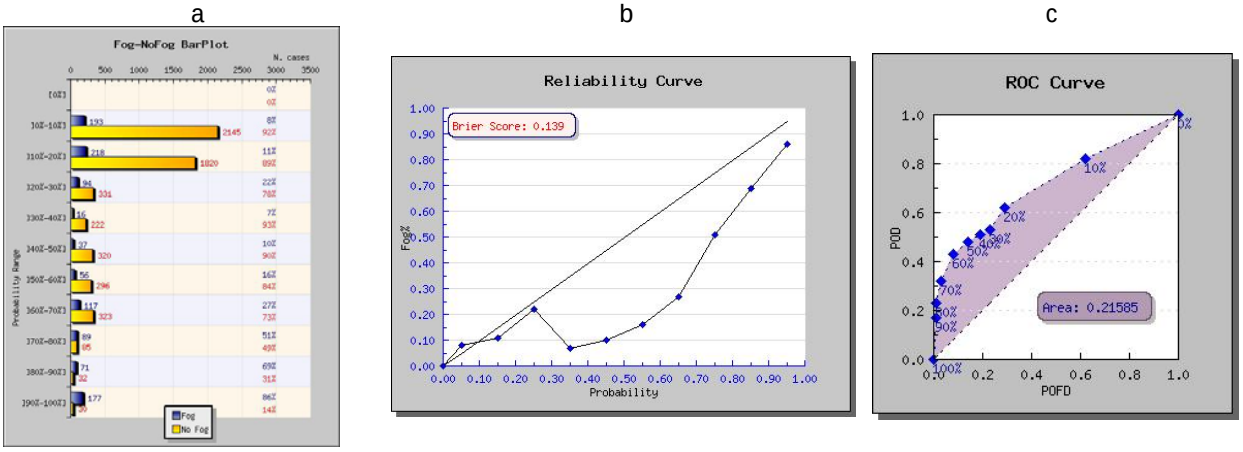


Figure 8. Field of probability of SAFNWC-based method. a) Probability class distribution of the events: blue bars are "Event" (fog) and yellow ones "no-Event". b) Reliability curve of fog estimation. c) Relative operating curve (ROC) of the estimations.

8.4 Product performance

In this section the results of the verification of the sub-products and the final, merged product are presented in form of a probability density distribution (Panels a of Figures 8, 9, 10, 11 and 12), a reliability curve (Panels b), and a ROC diagram (Panels c), which are discussed. Also, the indices POD and POFD in the ROC diagram can be used to evaluate the economic value for a chosen probability threshold. These descriptions allow to assess the added value of additional observations in the Fog Pilot.

8.4.1 SAFNWC-based product verification

The verification applied to satellite-derived fog probability shows a lack in reliability for middle ranges probability estimation (30-70%). However, the results are acceptable for our aims, even if this behaviour can be improved by tuning parameters of the satellite estimate model. At the current stage, the response of the SAFNWC derived fog estimates can be considered satisfactory, since the most part of the samples are classified in the probability range from 0% to 20%, that shows a good reliability, and the high probabilities range classify well the fog events, as shown in ***Error! Reference source not***

found.a, where the bar plot evidence the good distribution of probability estimation. The reliability appears to fall down in the middle probability estimations, suggesting an over-estimation of fog cases (***Error! Reference source not found.***b). For this reason the ROC curve shows a lack of probability of detection in the middle probability thresholds.

8.4.2 meteorological surface station-based product verification

CART processing of meteorological data from the ground stations gives a very reliable result for low probabilities range (0 to 30%), but does not gives information on higher

a b c

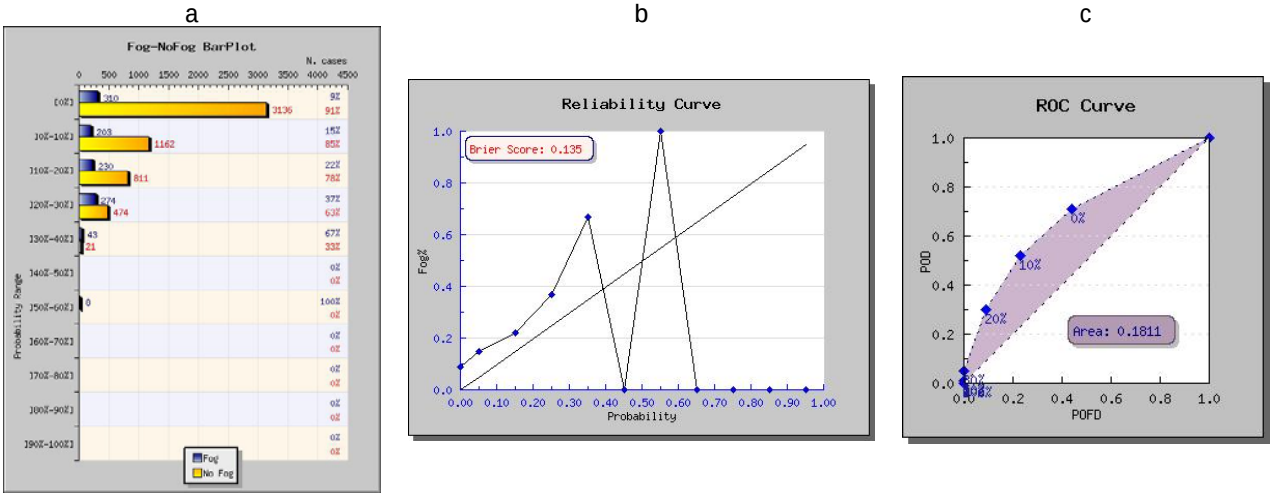


Figure 9. Field of probability of Meteorological surface station-based method. a) Probability class distribution of the events: blue bars are "Event" (fog) and yellow ones "no-Event". b) Reliability curve of fog estimation. c) Relative operating curve (ROC) of the estimations.

a b c

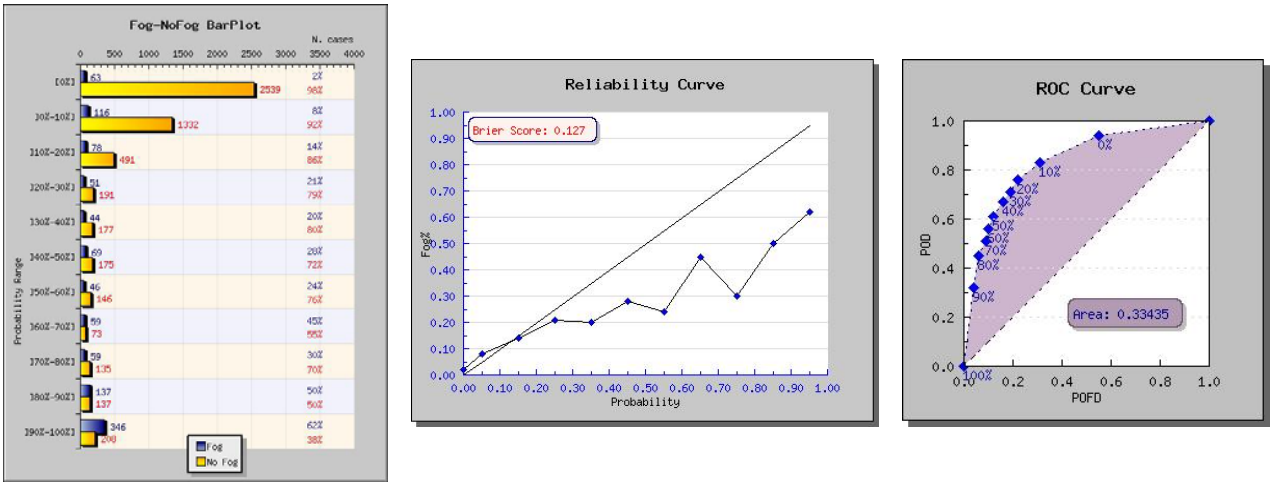


Figure 10. Field of probability of Visibilimeter-only-based method. a) Probability class distribution of the events: blue bars are "Event" (fog) and yellow ones "no-Event". b) Reliability curve of fog estimation. c) Relative operating curve (ROC) of the estimations.

probabilities (see Figure). This product does not discriminate between "very foggy" and "moderately foggy" situations, except for a very small number of cases with probability of

around 60 percent. That is why, in the final product, a quite low weight has been assigned to the CART

8.4.3 Visibilimeter-only-based product verification

The product derived from direct visibility observations shows a good discrimination between “event” and “not event”, but also an overestimation of fog increasing with the estimated probability (see Figure 10.). This behaviour is related to the spatialization process. In the final product, using the weight of this information (related to the distance of the instruments), the overestimation present in the field produced by the visibility network is corrected by the merging process.

a b c

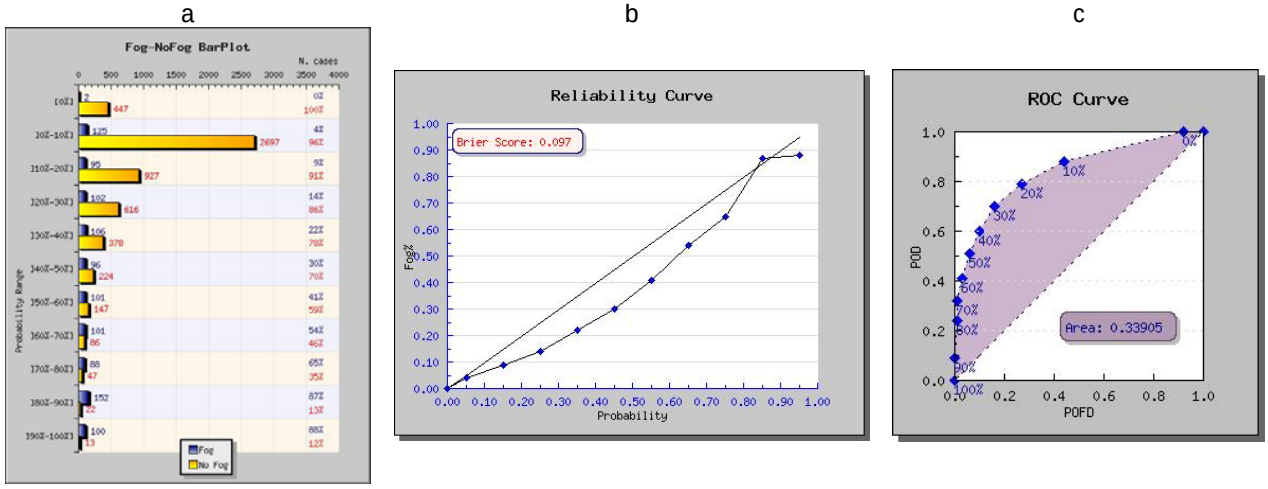


Figure 11. Field of probability of the Fog Pilot Merged final product. a) Probability class distribution of the events: blue bars are “Event” (fog) and yellow ones “no-Event”. b) Reliability curve of fog estimation. c) Relative operating curve (ROC) of the estimations.

8.4.4 The merged fog pilot product verification

The verification of the final product gives encouraging and consistent results, in that the overall performance is superior than for the single data source-based products, except maybe for a slight overestimation of the frequency of the events. The bar plot graphic shows as no-fog cases are well classified and only a few cases of fog are missed (Figure 11a). The reliability of the overall product is well distributed along the different probability values, despite a problem common to the most part of the distribution is a slight overestimation of the fog probability (probability > frequency, Figure 11.b). Finally the ROC shows the behaviour of the Fog Product, varying the probability threshold for fog alert; thresholds around 30-50% are a good compromise that allows having low false alarm in front of relatively high probability of detection (Figure 11.c).

To conclude, the Fog Pilot monitoring system, at the current stage, allows to detect the fog presence (defined as visibility less of equal 500 meters) by choosing the 30% of probability as the threshold for issuing a fog alert, with a Probability Of Detection of around 70% (we match the 70% of events) and False Alarm Rate of 20% (we meet the 20% of possible false alarm). As illustrated earlier, the relative economic value of such a

warning system is 0.36 for a cost-loss ratio of 0.2. This means that the total costs incurred when having no warning system can be reduced by 36% percent.

8.4.5 Verification estimates from a webcam located in the city of Venice

A coarse estimation of the actual visibility was performed through the analysis of an images archive recorded by a webcam located by ARPAV on a building roof (around 15 meters height) in the city of Venice. Thanks to the presence of several benchmarks in the picture it was possible to give an acceptable estimate of fog or no fog condition (visibility below or above 500 meters). The product output was extracted for the specific point (area of 4x4 km) corresponding to the city of Venice, for a dataset of 640 cases (640 hours) during the winter 2009-2010.

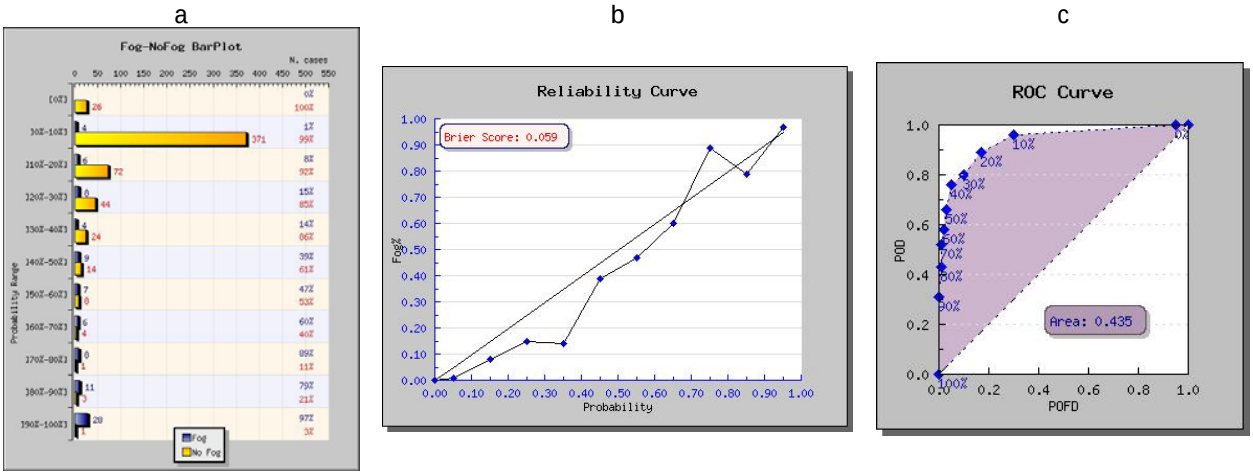


Figure 12. Field of probability of estimates from a webcam. a) Probability class distribution of the events: blue bars are "Event" (fog) and yellow ones "no-Event". b) Reliability curve of fog estimation. c) Relative operating curve (ROC) of the estimations.

These results, as shows Figure 12b, are very good, significantly better than the ones obtained through the cross validation procedure (see last paragraph). The reliability of the product is almost equivalent to the result of cross validation (Figure a); it becomes better in the high probability range, over the 60%. Indeed, in this case for a fog probability warning threshold of 30% the POD = 80% and POFD = 10% the system yields a relative economic value of 54% for a cost-loss ratio of 0.2 (see Figure c)! This encouraging result maybe due to a good correlation of the fog occurrence in Venice and the nearest visibilimeter in Cavallino at a distance of about 12 km. Also, the coherence of fog over this distance, about a third of the average distance of the visibilimeters in the network, maybe significantly larger than at longer distances.

More independent points of verification, like the webcam of Venice, have to be used, in order to investigate the performance of the product in different meteorological situations and geographical positions.

8.5 Conclusions

An extensive verification procedure was set up in order to assess the meteorological and warning quality of the Veneto Fog Pilot. The main findings can be summarized as follows:

- The extension of the Roadidea project allowed to extend the data set collected in the previous winter 2008-09, which exhibited an unusually low frequency of fog occurrence; the data set comprises some 760 hours, including some 20 days with fog occurrence, of observations with a fog-non fog ratio of about 1:5;

- the fog pilot is inherently a probabilistic warning system, thus requiring a probabilistic verification approach; the system has been characterized by the frequency distribution, a reliability diagram, a ROC curve constructed with the probability of detection (POD) and the probability of false detection (POFD), and a relative economic value;

- systematic cross validation of the products coming from the individual data sources reveals low value for the meteorological surface station values; acceptable values for the high and low fog probabilities coming from the satellite, and best values for the simply interpolated visibilimeter network;

- the merged product features a clear added relative to the individual products, with a POD of 70% and a POFD of 20% for a fog warning threshold of 30% resulting from the systematic cross verification which yields a relative economic value of 36% for a cost-loss ratio of 0.2;

- verification of the merged product with webcam imagery suggests even higher scores, in the case of the City of Venice for a fog probability warning threshold of 30% the POD = 80% and POFD = 10% which yields a relative economic value of 54% for a cost-loss ratio of 0.2;

- finally, Table 3 and Figure 13 summarize the economic value of various warning system setups, i.e. never warn, always warn, individual components as well as the merged version of the Fog Pilot, and the value for a site like Venice. It can be observed that all the individual components of the Fog Pilot have a value, and that the merged product yields the best performance, also in terms of cost minimization and relative economic value.

Table 3. Comparison of the performance indices (POD, POFD, total cost COST, and relative economic value) between the individual components and the merged Fog Pilot products for a cost/Loss ratio of 0.2; the reference values are the absence of action (never protect) and the maximum of the security (always protect). Result column "Venice" is referred to the value of the merged product with reference to the independent webcam observation of Venice city.

	never	always	CART	SAF	VIS	MERGED	Venice
BEST POD	0	1	0.52	0.43	0.76	0.73	0.76
BEST POFD	0	1	0.23	0.008	0.22	0.18	0.05
COST (L)	1	1.2	0.8254	0.74	0.6229	0.6049	0.4544
economic value	0	-20%	17%	26%	38%	40%	55%

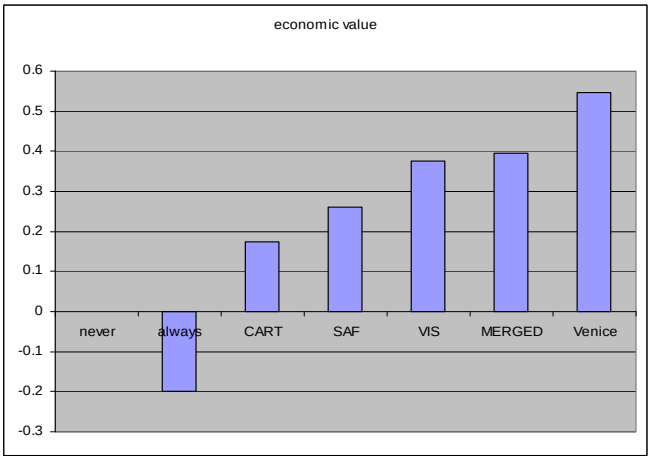


Figure 13. Relative economic value of the different warning systems listed in Table. 3. The merged product (last two values) gives the best results.

9 Optimization by tuning of model parameters

9.1 The range of influence of visibilimeters: The distance factor

The distance factor is an inner parameter of the Fog Pilot system; this parameter appears into the formula that gives the probability of fog from the visibilimeter network.
This probability comes from the single probabilities given from each visibilimeter, and is obtained by a weighted mean. This procedure is repeated for every point on the map. The weight of the probability of every single visibilimeter depends on the distance between instrument and the point.

$$probability(xpo\,\mathrm{int}, ypo\,\mathrm{int}) = \frac{\sum_k probability(vis\_k) \cdot weight(vis\_k)}{\sum_k weight(vis\_k)}$$

The *weight* (*vis_k*) of the measure in the weighted mean decreases, following the function represented in Figure .

Figure 14. In the X-axis the distance between the visibilimeter and the considered point, while in the Y-axis the weight that the information coming from the visibilimeter will have into the weighted mean.

The weight of the information coming from a visibilimeter decreases to a 73% at a distance equivalent to a distance factor, and fall down to a 27% at a distance double than the distance factor.
A cross-verification was performed for the complete dataset, on the final product, varying the distance factor. The results are showed in the following graphs, reporting the ROC Area, the BRIER Score and the Event Separation versus the distance factor.

INDEX	Test period (Distance Factor=10km)	After tuning (Distance Factor=20km)
ROC AREA	0.34	0.355
BRIER SCORE	0.097	0.095
SEPARATION	35	36.5
ECONOMIC VALUE	0.184	0.196

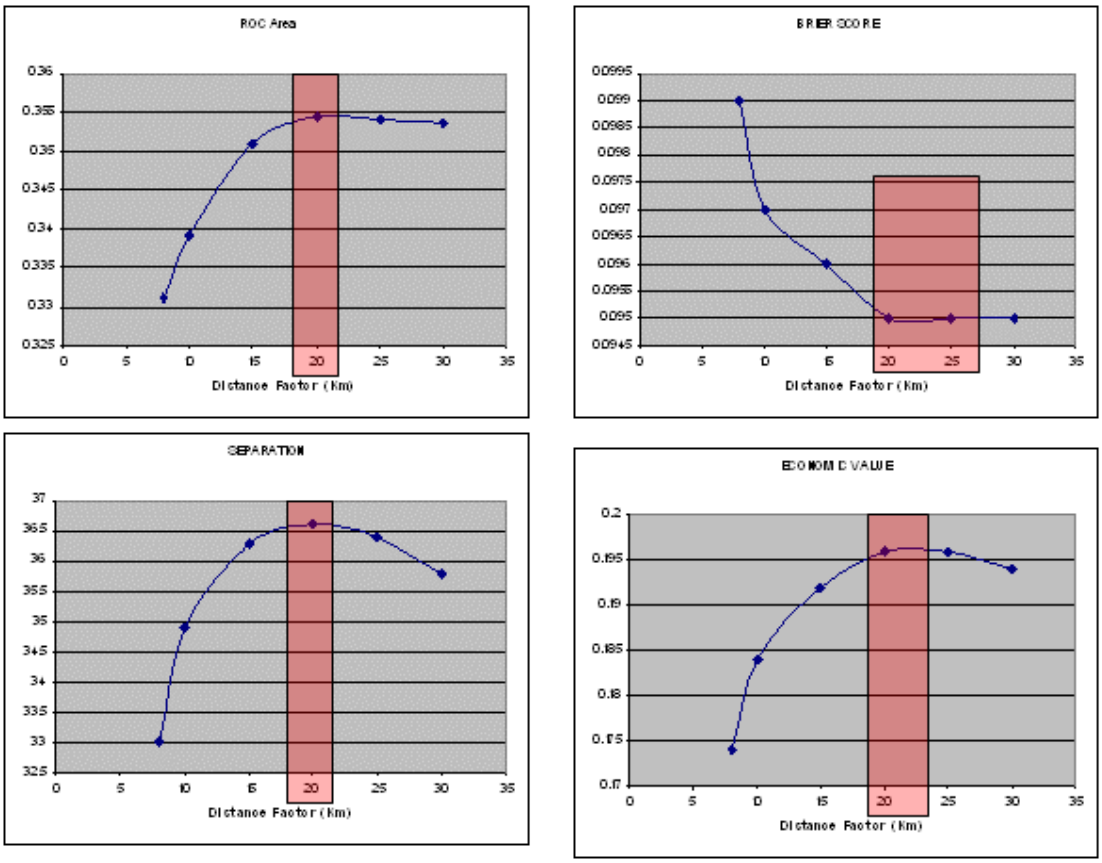


Figure 15. Diagrams of performance indexes for the fog probability field obtained with the Visibilimeters network, varying the Distance Factor; the optimal results are clearly obtained for a Distance factor close to 20 km. This value allows to maximize the ROC Area, the Economic Value and the Separation between Event and NotEvent;, while the Brier Score is minimized.

The results suggest the choice of a distance factor of around 20 km, instead the 10 km used for the first test and initial operational period of the system.
The indexes we were tuning are improved of a 3-4% in comparison to the initial values.

9.2 Tuning the parameters for the correction of SAFNWC-derived probability field

The SAFNWC software reports a product called “Cloud Classification Type”; this product provides a description of the present cloud cover, that for our aims can be divided in the three following sorts: overcast, clear sky, low and very low clouds.
Every point in our map will be associated to one of these three types of cloud cover; starting from the satellite cloud type classification we can assign to every point on the map a probability to have limited visibility at the ground.
Empirically we assigned to the three cloud type class the following probability of fog:
Overcast – 10%
clear sky – 5%
low and very low clouds – 70%.

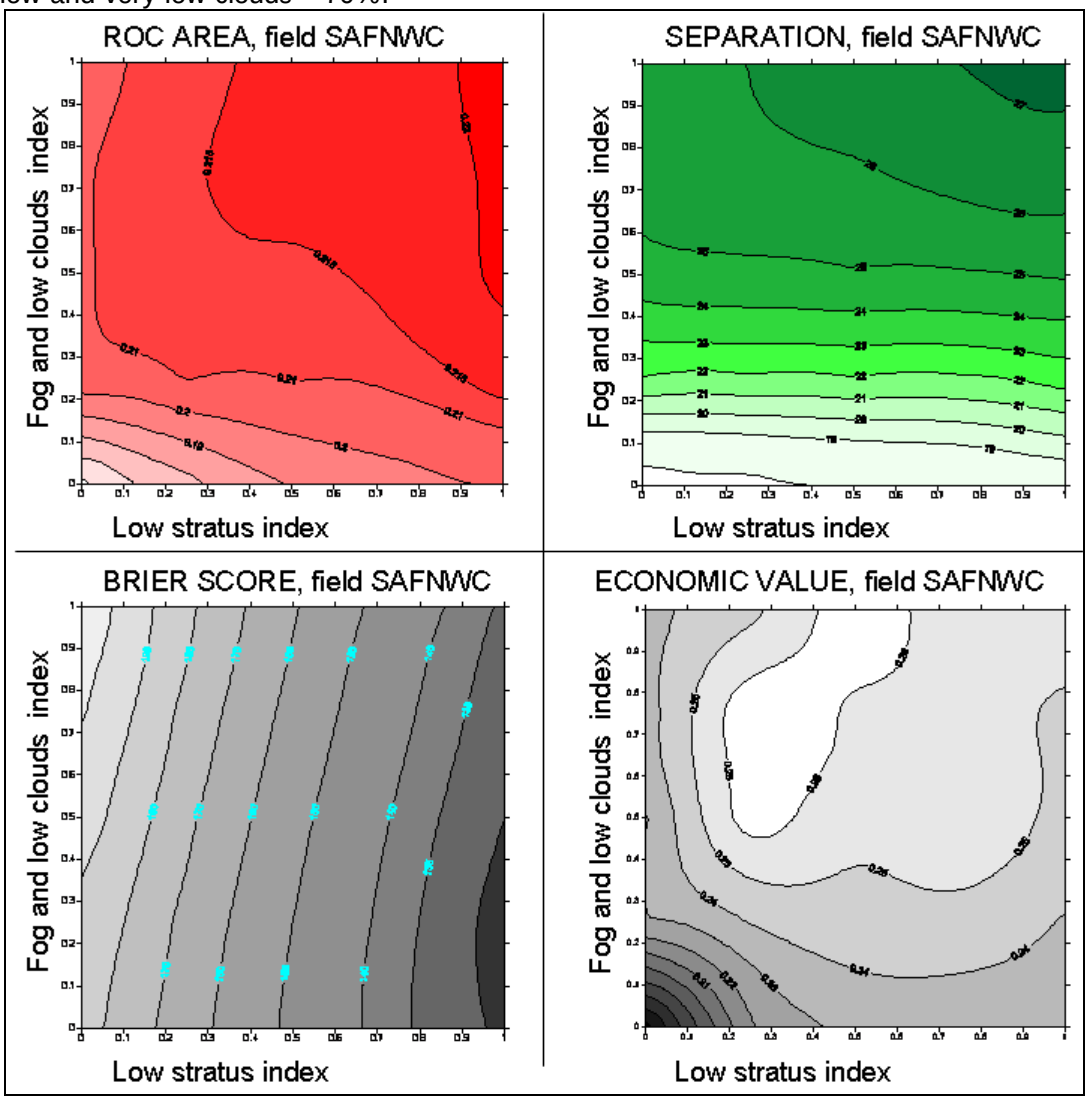


Figure 16. Diagrams of performance for the Satellite-derived field, varying the 2 indexes correcting the Satellite estimation with ground visibility records.

[It must be observed that this assignment of such a probability will require a deeper analysis, once a greater dataset will be available, with at least 12 month of visibility-satellite comparative data - end of summer 2010 -]

Since we know that the information coming from satellite is unsure, we have produced a correction method, comparing the satellite-derived estimation with the ground visibility records.
In the core system of Fog Pilot two indexes are produced, one measuring the coherence between low clouds detection from satellite, and visibility reduction at the ground, and another measuring the opposite condition (incoherence between satellite and visibilimeters).
We will call them respectively *Fog_and_Low_Clouds_Index* and *Low_Stratus_Index*.
These indexes are used to lift or lower the probability and the weight of satellite-derived field.
The performance of the system was tested, varying between 0 and 1 the largeness of the effect of the indexes.
The evaluation of the performance is tested calculating the following four quantities:

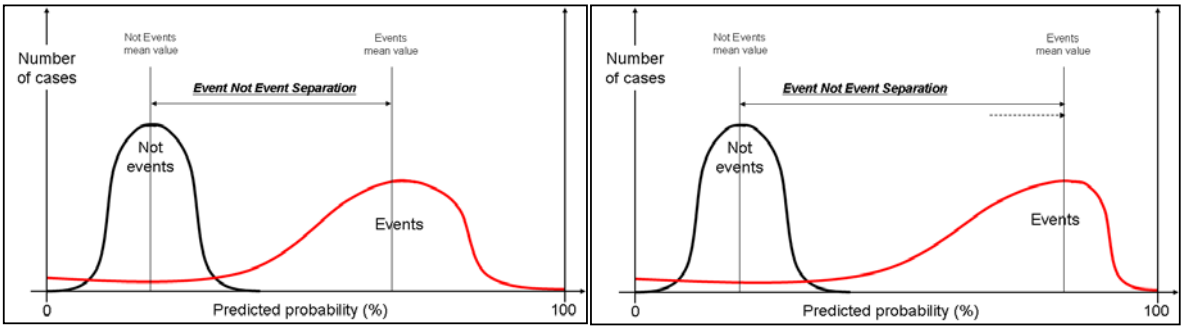


Figure 17. Increasing systematically the forecast probability of events, the Separation of mean values becomes greater, but it's reasonable to expect an overestimation of measures. This explains why the Separation and the Brier Score show a behaviour quite different in the diagrams.

- ROC area (higher is better)
- Brier Score (lower is better)
- Event Not-Event Separation, i.e. the distance between probability mean of events and not-events (higher is better)
- Maximum Economic Value of the warning system, with L/C = 0.2 (lower is better)

Results are presented in the graphs in Figure .

At a first look, is clear that the two correction components have a positive impact on the performances of the system.
It can be noted, in particular from the Brier Score graph, that increasing of the weight of the *Fog_and_Low_Clouds_Index* (index of coherence, lifting the value of the probability) doesn't give a positive contribution to the product, without increasing the weight of the *Low_Stratus_Index*. This can be explained observing the Reliability Diagram, which clearly shows that the system tends to overestimate in case the first index is significantly higher than the second.
By the other hand, the Event Not-Event Separation value takes advantage from the increasing of the effect of the *Fog_and_Low_Clouds_Index*, because increasing this index, higher probabilities are filled; because of that the distribution of events, even if present an overestimation of fog occurrence, is more separated from the not-event distribution (see Figure).

From the results of the tuning here described, we have opted for equal value of 0.75 for both indexes. These values seem to provide an acceptable compromise between the different evaluation methods of the performance.

The following table reports the advantages obtained applying this kind of correction to the crude satellite-derived field of fog probability:

INDEX	No correction (0; 0)	After System Tuning (0.75; 0.75)
ROC AREA	0.16	0.217
BRIER SCORE	0.184	0.142
SEPARATION	16	26.3
ECONOMIC VALUE	0.157	0.258

9.3 Tuning the relative weights for the final product production

As previously described, the merging method requires a weight for each different field that takes part in the process. The fields involved are:

- Satellite-derived estimation
- CART statistical method on CALMET output
- Visibilimeters network

As we can see in the following formula of the weighted mean, used to obtain the merged probability field,

$$P(x,y) = \frac{\sigma_{vis} \cdot w_{vis}(x,y) \cdot P_{vis}(x,y) + \sigma_{sat} \cdot w_{sat}(x,y) \cdot P_{sat}(x,y) + \sigma_{cart} \cdot w_{cart}(x,y) \cdot P_{cart}(x,y)}{\sigma_{vis} \cdot w_{vis}(x,y) + \sigma_{sat} \cdot w_{sat}(x,y) + \sigma_{cart} \cdot w_{cart}(x,y)}$$

every field of probability is weighted, point by point, by a local weight (the *w* values), and with a total weight (the sigma, σ, values), constant for the probability field.
This weight sigma represents the "importance" of the single probability field in the final merging process.
Varying these sigma values for the three principal fields, and running the fog pilot for the entire verification dataset, a map of evaluation index are obtained; these maps allows to select the best set of weight for the three probability fields.
It must be noted that we can divide the upper and the lower member of the formula for the σ_{vis}, and proceed varying the ratio σ_{sat} / σ_{vis} and σ_{cart} / σ_{vis}.
The formula becomes:

$$P(x,y) = \frac{w_{vis}(x,y) \cdot P_{vis}(x,y) + \left[\frac{\sigma_{sat}}{\sigma_{vis}}\right] \cdot w_{sat}(x,y) \cdot P_{sat}(x,y) + \left[\frac{\sigma_{cart}}{\sigma_{vis}}\right] \cdot w_{cart}(x,y) \cdot P_{cart}(x,y)}{w_{vis}(x,y) + \left[\frac{\sigma_{sat}}{\sigma_{vis}}\right] \cdot w_{sat}(x,y) + \left[\frac{\sigma_{cart}}{\sigma_{vis}}\right] \cdot w_{cart}(x,y)}$$

This operation allows a tuning with only 2 parameters, instead the three different weights.
In Figure 17 are reported the results of the tuning of 2 mentioned parameters. Since we taken the weight of visibilimeters field equal to 1, these parameters are equivalent to the CART field and satellite derived (SAFNWC) field weights.

The results obtained with the tuning of weights are the following:

- the impact of the variation of the weights is evident; in particular there are areas in the plots that show significant lost of system performance, while the lower-left side of the graph presents the best results;
- the *Roc Area*, *Brier Score* and *Economic Value* are optimized with weight values roughly between 0.1 and 0.4-0.5;
- the Event NonEvent *Separation* is higher when the CART weight is close to zero; this corresponds to an overestimation of the events;
- considering the previous observations our selected best weights for the system are 0.2 for the CART and 0.4 for the SAFNWC.

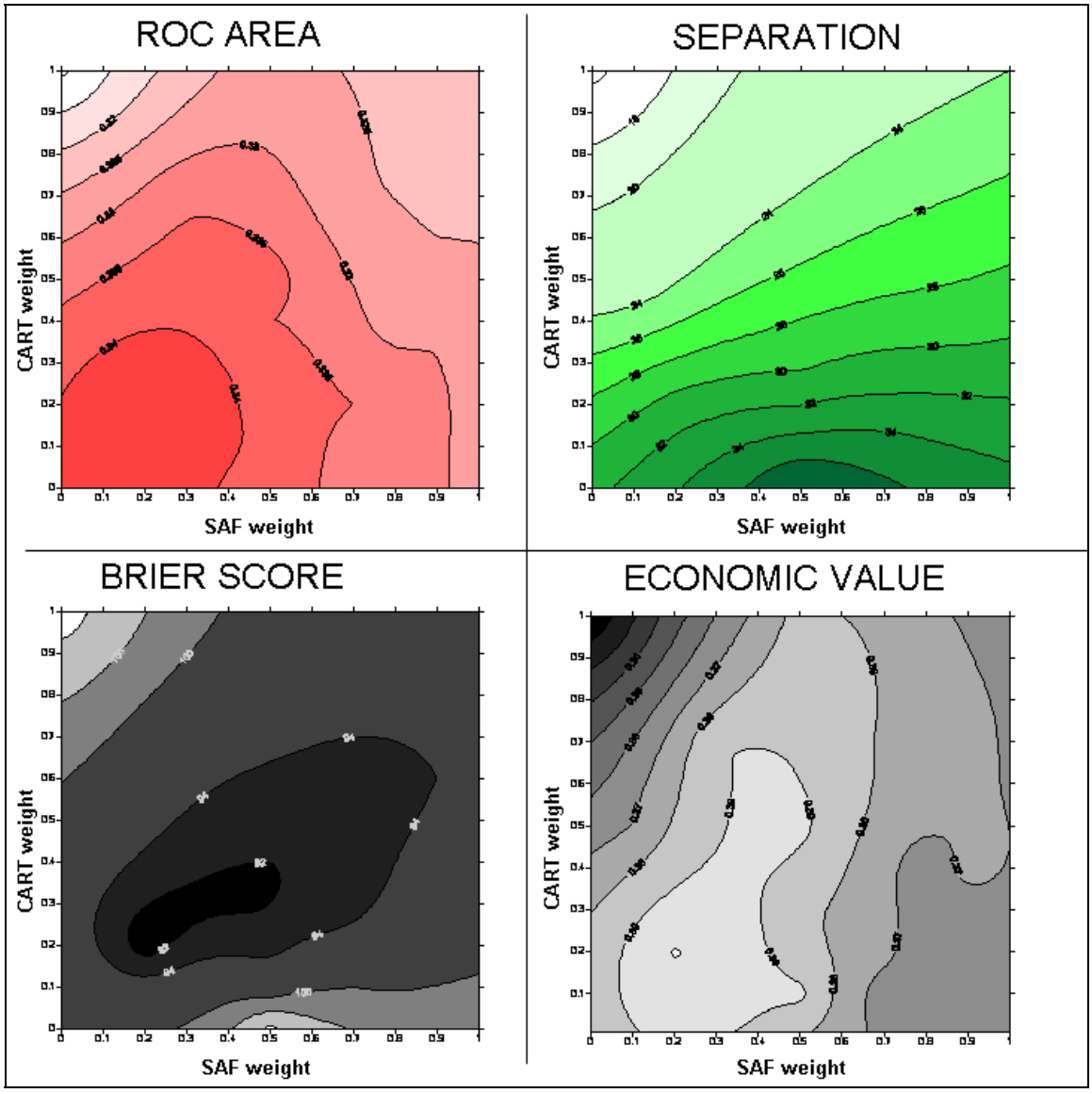


Figure 18. Diagrams of performance for the Final Product of fog probability (Merged field), varying the 2 relative weights (of CART and Satellite-derived fields) used in the weighted mean of the three basic products (CART, Satellite-derived, Visibilimeters Network).

In conclusion, in comparison to the basic setting of the system, with weight equal to (1; 1; 1) we have changed the weights to (1; 0.4; 0.2), improving the performance of the system as it can be seen below:

INDEX	Base Setting(1; 1; 1)	After System Tuning (1; 0.4; 0.2)
ROC AREA	0.32	0.34
BRIER SCORE	0.096	0.093
SEPARATION	24.1	31.0
ECONOMIC VALUE	0.365	0.393

The optimization is appreciable in all the considered scores.
With the triplet of weights obtained with the tuning the reliability of the system (Brier Score) is slightly improved, the discrimination (Separation) between event and non events distribution is higher, and the utility of the forecast (Roc Area and Economic Value) is increased.
New performance analysis will be essential for further improvements the system. The refinement of the system parameters may require years and years of data.

10 Impact of PBL parameters on the Veneto fog pilot

10.1 Introduction and main results

Assessing the impact of variables or parameters describing the state of the planetary boundary layer (PBL) was among the additional activities ARPAV has proposed. It makes sense consider two sets of parameters, which are not independent but are retrieved in a very different way. The vertical temperature structure of the PBL can be observed by the microwave radiometer network operated by ARPAV. The temperature profile gives important information on the stability of the PBL which, in turn, is intimately linked to fog conditions.

On the other hand, there is a set of parameters describing the turbulent state of the PBL. These cannot be directly measured, but need to be estimated by relatively complex models which account for the turbulent dynamics of the PBL. Most notably, the PBL mixing height (Hmix) denotes the height to which the part of the atmosphere which is in contact with the surface can be considered well mixed. This means that quantities like temperature, momentum, humidity, and air pollutants are distributed in the entire PBL by the process of turbulent mixing, which leads to characteristic profiles of temperature and wind. In relation to fog, the idea is that if turbulent mixing and therefore the mixing height is large, conditions for having fog are very unlikely. On the other hand, if the mixing height is very low (weak or absent turbulent mixing), the likelihood for fog conditions increase. The friction velocity (u*) and the inverse of the Monin-Obukhov length (1/L) are used as indicators of the amount of mechanical and convective turbulence. They both are not independent of the mixing height, but respectively reflect the action of wind and radiation on the turbulent mixing in the PBL.

The main results obtained from this analysis are anticipated here and discussed in more detail in the remainder of this section. They can be summarized as follows:

- knowledge of the local vertical profile of temperature (along with the relative humidity) is among the most efficient parameters which identify fog conditions from meteorological surface station network data;

- stability conditions as expressed by Hmix, u*, and 1/L do not pinpoint fog conditions, as a very stable PBL often has no reduced visibility;

- conversely, conditions of instability as expressed especially by 1/L seem to exclude fog conditions quite efficiently;

- subjective thresholds for these parameters were identified as Hmix < 2-300m, u* < 2 m/s, and -100m < L < 200m.

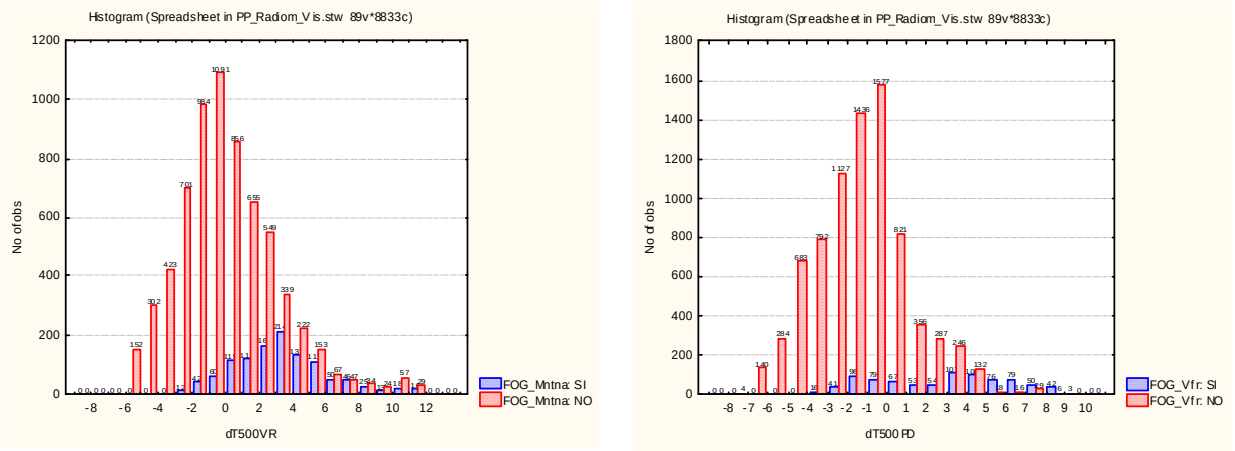


Figure 19. Fog event (blue) and non-event (red) frequency distributions as a function of the vertical temperature gradient (K/500m) for the station Montagnana and the 14km distant MWR Legnago (left panel), and for station Villafranca Veronese and the 88km distant MWR Padova (right panel).

10.2 PBL Temperature profiles

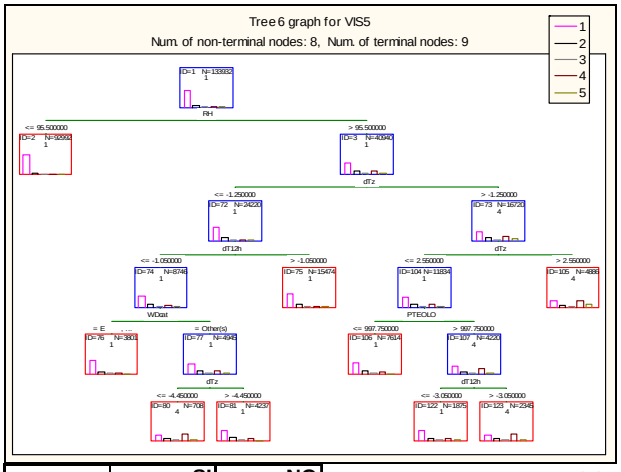
10.2.1 Approach

The impact of the PBL temperature profile, or vertical temperature gradient, on the identification of fog conditions is performed by the non-parametric classification and regression tree analysis (CART), as used for extracting fog probabilities in the Fog Pilot model and described in D3.4. There, the vertical temperature gradient was extracted from the two relatively close by surface stations Legnaro and Mt. Grande, which are located at about 10 and 470 m above sea level, respectively. Such a profile can be considered local for the area of Padova and surroundings, but less so for Verona, for instance. The value of this information as assigned by the CART analysis comes from the fact that positive temperature differences between the upper and lower level indicate temperature inversions and therefore stable to very stable PBL conditions. Together with relative humidity close to saturation such conditions are necessary, but not sufficient, ingredients for fog formation.

Figure illustrates the value of having local profile information, in that a relatively well peaked probability distribution for fog emerges from the CART for surface station Montagnana using the microwave radiometer (MWR) in Legnago, where the two locations are located about 14 km apart. As a matter of fact, fog occurrence is predominant for temperature differences between the upper (500 m) and the lower (surface) level greater or equal zero, i.e. stable conditions. The peak is between 2-4°C. If, on the other hand, the CART for surface station Villafranca Veronese using the microwave radiometer in Legnago (88 km distance), the probability distribution for fog is much flatter and indicate fog conditions on a much broader range of temperature gradients. In particular, the number of fog events related to conditions with no thermal inversions is about doubled.

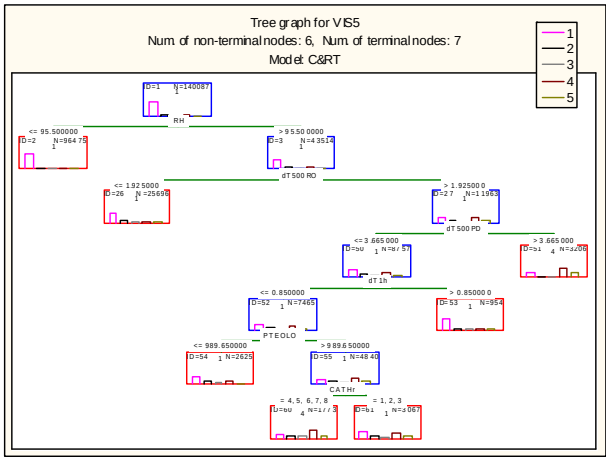
The differences in the actual CART tree are shown in Figure , along with the contingency tables for fog yes or no, based on the visibility threshold of 500 m, and a number of

scores. The version of the CART analysis with the more localized temperature profile information (derived from the MWR) have a higher CSI, POD and Accuracy, while the FAR, POFD, and BIAS are smaller. Both sets of scores differ quite significantly. The main difference of the two classifiers seems to be the massive reduction of false alarms (from 4128 to 1616) and misses (from 5392 to 3193).



	SI	NO
SI	3811	4128
NO	5392	120601

Accuracy	92.9
CSI	0.29
POD	0.41
FAR	0.52
POFD	0.03
BIAS	0.86



	SI	NO
SI	3363	1616
NO	3193	122431

Accuracy	96.3
CSI	0.41
POD	0.51
FAR	0.32
POFD	0.01
BIAS	0.76

Figure 20. Optimal tree based on the lowest cross-validated relative error: ground parameters and temperature difference from surface stations Mt. Grande (470masl) and Legnaro (10masl) (left panel), and temperature differences between levels 500m and surface from MWR Padova and Rovigo (right panel). Contingency tables and a number of scores are evaluated on the basis of the visibility threshold of 500m for fog conditions (CSI: critical success index, POD: probability of detection, FAR: false alarm ratio, POFD: probability of false detection, BIAS: bias).

The CART analysis yields the relative importance of the predictors used in the classification process. This means that CART will tell which variable are most useful for distinguishing, in our case, fog from non-fog, and which ones are of very little value. Figure shows this ranking for the two classifiers discussed in this section. In case of one single information on the vertical temperature gradient (Figure , right panel) the relative humidity emerges to be by far the most important distinguishing variable, followed by the vertical temperature gradient. In case of the two MWRs the ranking changes, in that now the vertical temperature gradient information of the Padova MWR appears to the most important selector, followed by the one in Rovigo and the pressure information, relative humidity and temperature tendency.

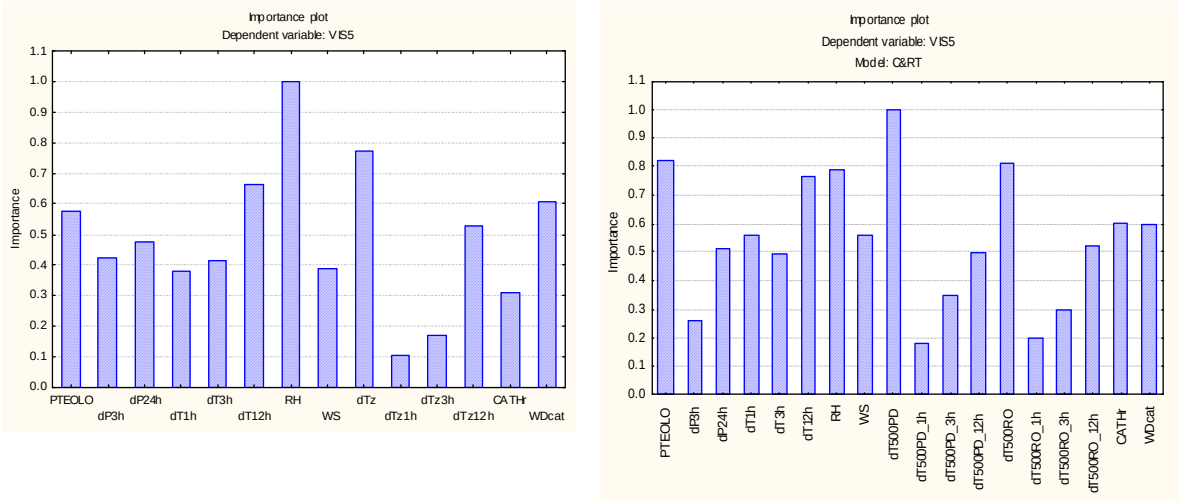


Figure 21. Relative importance of parameters in detecting visibility values below 500m as output by the CART analysis when respectively including the one pseudo temperature profile obtained from the surface stations at Mt. Grande (470masl) and Legnaro (10masl, left panel), and the temperature differences between levels 500m and surface from MWR Padova and Rovigo (right panel).

In conclusion, the more local information on the vertical temperature gradient seems to give an additional value to the fog – non-fog classification done by CART on the traditional meteorological observations. This highlights the potential of having PBL temperature profilers. In case no such devices are available, a set of relatively close by stations with an altitude difference of the order of 500m can provide useful information.

10.3 Turbulent state of the PBL

10.3.1 Approach

The remainder of this section is devoted to illustrating and discussing these main findings. The approach chosen to assess the impact of observations related to the turbulent state of the PBL is as follows:

- calculation of the PBL parameters Hmix, u*, and 1/L;
- visual inspection of the co-variance of the PBL parameters and visibility for a number of fog episodes and two stations;
- basic statistical description of the PBL parameters and visibility for the winter months (December, January, February) 2008-09 and 2009-10;
- subjective assessment of the potential use of the PBL parameters in the fog pilot model.

10.3.2 Results

Visual inspection of the co-variance of visibility, Hmix, u*, 1/L, relative humidity (RH), and temperature (T) give a clear indication that there is a close relation between the state of the PBL and conditions which are conducive to fog formation. Note that the extra horizontal line on plots denotes the visibility = 500 m threshold for indicating fog. The following findings emerge from the four fog periods spanning a total of 9 days:

- Values of Hmix above about 300 m are mostly related to good visibility; when 1/L is negative and not close to zero, visibility appears to be far from the 500 m threshold; the impact of u* on Hmix seems to be less effective on fog dispersion;

- Low values of Hmix and u*, and values of 1/L indicating atmospheric stability are not sufficient conditions for visibility to be reduced, even with high RH and low T;

- In evident fog conditions, short episodes of mixing, especially by thermally induced turbulence (1/L << 0), re-establishment of fog is lagging the stabilization of the PBL by several hours, i.e. fog formation appears as a slow process.

These findings are based on a limited number of cases as seen on two stations. They are, however, more or less confirmed by the basic statistical analysis given in the scatterplots in Figures 22, 23 and 24. All of the three inquired PBL parameters exhibit a scatterplot with a characteristic triangular shape which is very broad in terms of observed visibility values for stable to very stable PBL conditions, and becomes narrow towards very high visibility values as instability conditions increase. This confirms that the indication of stability conditions does not pinpoint fog conditions, but, conversely, conditions of instability seem to exclude fog conditions. For Hmix a good subjective threshold could be picked at 2-300 m, for u* at about 0.2 m/s, and for -100 m < L < 200 m.

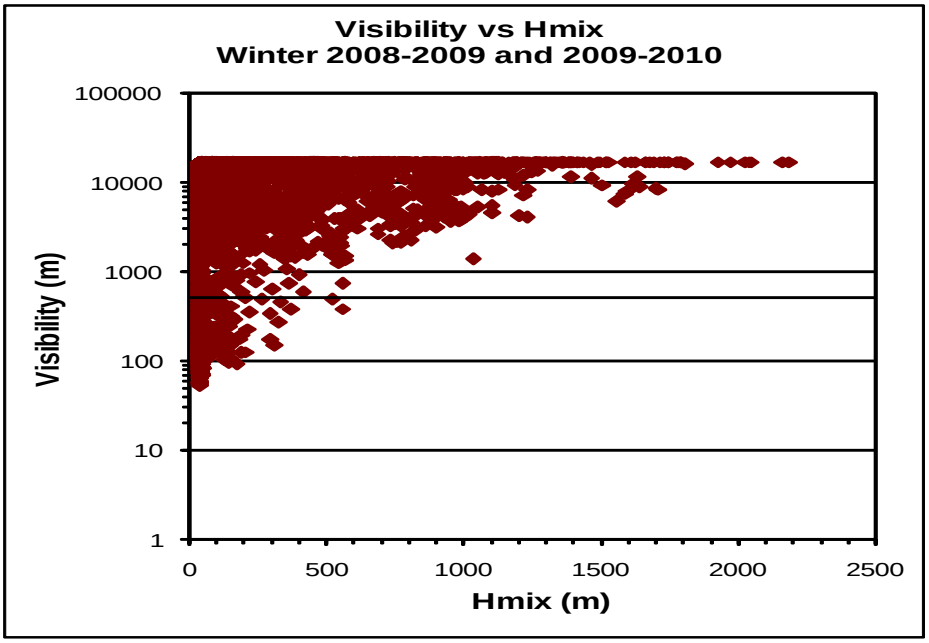


Figure 22. Visibility (log10(m)) versus Hmix (m) for the winter periods (DJF) 2008-09 and 2009-10.

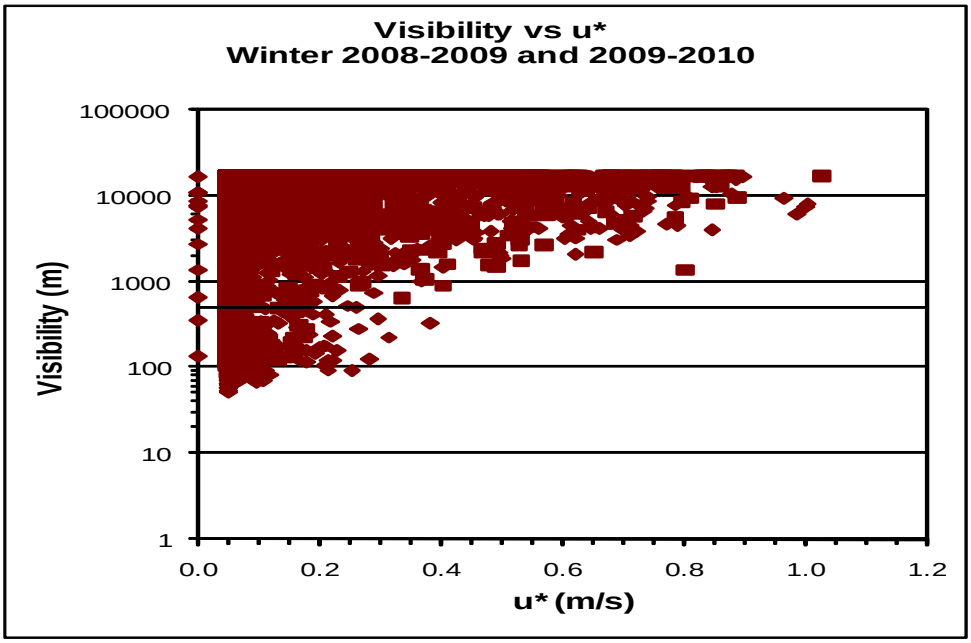


Figure 23. As in Figure 22 but for surface friction velocity u (m/s).*

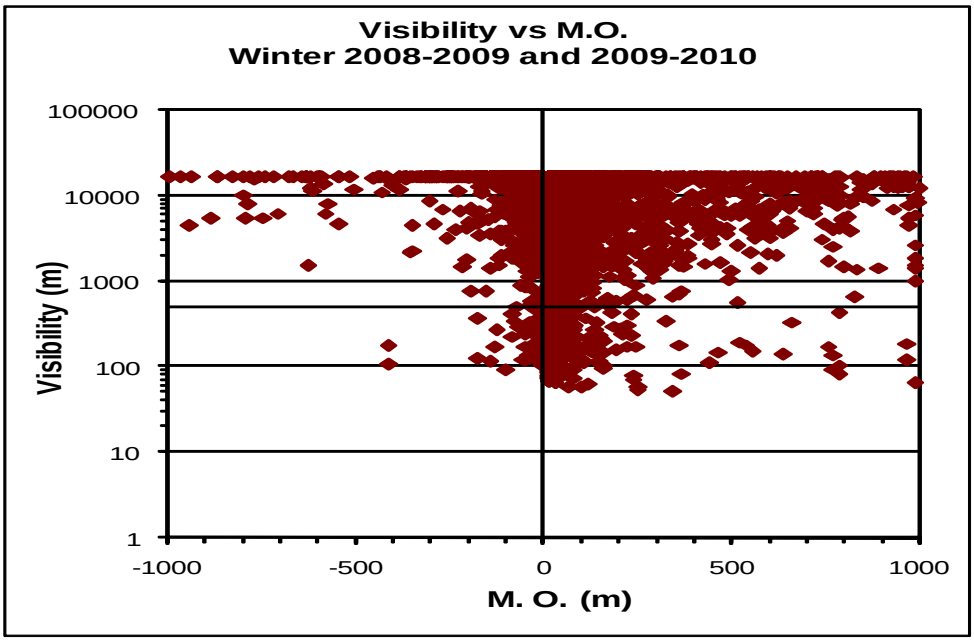


Figure 24. As in Figure 22 but for Monin-Obukhov length L (m).

10.3.3 Potential of application in the fog pilot

A conceptually straight-forward benefit of these findings consists in applying such thresholds in the context of a decision tree, especially where there is no direct observation of visibility. This points to the satellite-derived fog probability map, for which probability values for fog (visibility < 500 m) could be substantially reduced if the PBL parameters indicated instability conditions. A more delicate decision would have to be made if such conditions were analyzed where direct visibility measurements indicated fog. Full implementation of the PBL parameters into the Fog Pilot was not possible in the timeframe of this deliverable, but will be pursued.

11 Open fog pilot architecture: processing of additional information

11.1 Introduction

The Fog Pilot system in its current version real-time processes data retrieved from a dedicated visibilimeter network of 10 stations, the cloud classification scheme of the Meteosat Nowcasting Satellite Application Facility (SAFNWC), and a conventional meteorological observations of 30 stations. As laid out in deliverable D3.4 the data are of each source are transformed into a probability of fog occurrence based on a visibility threshold (500 m in our case) and assigned a data quality. Then the three fields are merged taking the relative quality into account.

Such an architecture can, in principle, accommodate additional data sources which can be transformed into fog probabilities and whose quality compared in relative terms to the other data sources. The Fog Pilot system architecture can, therefore, be considered an open architecture, and able to accommodate a potentially significant number of additional and novel information types. As such it very well fits into the Roadidea data exploitation philosophy. A list of such potential information sources includes:

- Visibility estimates from webcam networks;
- Occasional reports from visibilimeters outside the permanent network;
- Direct human visibility reports;
- Automatic direct and indirect reports from intelligent cars (e.g. direct measurements, indirect information from fog light activation or reduced speed).

The extra work required for each of these categories to be included into the current Fog Pilot structure consists in:

- webcam networks: devise automatic visibility estimates via suitable image processing, and determine relative weight w.r.t. direct visibility measurements from reference network;

- Occasional reports other visibilimeters: establish confidence of observation quality;

- Direct human visibility reports: potentially large number of reports of heterogeneous quality requires robust statistical pre-processing to establish actual information content;

- Automatic direct and indirect reports from intelligent cars: as for human reports.

It is beyond the scope of the Fog Pilot to include such information in a systematic manner. It is, however, the scope of Roadidea to pinpoint non-conventional, potentially valuable data sources which can benefit road transport on the European Scale. The Fog Pilot was designed to do this w.r.t. direct and indirect visibility information, potentially available along the major road transport routes throughout the continent. In this sense, the Fog Pilot would have the potential to become a community tool, which would allow, for example, a number of (professional) users to feed data into the system and benefit from the compositing done by the system.

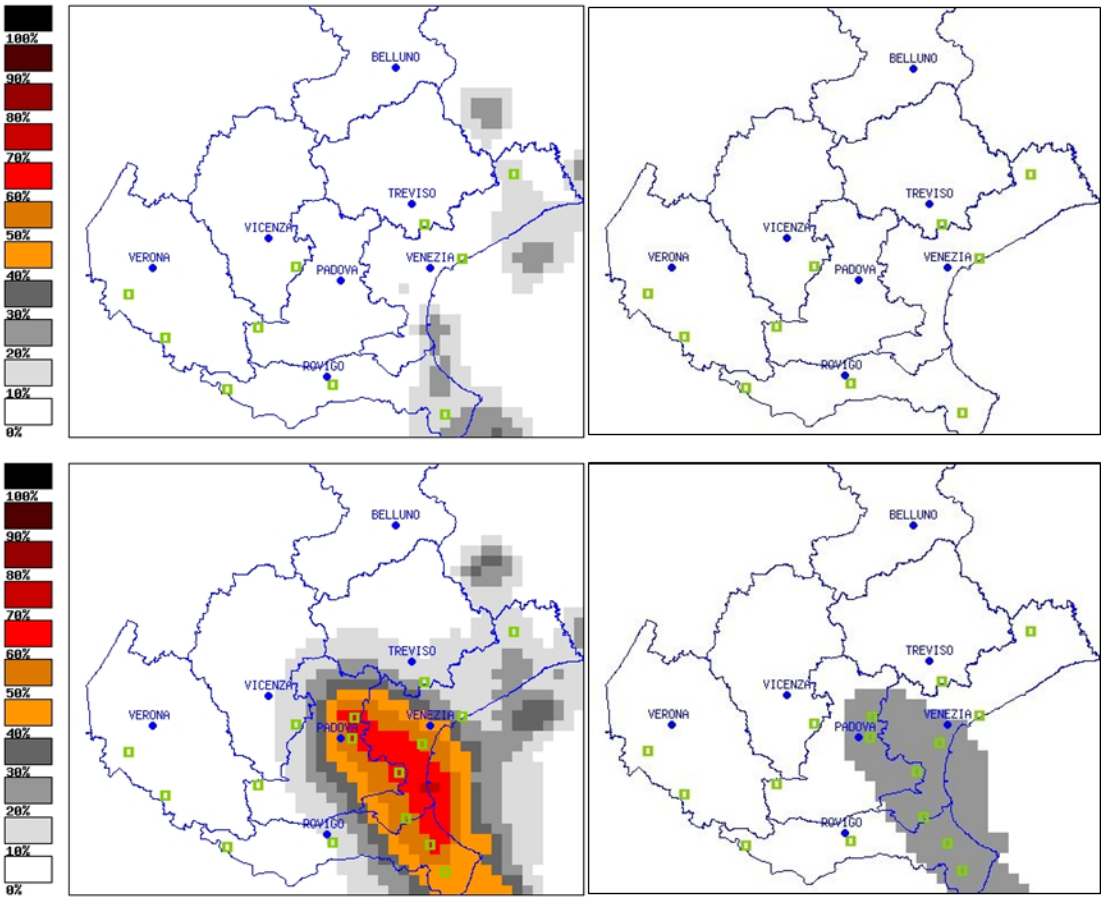


Figure 25. Fog probability (left column) and alert (right column) maps for the Veneto Fog Pilot reference system (upper row), and the system extended with additional visibility information (lower row). Political borders of the Provinces of the Region Veneto are drawn in blue, the points in which visibility observations are available are drawn as green squares.

The remainder of this section is dedicated to illustrating how extra information can benefit the performance of the Fog Pilot on the regional scale of Veneto. Two scenarios were selected:

- sparse fog which was non detected by the reference system, but additional information did;

widespread fog in which the scale of the reference system did not pick up smaller-scale variation of the phenomenon

11.2 Fog not observed by the reference system

The reference visibilimeter network currently features an effective horizontal resolution of some 30 km. For the potentially highly variable phenomenon fog such a resolution can be insufficient even to pick up relatively large patches of fog. Figure shows the fog

probability map (upper left panel) and fog alert map (upper right panel) of the reference system. The resulting low fog probabilities in the coastal regions do not trigger a fog alert (e.g. based on a 40% fog probability threshold).

Additional direct visibility reports in the coastal southern regions detect probabilities for reduced visibility conditions in the range of 70%. This signal is now sufficiently strong to trigger a fog alert in the south-eastern part of Veneto.

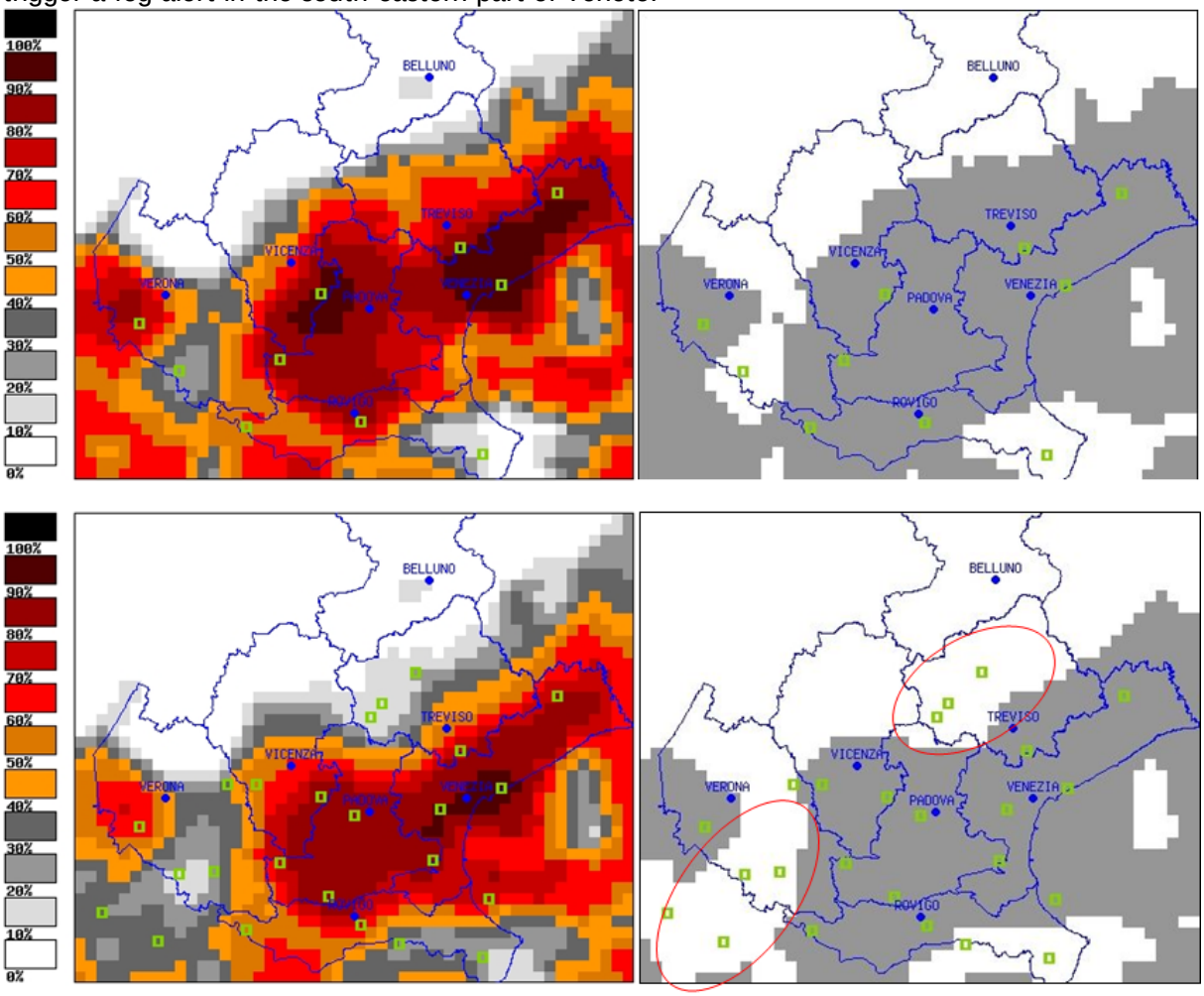


Figure 26: as for Figure .

11.3 Add small-scale information in a widespread fog situation

The second example in which additional visibility observations can be valuable is illustrated in Figure . The upper left panel shows widespread high values of fog probabilities and the corresponding widespread fog alert (upper right panel). In particular, the entire area bordering the visibilimeter network to the north is included in the fog warning.

Additional visibility information becoming available in the northern and south-western part of the visibilimeter network changes the picture by reducing the fog probabilities below the alert threshold. Now the fog warnings along the northern border is better delineated, as well as in the south-west, where the additional reports reveal an area with significantly lower fog probabilities.

11.4 Open issues

As stated in the introduction of this section, inclusion of non-conventional data sources into the concrete Fog Pilot is beyond the scope of Roadidea, highlighting the potential of such data sources, however, is at its heart. The Fog Pilot has, in principle, characteristics that would allow it to become a community tool. Such a tool would give a benefit which is increasing with an increasing number of contributors (road traffic community). However, in order to include an unknown number of pieces of information from unknown sources would require paying close attention to at least the following issues:

- Robust statistical pre-processing, especially in view of plausibility;
- Treatment of incoherent data in the same area;
- Attribution of a quality, or weight, to unknown data relative to the reference network;
- Handle differing data density, i.e. how to handle areas with few or many data points.

This list is certainly not complete and only points towards a few theoretical issues in handling and processing information, which would need additional attention.