15th International Road Weather Conference
Quebec City, Canada, 5-7 February 2010

ROADIDEA **Session**

Evaluation of FMI's new forecast model of road surface friction

Pertti Nurmi
Marjo Hippi
Ilkka Juga

Finnish Meteorological Institute (www.fmi.fi)
Meteorological Research Applications
e-mail: pertti.nurmi@fmi.fi

Abstract

The Finnish Meteorological Institute (FMI) is in the process of developing a physical-statistical forecast model of road surface friction (see, preceding paper by Hippi et al (2010)). The model is based on correlation analysis between observed road surface friction as measured by Vaisala DSC111 optical sensors and other, more conventional, road weather observations. The statistical relationship was derived utilizing measurements from winter 2007-08, and then validated using independent data from the following winter, 2008-09. The initial results showed high correlations between the observed and modeled friction under both icy/snowy and wet/damp road conditions. The resulting statistical relationship is linked to FMI's operational road weather forecast model, which takes its input from a physical numerical weather prediction model. Friction is thus a new forecast parameter resulting from a numerical forecast model. It has up till now been interpreted from the more conventional weather parameters by professional end users. This totally new forecast product is under extensive analysis, testing and validation. The model is in a test phase in the operational forecasting environment of FMI during winter 2009-10. The forecasts are targeted at a number (c. 100) of road weather stations in Finland, operated by the Finnish Road Administration, which are equipped with optical sensors facilitating comprehensive verification. The modeled friction data are spooled into categories of explicit, pre-defined values exceeding given threshold relevant for different categorized slippery conditions of the road surface. Some new, emerging meteorological forecast verification techniques like a modified Extreme Dependency Score (EDS) and the Symmetric Extreme Dependency Score (SEDS) are adopted to complement the more conventional verification metrics. In addition to providing feedback on the capabilities of the statistical friction model, this study will provide valuable information on the behavior of these new verification measures. The SIRWEC Conference 2010 is the first forum to showcase these early results.

This study is associated with the EU/FP7 Project ROADIDEA and the EU/COST Action TU0702. The major goals of these undertakings are to study the adverse effects of weather on traffic and to develop new and innovative methods and tools to increase traffic fluency and safety.

1. Introduction and scope of the study

Slipperiness caused by freezing of the road surface is probably the most notorious phenomenon affecting road traffic and transport of wintertime Finland. The most slippery situations are experienced when having water or snow above an ice layer covering the road. Then the friction, i.e. the grip between car tires and the underlying road surface reaches its minimum. Friction, by definition, is the force resisting the relative lateral motion of solid surfaces, fluid layers, or material elements in contact. In our context, friction is defined as a coefficient ranging from 0.1 to 0.8. The general driving conditions, and the so-called road weather index, have been defined by the Finnish Road Administration (Finnra) in three categories based on surface friction: (i) very bad, (ii) bad, and (iii) normal road weather. The corresponding friction thresholds are 0.15 and 0.3, respectively. There is also a more detailed slipperiness classification within this broader categorization (see Figure 1). The road weather index is the basis for providing warnings against adverse driving conditions to the general public in the whole of Finland.

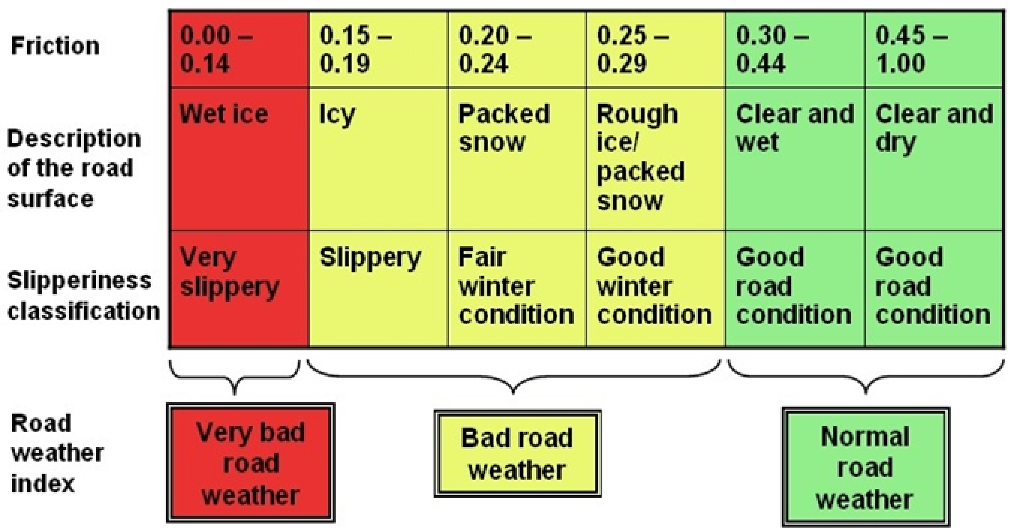

Figure 1. Road weather classification in Finland.

To further emphasize the importance of the notion of both present and expected road surface friction conditions it should be pointed out that friction criteria are used by Finnra when contracting out their winter maintenance services and activities. Quality demands on road maintenance activities require specified target friction levels to be maintained along the Finnish road network, and the realization of these targets is monitored closely (Pilli-Sihvola, 2008). Consequently, the road network needs to be well maintained to allow for safe driving under various weather conditions. On the other hand, unnecessary or superfluous activities like erroneous or extraneous salting, are extremely costly.

All this means that the observed state of the road network should be constantly, efficiently and reliably monitored. Moreover, producing detailed and reliable forecasts depicting the future state of road weather conditions is of utmost importance. Under test conditions friction can be evaluated fairly straightforward mechanically by braking tests utilizing suitably equipped cars. This approach would not be easily applicable in a real-time monitoring (or forecasting) environment. However, there are available relatively new remote measuring techniques for road surface state sensing. One of these is the Vaisala DSC111 equipment, which is based on laser technology and, adopting a spectroscopic (optical) measuring principle identifies, individually, the presence of water, ice, slush, snow/frost on the underlying surface. This sensor provides also a continuous estimation of surface friction (Vaisala, 2005a). DST111 is a complementary sensor which measures air, dew point and road surface temperatures (Vaisala, 2005b). The Finnish road network boasts well over 500 road weather stations operated by the Finnra. More than 100 of these stations are equipped with DSC111 and DST111 installations. Data from a selection of these stations are utilized in the friction modeling and validation efforts presented in this study.

FMI operates, since 2000, a dedicated fine-scale road weather forecasting model which is driven by operational larger scale atmospheric numerical weather prediction (NWP) systems. It is a one-dimensional energy balance model which estimates the vertical heat transfer within the ground and at the ground-atmosphere interface, taking also into account the prevailing traffic. Forcing from the upper boundary of the domain is driven by input taken from forecasts of the NWP model(s). The road weather model does not produce as direct model output a surface friction parameter per se. This means that the friction, or slipperiness, has up till now been interpreted from other model forecast parameters.

The basic concept and goal of this study and the one presented by Hippi et al (2010) are to first try to formulate a statistical relationship between the friction as observed by the Vaisala DSC111 instrument and parameters believed to be relevant in estimating slipperiness: water, ice and snow/frost content on the surface as well as road surface temperature. If such a relationship can be defined and validated to have sufficient quality, this statistical model is expected to be applicable for estimating prevailing friction/ slipperiness conditions more generally. The follow-up challenge would then be to utilize the statistical model in a forecasting mode by using output from the road weather forecast model as input parameters in the statistical relationship. Hence, the approach can be considered a perfect prog application, i.e. encompassing a non-trivial assumption that the driving forecast parameters originating both from the NWP and from the road weather model were perfect. This is definitely a non-valid assumption, but depending on model verification outcome the potential usefulness of the eventual explicit friction forecasts can be analyzed.

This study focuses on the first entity as defined above, i.e. to the validation of the statistical relationship for friction. Prospects for practical forecasting are briefly elaborated in the latter part of this paper.

2. The statistical model

The statistical method to depict the relationship between friction and thickness of ice, snow and water content on road surface and road surface temperature, as observed by the DSC111 and DST111 sensors, was introduced in the accompanying paper by Hippi et al. (2010). This statistical association was estimated adapting regression analysis at selected road weather stations by utilizing observation data from winter 2007-08. Also other observed weather parameters were considered and tested but their significance turned out negligible during the regression analysis. The best fit model(s) resulted in three mutually exclusive equations, separately for cases when there is (i) ice and/or snow on the surface, (ii) liquid water on the surface, (iii) dry and clear road surface. The respective regression equations are of the form:

$$F_1 = a_1 * f(IS) + b_1 * T_{rs} + c_1 \qquad \text{(i)} \qquad \text{ice/snow}$$

$$F_2 = a_2 * f(W) + c_2 \qquad \text{(ii)} \qquad \text{water}$$

$$F_3 = 0.82 \qquad \text{(iii)} \qquad \text{dry}$$

where a_i b_i c_i = regression parameters
T_{rs} = road surface temperature
IS = thickness of ice/snow [in mm]
W = water content [in mm]

The terms f (IS) and f (W) indicate a function, rather than linear fit, of the thickness parameter. In cases of dry road surfaces the friction value is set explicitly to 0.82, which corresponds to the maximum measurable value of the DSC111 sensor. The minimum value is set to a discrete value of 0.1 even if the regression formulae can produce lower values.

As an example of the usage of the regression equations, Figure 2 (left) shows the correspondence between observed friction (y-axis) and the thickness of ice/snow on the road surface (x-axis) at a given road weather station (Anjala) based on winter 2007-2008 observations (dependent dataset). Figure 2 (right), accordingly, shows the realization of the ice/snow Equation (i) when applied to the independent dataset of the following winter, 2008-09 at this station. Similar distributions are shown for another station, Utti, in Figure 3. The regression parameters a, b, c differ typically somewhat from station to station. Nevertheless, there is a reasonable fit - correlations are of the order of 0.9 in both of the datasets.

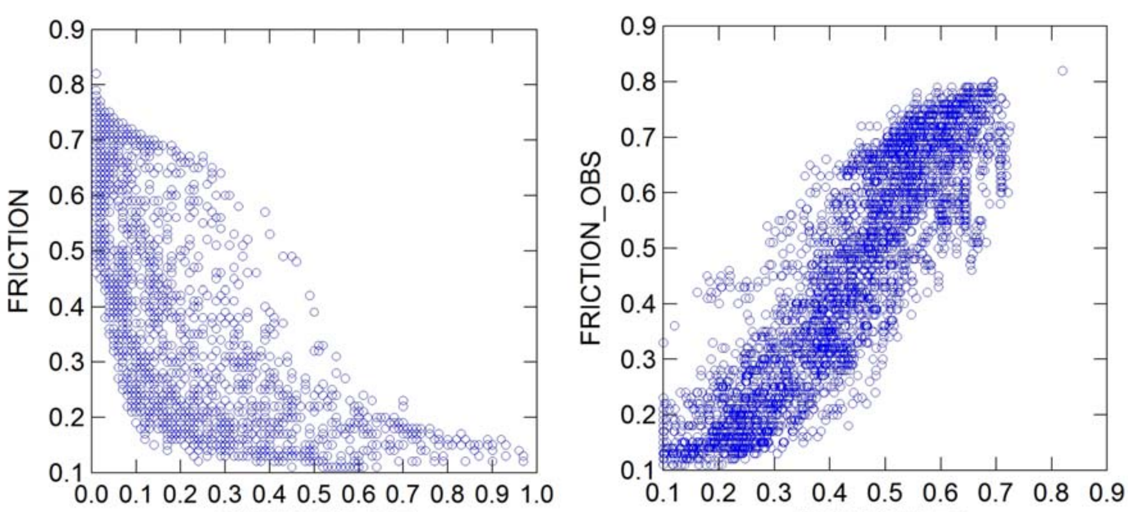

Figure 2. Dependence between observed friction and thickness of ice/snow layer on road surface at station Anjala during winter 2007-08 (left), and correspondence between observed and modeled friction based on winter data 2007-08 by applying Equation (i) on the independent dataset from winter 2008-09 (right).

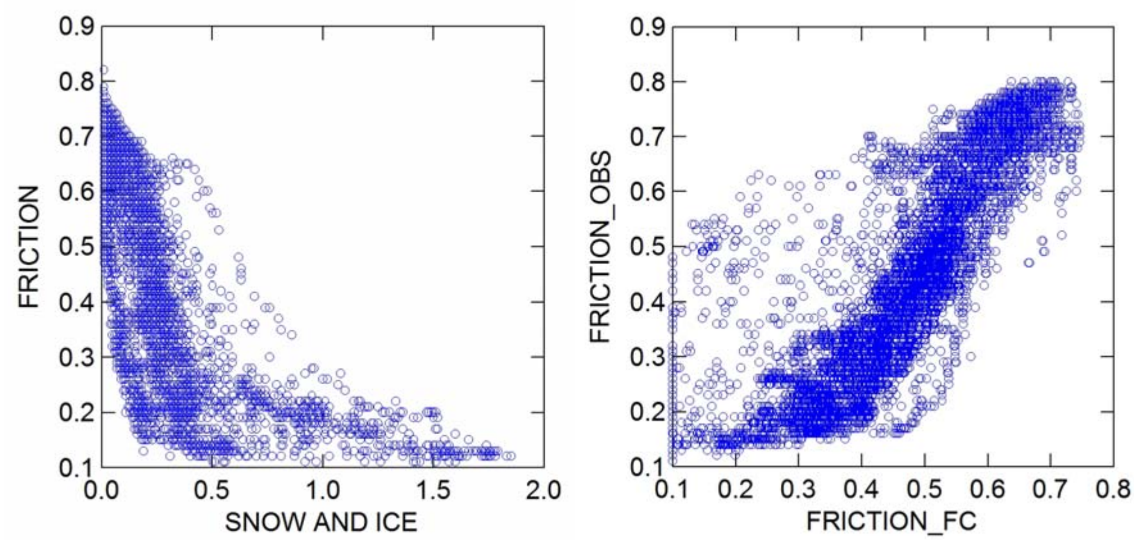

Figure 3. As in Figure 2 but for another station, Utti.

3. Verification issues

Forecast verification is an active, dynamic and rapidly evolving area of research. There are presently more meteorological forecast models available than ever, boasting increasing complexity and continuously enhanced scale resolution. Also new types of observations to evaluate forecasts are becoming available. Before reaching a mature enough stage to be applied in quality assessment such observation means need to be thoroughly analyzed and evaluated. Optical friction measurements obviously fall into this novelty category. This study will introduce and address features and properties of some of the "traditional" forecast verification measures along with decidedly new metrics. The latter will need thorough further investigations, but may turn out fitting for applications like the one presented here. Therefore, the present study is an appropriate test bench for these new verification tools, in addition to providing, hopefully, useful information on the capabilities of the derived statistical friction model.

The root mean square error (RMSE) is the most commonly used measure of accuracy, and the correlation coefficient for evaluating a match between forecast and observed continuous distributions. Both of them value equally an event being forecast, but not observed, as they do for an event being observed but not forecast. In reality, however, there can be an enormous difference, e.g. if slippery conditions occur without having been predicted as opposed to making false forecasts of slipperiness. Consequently, we concentrate here on explicit events, defined by specified friction thresholds, and whether they have been exceeded or not - we are dealing with "yes/no" statements referring to pre-defined thresholds as detailed in Figure 1. Categorical statistics are needed to evaluate such binary, or dichotomous, events. The first step in verifying binary events is to compile a 2*2 contingency table indicating the frequency of "yes" and "no" forecasts and corresponding observations (see below):

Event forecast	Event observed		
	Yes	No	Marginal total
Yes	a	b	a + b
No	c	d	c + d
Marginal total	a + c	b + d	a + b + c + d = n

The seemingly simple definition of a dichotomous event and the related contingency table embraces surprising complexity, and there are numerous established measures to tackle the associated binary verification issue. Some of the most common "traditional" verification measures (e.g. Nurmi, 2003) can be written simply as:

$PC = (a + d) / n$ (Proportion Correct)

$POD = H = a / (a + c)$ (Probability Of Detection, or Hit Rate)

$FAR = b / (a + b)$ (False Alarm Ratio)

$F = b / (b + d)$ (False alarm rate [note the difference to above] also called Probability of False Detection)

$KSS = POD - F$ (Hanssen-Kuipers Skill Score; also called True Skill Statistics or Peirce Skill Score)

$TS = a / (a + b + c)$ (Threat Score; also called Critical Success Index)

$ETS = (a - a_r) / (a + b + c - a_r)$ where $a_r = (a + b)(a + c) / n$ (Equitable Threat Score; also called Gilbert's Skill Score)

$HSS = 2 (ad - bc) / \{ (a + c)(c + d) + (a + b)(b + d) \}$ (Heidke Skill Score)

All of these scores have been derived many decades ago and they have been used very extensively in meteorological forecast verification. However, all of them have their pitfalls and severe shortcomings. It is out of the scope of this study to go into any details, suffice to say that they are at their worst and quite often useless in the quality assessment of forecasts of rare meteorological events having a low probability of occurrence (i.e. a low base rate).

There are a couple very recent measures for the verification of binary events, with a specific purpose to address the difficulty in properly taking into account the scarceness of the event under evaluation:

$$EDS' = \frac{\log F - \log H}{\log F + \log H} \qquad SEDS = \frac{\log [(a+b)/n] + \log [(a+c)/n)]}{\log (a/n)} - 1$$

(Chris Ferro, pers.comm.; Hogan et al, 2009). Both of them are modifications of the so-called Extreme Dependency Score (EDS) (Stephenson et al., 2008) by aiming at adjusting some of the potentially undesirable properties of the original EDS metrics.

These new measures together with some of the more traditional ones were used to test the behaviour of the statistical friction model introduced in the previous chapter. Figure 4 shows verification statistics at the same two road weather stations as presented before, Anjala (left), and Utti (right). Results based on five verification measures, TS, ETS, HSS, EDS' and SEDS, are shown as function of different friction thresholds (recall Figure 1). In all cases, the verification measures follow the same qualitative structure. Highest scores are reached with the novelty measures, EDS' and SEDS. TS and ETS are lowest and go aligned with each other as expected, and HSS lies inbetween. What is quite interesting is the behavior of all scores when we go to the more adverse, slippery cases, i.e. with friction thresholds below 0.3 and, especially, below 0.15. The new and, supposedly, more advanced measures EDS' and SEDS still score in the range 0.7 - 0.9, whereas the "traditional" scores fall even down to 0.25 – 0.35 (at the friction threshold 0.15). Thus they seem to follow the relatively well-known notorious behavior of these scores to converge asymptotically towards zero when the event becomes rarer. The difference between the two stations also calls for interpretation. The "spoon shape" structure of EDS' and SEDS at low friction thresholds (Figure 3, right), if not random, may be due to the Utti station being located along maybe the best maintained highway in Finland. Therefore the shape of the curves might also reflect the local road maintenance activities.

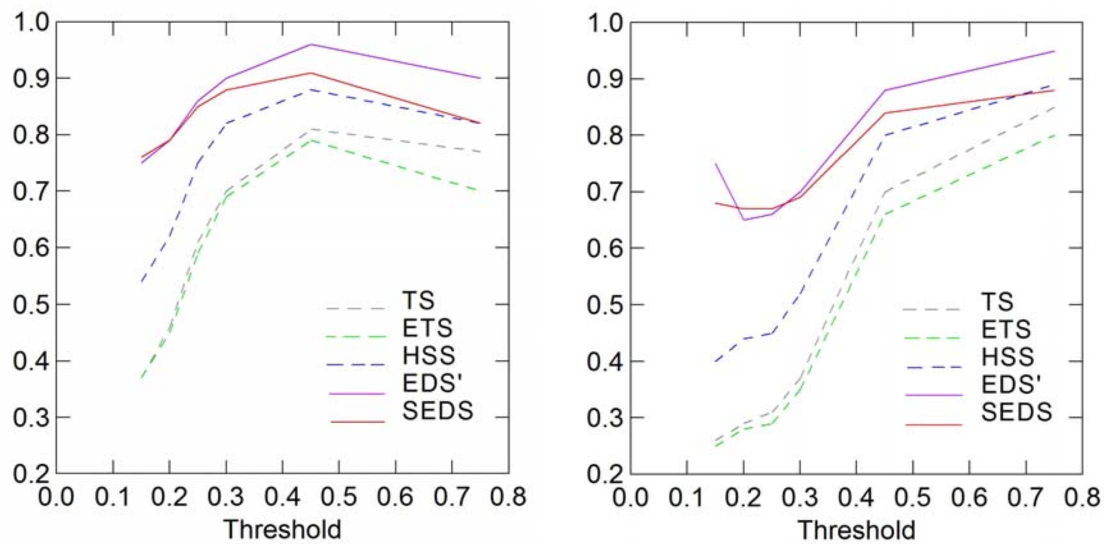

Figure 4. Contingency verification of the statistical friction model for winter 2008-09 for road weather stations stations Anjala (left) and Utti (right).

4. Conclusions and future

If we were able to define the statistical model in the form presented in chapter 2 in a perfect way to interpret friction from the other regression quantities (ice, snow, water layer thickness, temperature) we would expect to reach verification values close to unity (=1), provided that the applied verification measures were sound and sensible. However, since we are in the range 0.7 – 0.95 by applying the scores EDS' and SEDS (cf. Figure 3) there remains room for improvements on the derivation of the statistical relationships.

In this paper, we have not even touched the question of friction forecasting in the time domain by using the road weather forecast model to drive the statistical model. This is the natural next step in the end product development chain, which is actually already in progress. There are also certain challenging bottlenecks as already pointed out in the preceding paper by Hippi et al (2010). From the verification point of view, the initial results shown in this paper represent the quality to expect from the eventual perfect prog end product, under assumptions presented.

References

Hippi M, Juga I, Nurmi P. 2010. A statistical model for road surface friction. *Proc. SIRWEC 15th International Road Weather Conference*. Quebec City, Canada, 5-7 February 2010.

Hogan RJ, O'Connor EJ, Illingworth AJ. 2009. Verification of cloud-fraction forecasts. *Q. J. R. Meteorol. Soc.* 135,643, 1494-1511.

Nurmi P. 2003. Recommendations on the verification of local weather forecasts. *ECMWF Tech. Memorandum* #430, 19pp.

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

Stephenson DB, Casati B, Ferro CAT. 2008. The extreme dependency score: A non-vanishing measure for forecasts of rare events. *Meteorol. Appl.* 15: 41-50.

Vaisala. 2005a. Remote Road Surface State Sensor DSC111. *Vaisala Tech. Report*.

Vaisala. 2005b. Remote Road Surface Temperature Sensor DSt111. *Vaisala Tech. Report*.