ID: 0018

16th International Road Weather Conference
23-25 May 2012 Helsinki, Finland.
Operational OpenRoad Verification.

R.Coulson [1], B.Evans [2], and A.Skea [3]

[1,2,3] Met Office, Exeter, UK

E-mail: robert.coulson@metoffice.gov.uk,
benjamin.evans@metoffice.gov.uk, alasdair.skea@metoffice.gov.uk

Keywords: operational, cost/loss, verification, OpenRoad.

1 INTRODUCTION

OpenRoad is a weather forecasting package that helps road maintenance decision makers to plan, manage and minimise the effects of winter weather by the use of accurate weather forecasts. Road travel on a timely basis is essential for economies globally in the modern world. Cold winter climates impact on safety with dangerous travelling conditions causing delays. Therefore, the prediction and prevention of ice formation is imperative to mitigate against these.

The benefits of winter maintenance have been estimated at approximately eight times the costs of prevention. The estimated world expenditure (SIRWEC 2006) on winter maintenance was put at £6 billion (€10 billion), and clearly the realised benefit from this is significantly material. As a consequence, the verification of road forecasts have been evaluated using cost/loss models. This enables a strong link to be demonstrated to commercial customers for the use of accurate weather forecasts in order to save money.

Experiments have shown that four times more salt is required to melt snow and ice than to prevent its initial formation on road surfaces. Conversely, if the treating agent is applied to the road surface too soon, traffic and precipitation may remove it. In the UK, which has periods of marginal temperatures during the winter months, an estimated (SIRWEC 2006) £150 million pounds per year (at about £1250 per km) has been spent historically. The accuracy and temporal verification of all events is analysed and is an essential part of the operational OpenRoad tools. Due to the progressive research in weather modelling, verification diagnostics are critical for the assessment of improvements for each additional model enhancement that is released operationally.

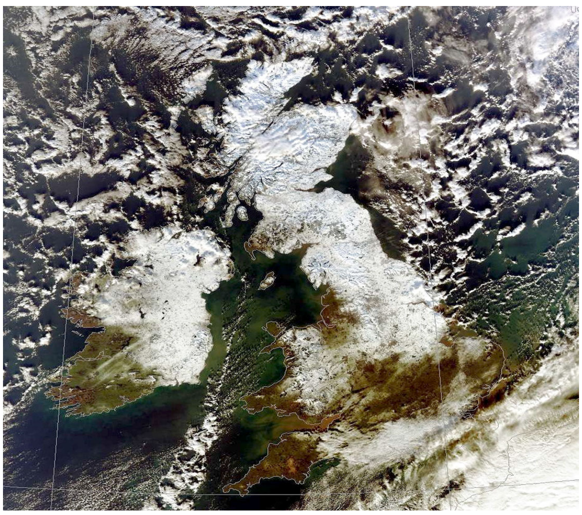


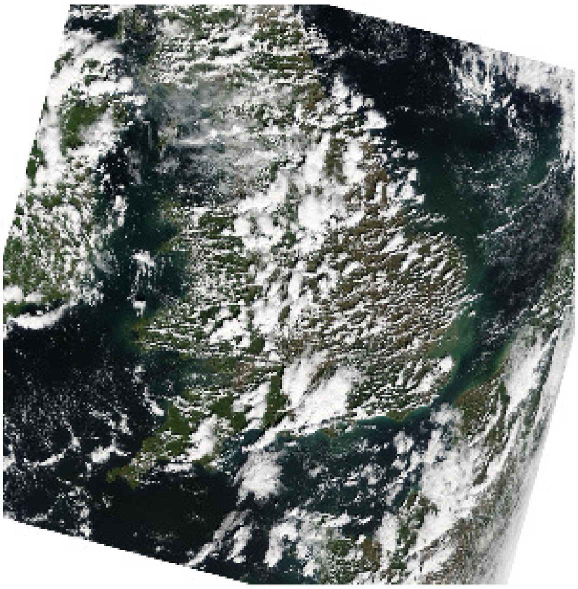
Figure 1. A UK winter satellite image.

2 Recent Improvements in Weather Prediction for Road Forecasting

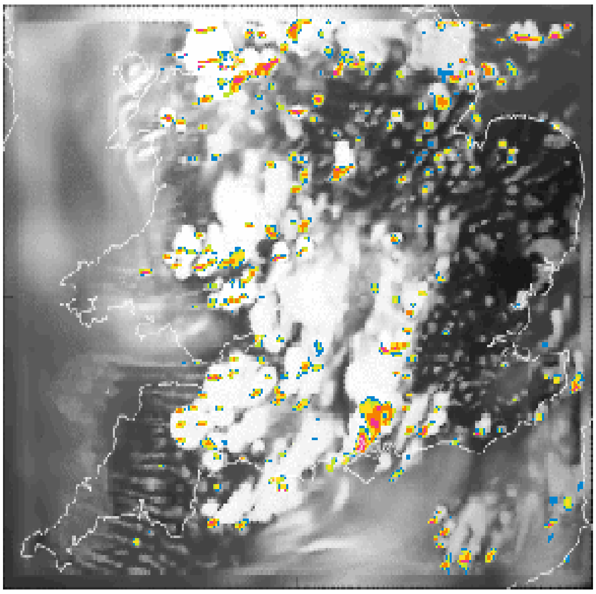
The majority of road weather forecast centres produce forecasts using an energy balance model together with a statistical correction scheme driven by meteorological variables produced from a Numerical Weather Prediction (NWP) model.

The most important factor for providing accurate road weather forecasts is the accuracy of the meteorological variables from the driving NWP model. Since the delivery of the IBM Power 6 in 2010 the U.K Met Office has routinely run a 1.5km horizontal resolution non-hydrostatic version of the Unified Model (UM) over the U.K. [1]. The model has an increased ability to represent local detail such as urban heat island, coastal and orography effects.

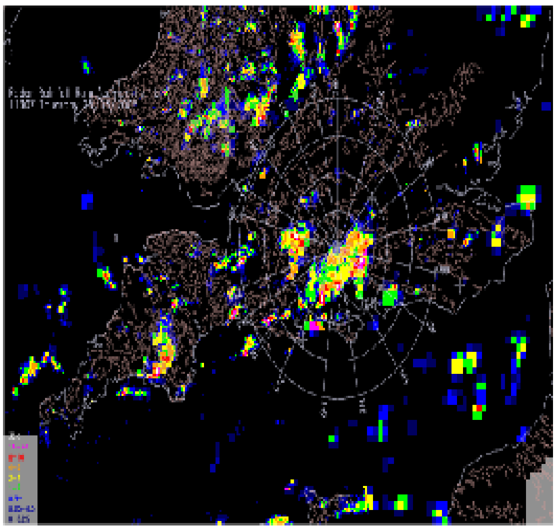
The accurate prediction of the location of cloud and precipitation remains critical for forecasting realistic minimum road surface temperatures and conditions. The high horizontal resolution model coupled with a high vertical resolution (70 levels up to the tropopause) has increased the forecast accuracy of patchy cloud coverage together with precipitation type.



Frame 1.



Frame 2.



Frame 3.

Frame 1, Satellite picture.

Frame 2, Long-wave radiation and precipitation rate from 1km model.

Frame 3, Radar

Figure 2. IOP18 case study comparison with 1km model with satellite and radar.

3 Verification methodology

3.1 Categorical verification

In verifying forecast and model diagnostics, it is fundamental to decide what we are comparing the required attributes against. The question of "truth" is critical in verification, and it is in this case the road weather forecasts are verified against observations from road side sensors at each site. In the winter season 2010/11, this amounted to approximately 342 sites. To ensure that the validation of the forecasts are reliable, a large sample size of matching forecast/observation pairs are collected over individual months and for the whole winter season for each site.

The method for verifying frost events is done by categorising each event by using a 2x2 contingency table. The event threshold is set so that if the road surface temperature observation or forecast is less or equal to zero degrees Celsius then an event is taken as being observed or predicted respectively.

	Frost Observed	**Non Frost Observed**	**Total**
Frost Forecast	Hits (a)	False Alarms (b)	Forecast Frost Events (a+b)
Non Frost Forecast	Misses (c)	Correct Rejections (d)	Non Frost Forecast Events (c+d)
Total	Observed Frost events (a+c)	Non Observed Frost events (b+d)	Number of Events (a+b+c+d)

Table 1. The structure of a 2x2 contingency frost event table.

The accumulated total of the categorised events, from the Table 1, for each site over the month and whole season can then be used to calculate performance metrics. The elements of Table 1 are defined below,

Hit events are recorded when a frost is forecast and a frost is observed.
False Alarm events are recorded when a frost is forecast, but no frost is observed.
Miss events are recorded when a frost is not forecast and frost is observed.
Correct Rejection events are recorded when no frost is forecast and no frost is observed.

A number of scores can be calculated as follows,

Hit Rate (H) = Hits/(Hits+Misses) = a/(a+c) (1)

This translates that fraction of observed frosts that were actually forecast as frosts, and can also be thought of as a conditional probability of the frost being forecast given that a frost was observed. Hit rates range from 1 to 0, with 1 corresponding to a perfect forecast. Often the values are multiplied by 100 to give them in % terms. This is also referred to as the **Frost Detection Rate.**

False Alarm Rate (F) = (False Alarms)/(False Alarms + Correct Rejections) = b/(b+d) (2)

This is the proportion of non-observed frost events that were forecast incorrectly. False Alarm rates range from 0 to 1, with 0 corresponding to a perfect forecast. Again, the values are often multiplied by 100 to give in % terms. It is also an indication of over-forecasting frost events if the value is too high, and possible poor forecasting technique. This is also know as the probability of false detection.

Accuracy = (Hits+Correct Rejections)/(Total Number of events)= (a+d)/(a+b+c+d) (3)

This is basically the fraction of correct forecasts and also know as the **Percentage Correct** if multiplied by 100.

Peirce's Skill Score = (Hit Rate – False Alarm Rate) = **H-F** (4)

Equation (4) is the Peirce's Skill Score also know as the **Kuipers' performance index**. It is an equitable skill score, which means that constant forecasts and random forecasts all give a value of zero. It also has no dependence on sample climate. Good forecasts have values with scores closer to 1, but a range of values for the score can be between -1 to +1.

Critical Success Index = Hits/(Hits+Misses+FalseAlarms)= a/(a+b+c) (5)

The is also know as the **Threat Score**, and is commonly used as a measure of performance for rare event forecasts. It can be thought of in terms of conditional probability, as a sample estimate of a hit given that event being analysed was either predicted, observed, or both. Values range from 0 to 1, with 1 being achieved with perfect forecasts.

3.2 Cost/loss verification using categorical events

The value of a forecast is closely linked to the user requirements. Any maintenance engineer may have a number of courses of action from which to choose due to winter weather, and each action has an associated linked cost. The economic benefits or losses associated with each forecast issued can modelled, two such cost/loss models examples are discussed in this paper. The **Value Index** cost-loss model is used on an operational basis.

3.2.1 Value index cost/loss model

The **Value Index** of a forecast system, (Thornes & Stephenson 2001), is based on criteria based on earlier work (Thompson & Brier, 1955). Let **C** be the cost incurred by the action of treating the road surfaces, and let **L** be the potential loss (predetermined) due to non treatment of the road surfaces (causing accidents and delays in transportation), even after taking into account the discounts achieved by not treating the road network. Hence the cost/loss ratio is denoted by **C/L**. If **p** is the probability on a given night that a frost will occur, it can be shown (Thompson & Brier 1955) that

If **p > C/L** it will be economical to treat the road surfaces.
If **p < C/L** it will not be economical to treat the road surfaces.
If **p= C/L** there is no clear economical direction.

Here, it is assumed that **0 < C/L <1**

Using Thornes (1999), taking a benefit/cost ratio of 8:1 for winter maintenance of roads, **C/L = 1/8**.
(Assumed cost (**C**) values of salting the road for one night could be £20,000, and loss (**L**) due to not salting to be £160,000).

Given an analysis of two types of error in forecasts:

1. The **Miss** event which is the most dangerous for transport systems as potentially accidents may occur, as well as economical impacts and litigation. Cost incurred in the model c**L.**
2. The **False Alarm** event in which the transport network may be treated unnecessarily, which is a waste of treating agents and financial resources. Cost incurred in the model b**C.**

By using **E** to denote the expense of an action taken, it can be shown that with a perfect forecast **E(P)**, where the transport network was only treated when there was a frost is given by,

E(P) = (a+c)**C** . (6)

Using a similar argument, it can be shown that the expense by using issued forecasts from a forecasting system, can be given by,

E(A)=a**C**+c**L**+b**C** (7)

Given that a duty of care is required by most transport maintenance networks, the only option in the absence of a forecasting system is to treat the transport network every marginal frost night. This would be at an expense,
E(S)=(a+b+c+d)**C** (8)

From Richardson (2000), and using equation(14) a **Value Index(VI)** can be defined as,

VI = E(without a forecast) **– E**(forecast issued) **/(E**(without a forecast) – **E**(with a perfect forecast)**)** (9)

VI = E(S) – E(A)/(E(S) – E(P)) (10)

Therefore, in comparing the use of the forecast to the expense of salting every marginal night, and so always protecting the transport network, this can be shown to be simplified to,

VI = [(c+d)-(c/**p**)]/[n-W] (11)

where,

n = total number of nights the road surface temperature ≤ the marginal temperature threshold
W = winter severity (a+c)
p = **C/L =** the cost-loss value

Hence, taking only nights where the road surface temperature ≤ the marginal temperature threshold.

VI = [(c+d)-(c/**p**)]/[b+d] (12)

The range of values for the **Value Index** would be expected to be between 1 to zero, and negative values would suggest that treating the network every marginal night would be more economical.

3.2.2 Relative value cost/loss model

The **Relative Value (RV)** (Richardson 2000) is a cost-loss ratio decision model which compares the **Mean Expense (ME)** of using a forecasting system to one based purely on climatology. Here, the cost-loss ratio is defined as the cost of taking action **(C)** , i.e. treating the transport network against frost, as a fraction of that part of the loss which has been saved by taking the action (**ζ**), e.g. delays in the transport network and accidents. Giving a cost/loss ratio as

α = (the cost of taking the action)/ (potential loss saved by taking the action) **= C/ζ** (13)

The relative value of a forecasting system is defined as the reduction in expense as a proportion of that which would be obtainable by a perfect forecast:

RV = {**ME**(climate)-**ME**(forecasting system)}/{**ME**(climate)-**ME**(perfect forecasts)} (14)

The upper value that can be obtained for **RV** is **1** and this would be for a perfect forecasting system, and a value of zero would indicate that the system is no better than climatology. Any values of **RV** > 0 would indicate that the user of the forecasting system will have an economical benefit over using climatology. Therefore, if a perfect forecast will save a maintenance engineer an amount **X** , then the forecasting system (in comparison to climatology) will save in maintenance 100**RV**% of **X**.

It can be shown that the maximum value realisable from the **RV** cost-loss model for a forecasting system can be obtained by calculating the **Kuiper's performance index (Peirce's Skill Score)** equation (4).

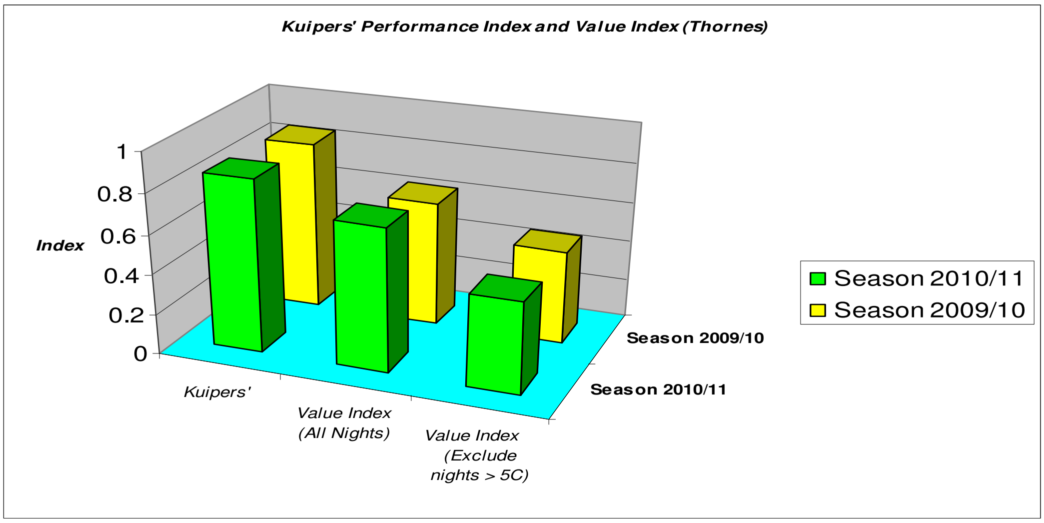


Figure 3. Results for the last few winter seasons showing Kuipers' Performance Index and the Value Index.

4 Operational verification results and discussion

4.1 Scatter-graphs

The operational verification results must be displayed in a way that can be easily understood quickly. After discussions, it became apparent that a much more visual way of presenting the results to customers and forecasters was needed. A scatter-graph is a simple, but also richly detailed, way of displaying for each of the verifying events, forecast and observed pairs. Also, colour coding was applied to enhance and draw the eye to the different contingency categories associated with each event in the forecast-observation space, on the scatter-graph. A traditional traffic-light colour coding approach was used, with a green zone showing expected good forecasts of frosts, a grey zone for correctly forecast non-frost events, a red zone for poor performance events of Misses, and amber for False Alarms. A pie chart also displays the proportion of the population associated with each area in the forecast–observation event space.

The Root Mean Square Error (**RMSE**) provides a measure of the spread of events about the central mean, and Mean Error gives the bias of the sample. A negative bias could indicate that forecasts are too cold, conversely a positive bias could indicate forecasts being too warm. A least squares regression line is also plotted on the graph using all the events in the population, giving the linear trend. This is useful for spotting systematic errors that may creep into the forecasts via the model etc.

4.2 Histogram distribution charts for temporal verification

The temporal verification of forecast events are visualized by using a histogram distribution of error bins. For each forecast event, the difference in time between the forecast temperature cross-over point and the observed temperature cross-over point is calculated. The error associated with an event can be allocated to an error bin, as an incremental tally in each bin. The cross-over threshold temperature can be chosen by the user and the results are displayed as a plot. Crossing–up timing errors, as well as crossing-down timing error plots can be displayed separately as required.

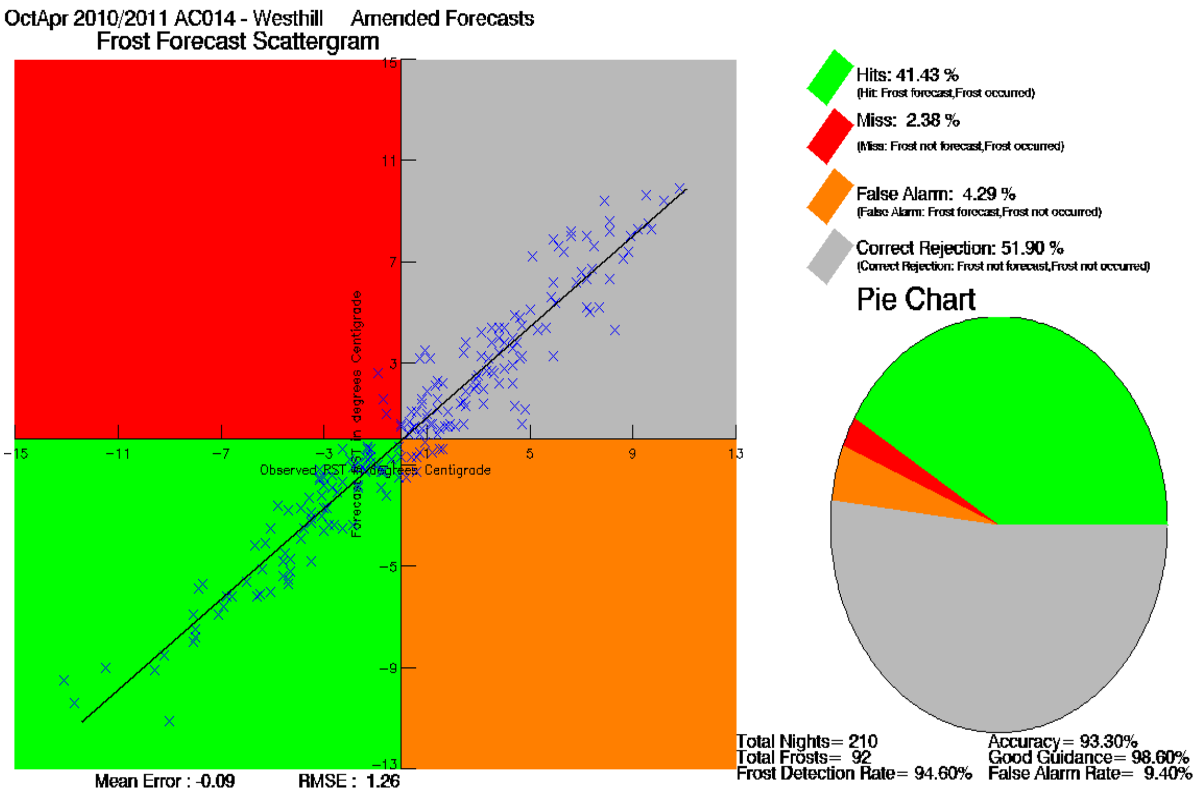


Figure 4. Operational verification scatter-graph of forecast against observed road surface temperatures.

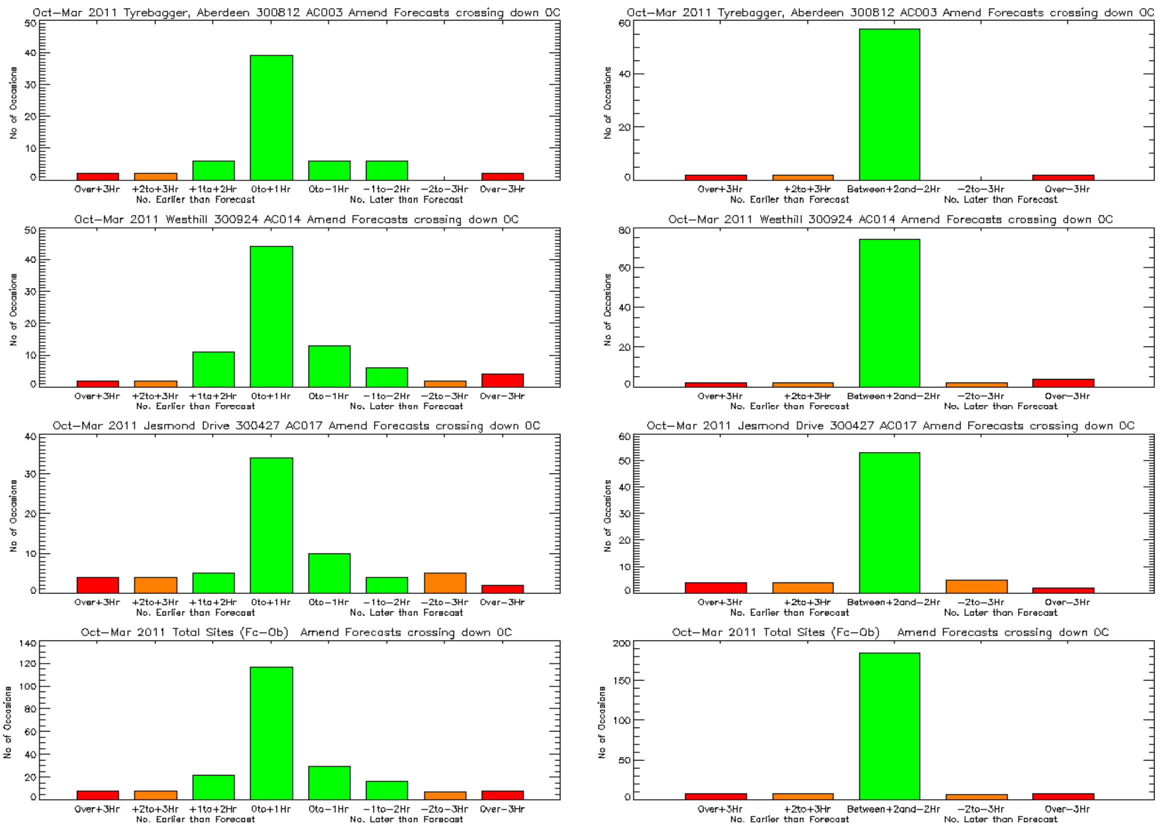


Figure 5. Histogram distribution of (forecast-observed) timing differences crossing down below 0C.

4.3 Road state verification contingence tables

The road surface sensors have for sometime now enabled the actual state of the road surface to be deduced, and this enables a time-series profile to be verified against the forecast road state. This enables a n by m verification contingency table to be derived for a single site. Figure 6, a colour coded example, highlights which correct states were forecast, and also when the road surface had been actively treated against the forecast weather element. The green states for ice-ice, frost-frost, and snow-snow should be as close to zero as possible.

OpenRoad - Road state verification
Month: Nov2010

Site: AC014		Forecast (Amended) State							Total
		DRY	DAMP/MOIST	WET	WET&RAIN	FROST	ICE	SNOW	
Obs State	DRY	23	2	1	1	0	0	0	27
	MOIST	56	54	28	33	0	14	0	185
	WET	10	27	11	15	0	0	0	63
	WET& TREATED	23	32	5	28	22	50	76	236
	FROST	0	0	0	0	0	0	0	0
	ICE	0	0	0	0	0	0	0	0
	SNOW	0	0	0	0	0	7	14	21
	SLUSH	0	0	0	0	0	0	0	0
	TRACE	57	51	15	25	4	20	16	188
	INVALID	0	0	0	0	0	0	0	0
	UNKNOWN	0	0	0	0	0	0	0	0
	FAULTY	0	0	0	0	0	0	0	0
Total		169	166	60	102	26	91	106	720

Figure 6. A verification road state contingency table for a single site.

5 REFERENCES

[1] Lean, Humphrey W., Peter A. Clark, Mark Dixon, Nigel M. Roberts, Anna Fitch, Richard Forbes, Carol Halliwell, 2008: Characteristics of high-resolution versions of the met office unified model for forecasting convection over the united kingdom. Mon. Wea. Rev., 136, 3408–3424. doi: http://dx.doi.org/10.1175/2008MWR2332.1

[2] Richardson D.S., “Skill and relative economic value of the ECMWF ensemble prediction system”, Q.J.R. Meteorol. Soc. Vol. 126, pp 649-667 (2000).

[3] Thompson, J.C. and Brier, G.W., ”The economic utility of weather forecasts”, Mon. Wea. Rev., Vol 83, pp 249-253 (1955).

[4] Thornes, J.E. (1999). UK road salting – an international benefit/cost review. Journal of the Institute of Highways and Transportation, July/August 22-26.
Meteorol. Appl. Vol. 8, pp 307-314 (2001).

[5] Thornes, J.E. and Stephenson, D.B., ”How to judge the quality and value of weather forecast products”, Meteorol. Appl. Vol. 8, pp 307-314 (2001).

[6] Jolliffe, I.T. and Stephenson, D.B., ”Forecast Verification – A Practitioner’s Guide in Atmospheric Science”.

[7] White S., Thornes J., and Chapman L. “A guide to Road Weather Systems”, SIRWEC 2006.