Norwegian Meteorological Institute
met.no

SMHI

METCOOP MEMO No. 01, 2013

Verification of cloud simulation in HARMONIE AROME

A closer look at cloud cover, cloud base and fog in AROME

Karl-Ivar Ivarsson, Morten Køltzow, Solfrid Agersten

Front:
”Low fog over snow”. Photo: Karl-Ivar Ivarsson

MetCoOp

Meteorological Co-operation on Operational NWP (Numerical Weather Prediction)

Norwegian Meteorological Institute (*met.no*)
and
Swedish Meteorological and Hydrological Institute (SMHI)

Date: 2013-01-18
ISSN: 1893-7519

Technical Memorandums (MEMO) from the MetCoOp collaboration are available from **http://metcoop.org/memo**

METCOOP MEMO (Technical Memorandum) No 01, 2013

Verification of cloud simulation in HARMONIE AROME

A closer look at cloud cover, cloud base and fog in AROME

Karl-Ivar Ivarsson, Morten Køltzow, Solfrid Agersten

Summary

The AROME model used as a benchmark for the MetCoOp project has been verified against Swedish automatic stations. Only stations capable of detecting clouds up to 7.5 km have been selected (about 45 stations). Satellite pictures have also been used for subjective evaluation only. Comparison with AROME has been done for the HIRLAM, ALARO and ECMWF models.

Low cloud, detectable cloud, cloud base and fog have been verified in the study. A 'low cloud' means a cloud up to 2.5 km height. A 'detectable cloud' means a cloud up to 7.5 km. A cloud at the lowest model level has been regarded as 'fog' in this study.

The prediction of a cloud base below one kilometre is more important than predicting a higher cloud base. AROME has the best skill for cloud base less than about one kilometre except in cold winter conditions, where HIRLAM has the best skill (equitable treat score, ETS has been used).

The same result is seen for low cloud. AROME has the largest RMS-error for both cloud base and low cloud. This contradicting result is explained by a larger variability of the AROME forecasts compared to the other models.

The RMS-error for detectable cloud is largest for AROME. The reason is partly due to larger variability of the AROME forecasts. The best skill is seen for ECMWF and HIRLAM.

Fog seems to be somewhat over-predicted by AROME. The skill is the best of the models anyway, expect for cold weather in winter.

There is a large over-prediction of ice-fog and low cloud in case of very cold weather (about -30º or colder). The reason for this is not known.

Table of contents

1 Introduction

This document is intended to give a brief summary of the quality of the cloud forecasts from the AROME model. This is the planned operational model for MetCoOp, see also 01/2012 METCOOP MEMO.

Cloud cover and cloud base (lowest height of cloud cover) are two important parameters in a numerical weather forecast. Previously height of cloud cover has not been evaluated due to lack of observations. New automatic stations which have the ability to detect clouds up to 7.5 km have recently been introduced. Data from these stations are useful for verification.

The main focus in this report will be on case studies, but some statistics will also be presented. The different types of verification scores used in meteorological verification statistics are often difficult to interpret. For this reason an explanation of the commonly used scores is included. The model results will be compared with other models available for the weather services of both countries.

This document is organized as follows: In chapter 2 a number of cases studies are presented, chapter 3 gives some verification statistics are summarized, chapter 4 there is a conclusion and in the annex some commonly used verification scores are explained.

Thanks to Rebecca Rudsar and others in the MetCoOp group for contribution and feedback regarding this report.

2 Case studies

2.1 Explanation of the figures

Cloud base:

The figures show forecasted cloud base as low cloud in yellow, middle high cloud in brown, high cloud in blue and precipitation in green. In case there where no post-processed cloud cover from the models, the cloud cover between the lowest and the highest level of interest, has been computed from the model level data.

Observations are also plotted.

Cloud cover:

Automatic stations observing ‘cloud cover’ are marked as triangles:

- ‘ No cloud’ is shown as unfilled (cloud cover 0-2 octas)
- ‘Partly cloudy’ is shown as partly filled (cloud cover 3-5 octas)
- ‘Overcast’ is shown as filled (cloud cover 6-8 octas) ,

Manual observations stations observing ‘cloud cover’ are marked as circles:

- ‘ No cloud’ is shown as unfilled (cloud cover 0 octas)
- ‘Partly cloudy’ is shown as partly filled (cloud cover 1-7 octas)
- ‘Overcast’ is shown as filled (cloud cover 8 octas) ,

Fog:

Forecast fog is marked with dark gray lines and observed fog is marked with three lines to the left of the circles or triangles.

The parameter fog is obtained only from the HIRLAM model. For the other forecast models it is not possible to distinguish between low cloud and fog.

2.2 Autumn

2.2.1 November 9 2011 06UTC

This is a typical case for autumn with low cloud and fog. A high pressure system with weak wind is present over Scandinavia, giving a typical condition for fog and low cloud. The satellite picture is shown in Figure 1. In Figure 2 the corresponding forecasts with AROME and a semi-operational version of ALARO, and the ECMWF forecast and the Swedish HIRLAM 7.1.2 forecast are shown.

Figure 1: Satellite picture over southern Scandinavia, Baltic Sea and northern parts of Germany and Poland. Low cloud and / or fog in white, high cloud in blue, black or dark red.

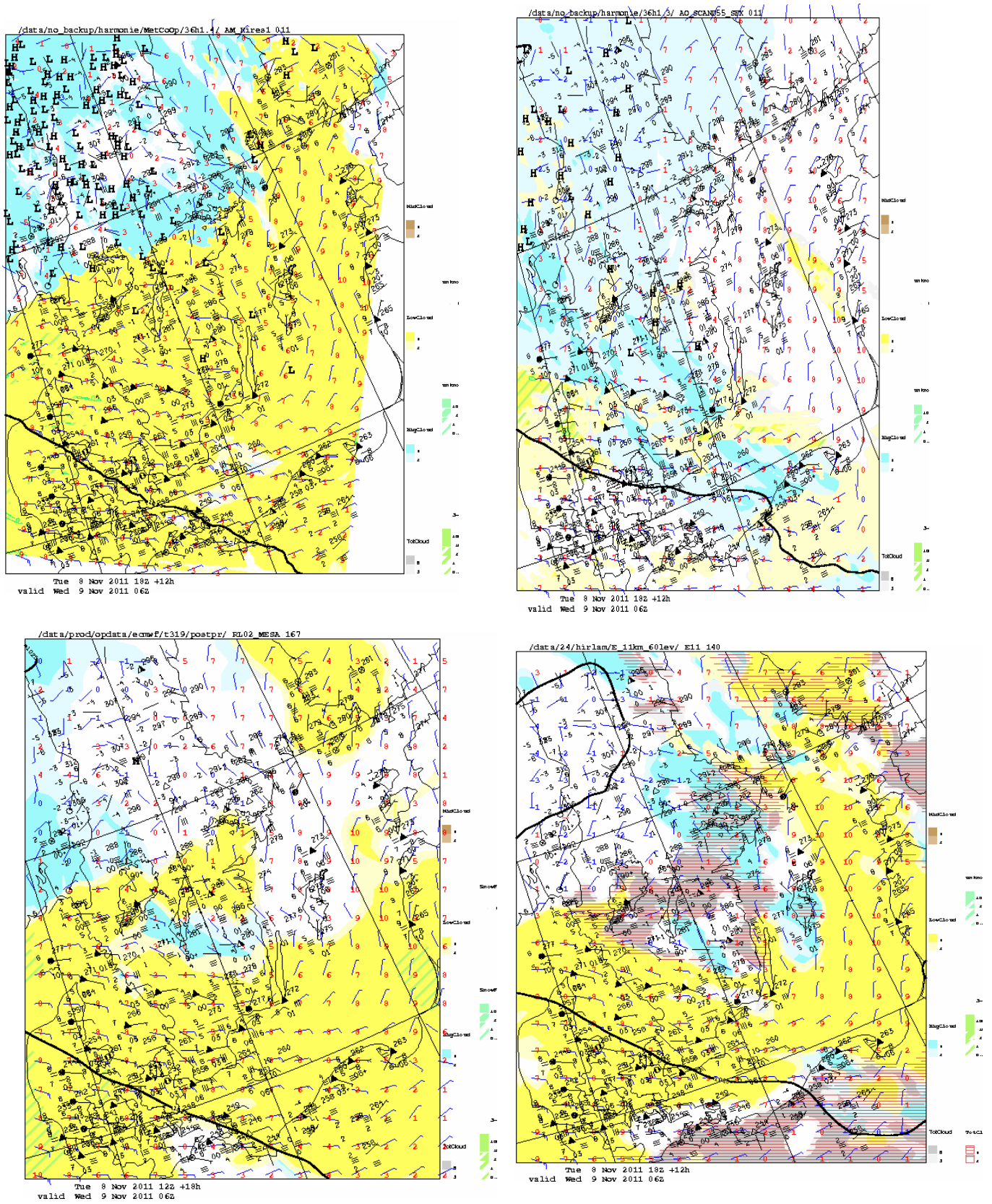

Figure 2: Upper pictures; AROME forecast to the left and ALARO to the right. Lower picture: ECMWF forecast to the left and HIRLAM to the right. Explanation of the legend and colours in the figures, see chapter 2.1.

The AROME forecast of low cloud is fairly accurate in this case. There is however a tendency to over-predict the amount of low cloud. The opposite is seen in the ALARO forecast, where the tendency is to forecast too small amounts of low cloud. The forecasts from the ECMWF and HIRLAM models have no major errors, but there are several smaller errors for example no low cloud north of Oslo and the position of low cloud over northern parts of Germany and Poland.

2.3 Winter

2.3.1 January 31 2012 00 UTC

This is also a case with a high pressure system over Scandinavia. Dry, cold air is blowing northeastward from central Europe towards southern Sweden and southern Norway. The air is warmed over the open water in the Baltic Sea, so it is not a case with particularly cold weather. The low cloud from the different models is shown in figures 4 and 5.

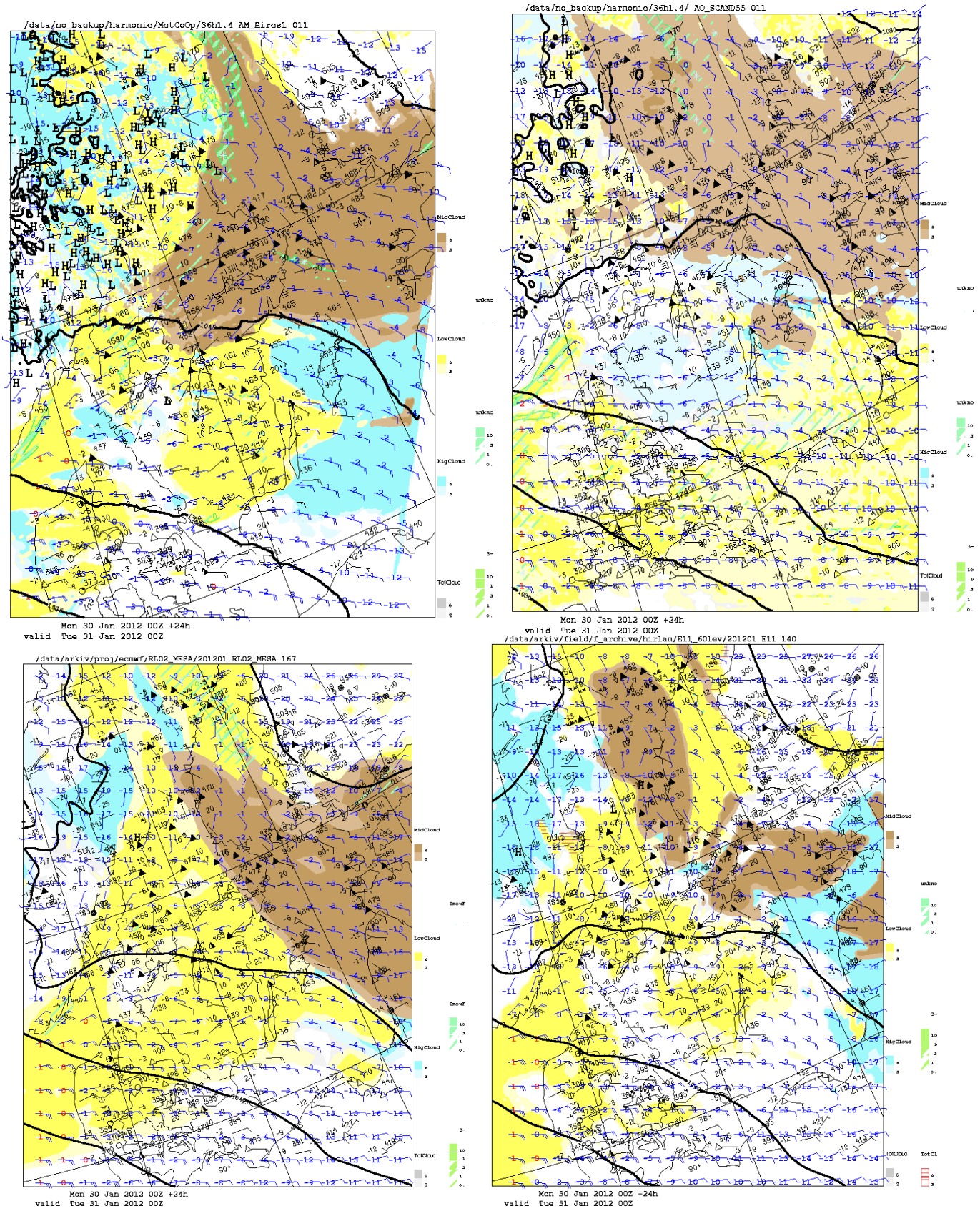

Figure 3: Upper pictures; AROME forecast to the left and ALARO to the right. Lower picture: ECMWF forecast to the left and HIRLAM to right. Explanation of the legend and colours in the figures, see chapter 2.1.

The results from the different models are similar to those in the autumn case. The models AROME, ECMWF and HIRLAM give a fairly accurate forecast of low cloud although there are some small errors. ALARO forecast has a too small amount of low cloud. One exception is over Germany and Poland where ALARO forecasts low cloud together with a small amount of precipitation. Neither the precipitation nor the low cloud are supported by observations.

2.3.2 February 4 2012 06 UTC

This is a case with more pronounced cold weather. There is a cold high pressure system over northern Russia and a ridge over Scandinavia. A minor low pressure system is seen over the Baltic Sea. Low cloud for different forecasts are seen in figures 6 and 7.

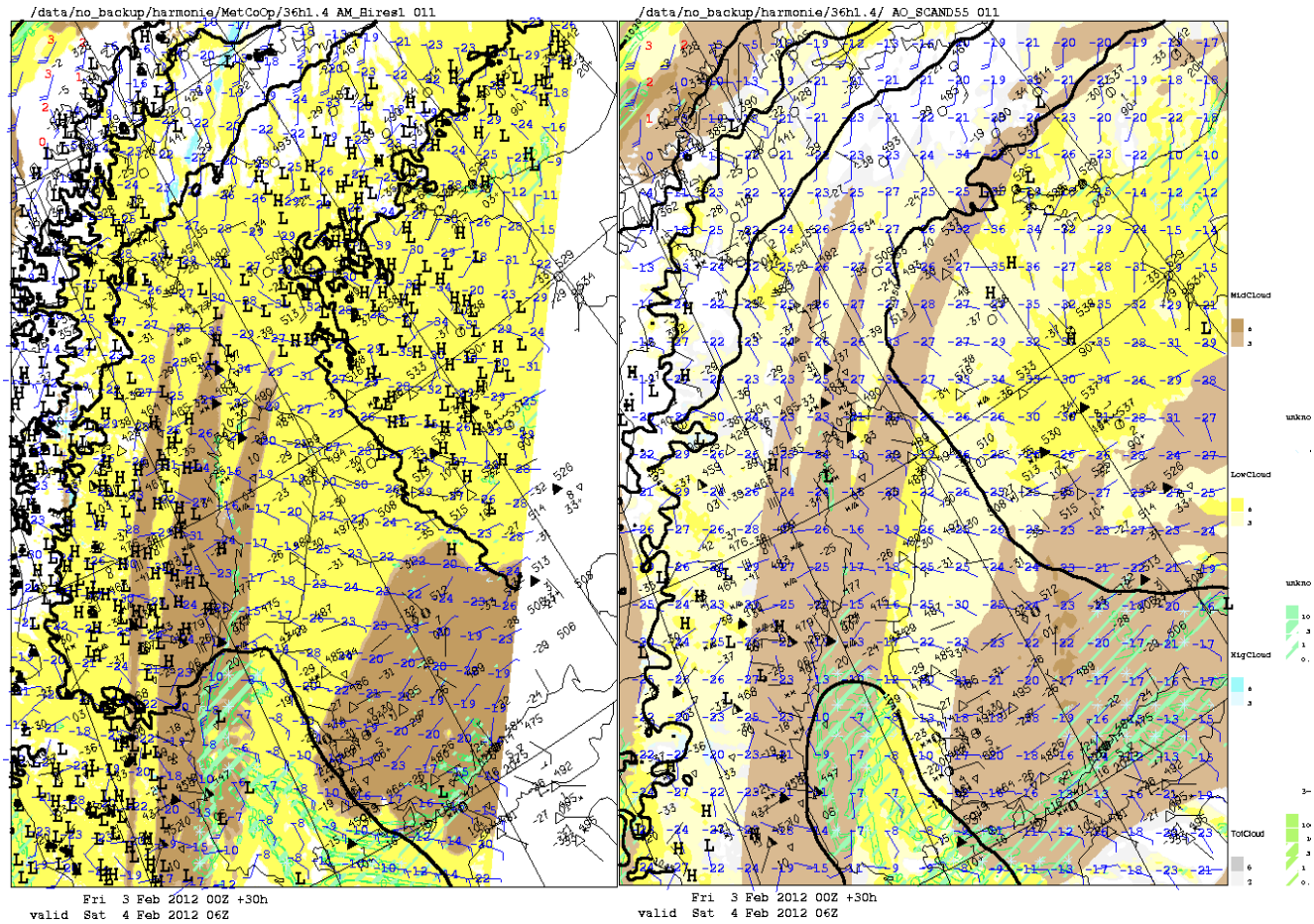

Figure 4: AROME forecast to the left and ALARO to right. Explanation of the legend and colours in the figures, see chapter 2.1.

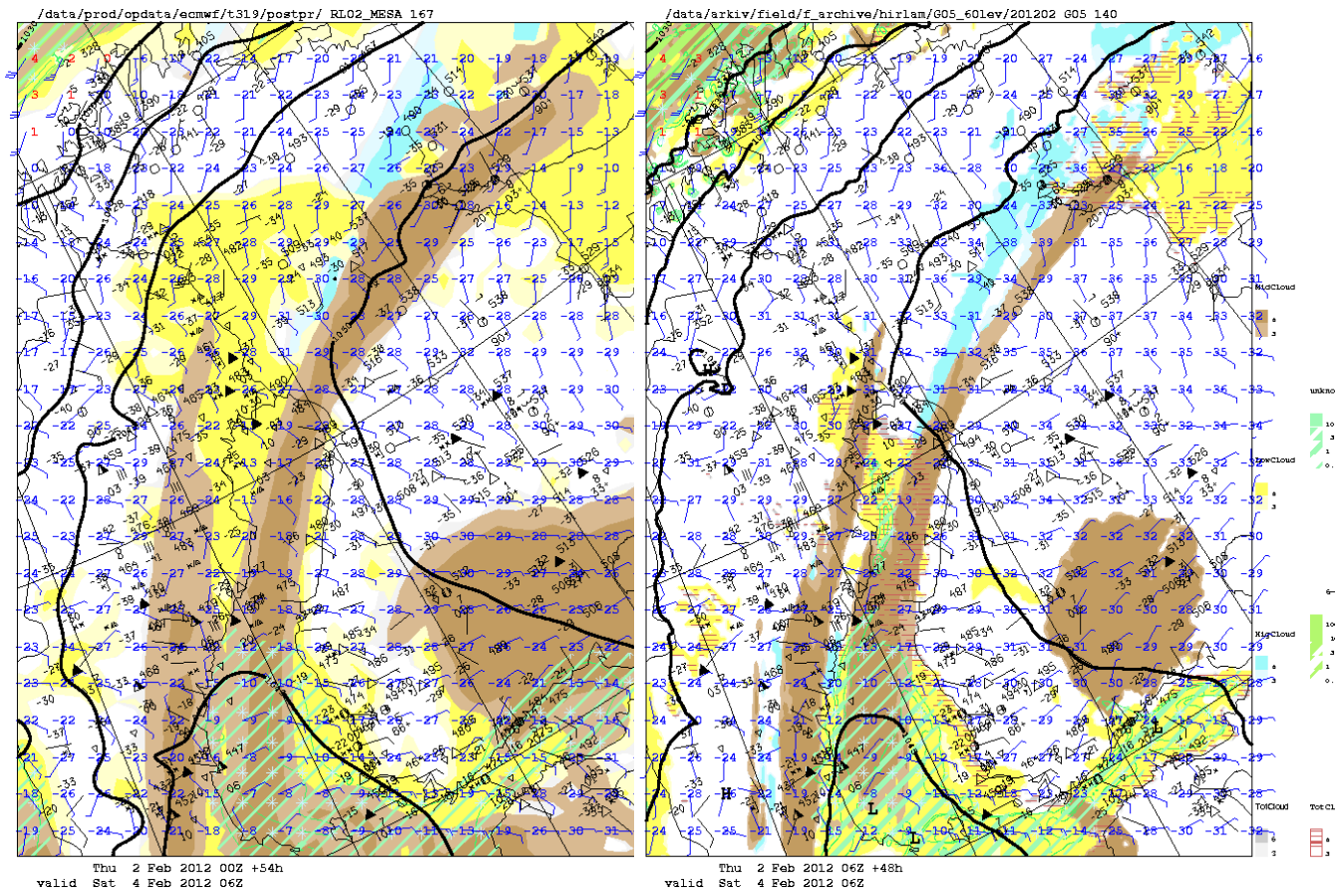

Figure 5: ECMWF forecast to the left and HIRLAM to right. Explanation of the legend and colours in the figures, see chapter 2.1.

In this case of very cold weather there is a problem with the AROME forecast of low cloud. The occurrence of low cloud is severely overestimated. The 2m-temperatures are not cold enough. The same temperature problem is seen in the ALARO forecast, but the amount of low cloud is not overestimated. The ALARO model has a tendency to forecast the fraction of low cloud as well as middle level cloud too often near 0.5 (4 octas). ECMWF has no particular bias in the low cloud cover amount but the low cloud is often misplaced. The HIRLAM forecast has too little low cloud cover amount in this case.

2.4 Spring

2.4.1 March 23 00 UTC

During spring time an over-prediction of fog is sometimes apparent. An example is shown in

Figure 6.

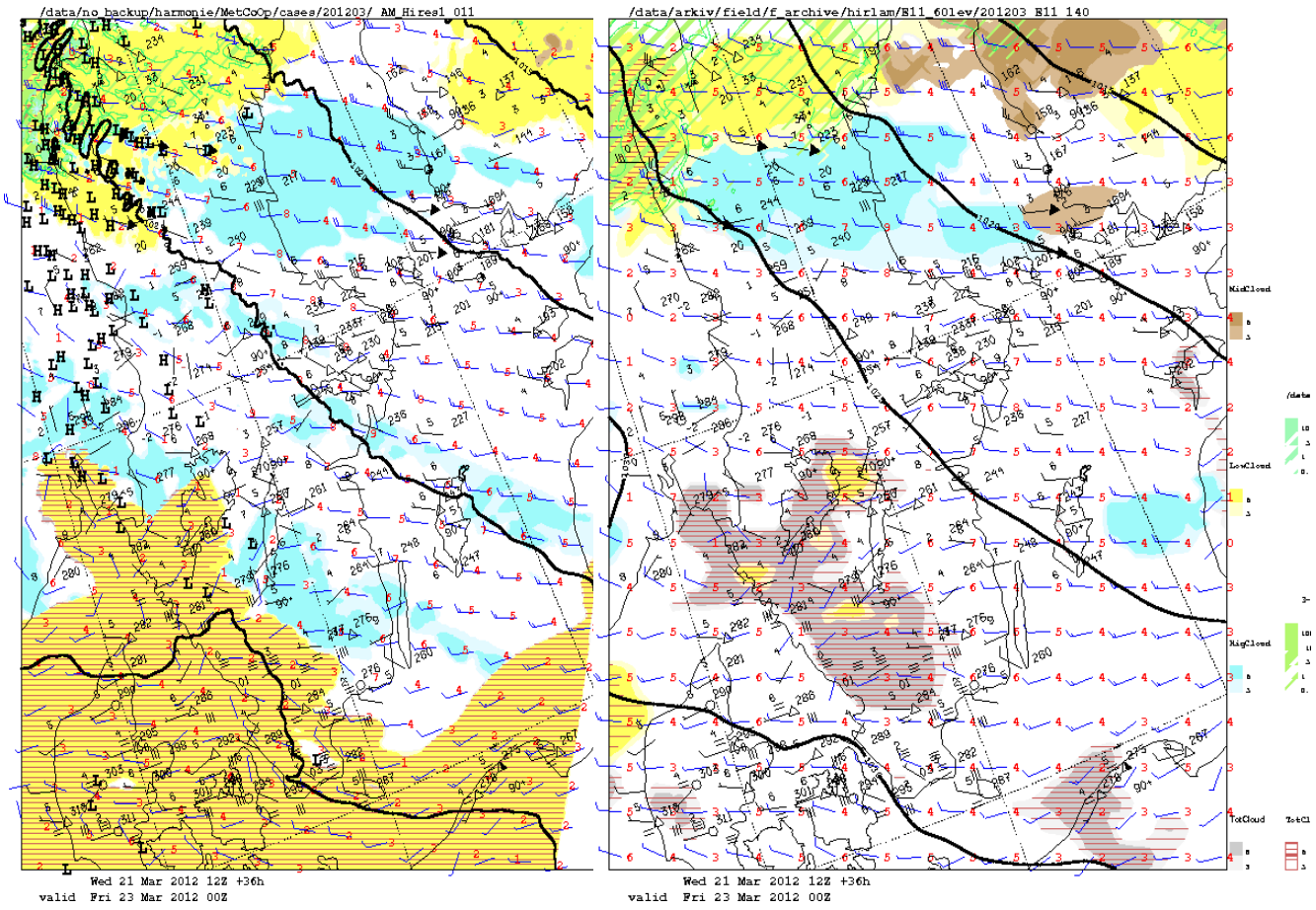

Figure 6: AROME forecast to the left and HIRLAM to right. Explanation of the legend and colours in the figures, see chapter 2.1.

Generally the AROME forecast of low cloud is fairly accurate, although some over-prediction is seen. An example of this is seen in the figure. According to the observations there is little or no low cloud over Denmark (in the left corner).but the AROME forecasts low cloud over most of the area. The HIRLAM forecast has less over-prediction, but there is too much fog over parts of south-western Sweden.

2.5 Summer

2.5.1 July 25 12 UTC

The AROME model was updated in summer 2012. The new version, 37h1.1, replaces the former version 36h1.4. The model domain was also extended. During summer 2012, the Arome forecast was regularly compared with several other models. This was done for a number of forecast lengths. An example is shown in Figure 7.

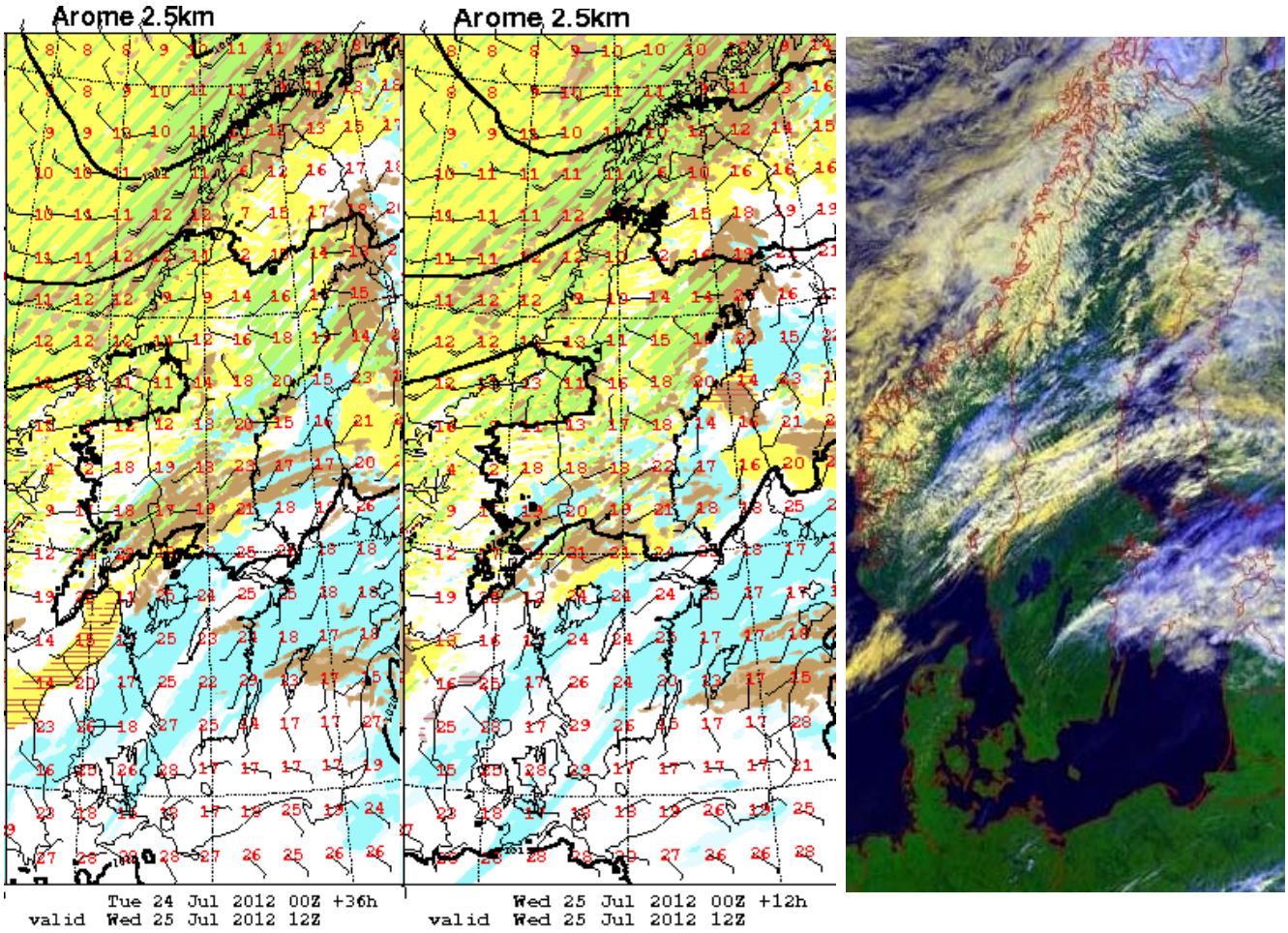

Figure 7: A 36 hour, a 12 hour AROME forecast (Explanation of the legend and colours in the figures, see chapter 2.1) and a satellite picture at valid time to the right.

In Figure 7 the two different AROME forecasts are compared with a satellite picture. Comparing the two forecasts and the observations over the Skagerrak area we see that the 12 hour forecast is better than the 36 hour forecast. This is a typical example of over-prediction of low cloud for the longer forecast lead times. Comparisons have also been done between the old (36h1.4) version and the new one (37h1.4) with respect to forecasts of low cloud, but no significant change is seen.

The other models (ECMWF, ALARO and HIRLAM) had no low cloud over Skagerrak. (Not shown.)

3 Verification statistics

3.1 Spring 2012 (AROME cycle 36h1.4)

Due to technical reasons AROME forecast are only available for a limited period in late April and May. Only automatic stations which can measure cloud cover and cloud base up to 7500 m are used. There are approximately 45 such stations and they are all located in Sweden. The observed and forecast frequencies of **cloud base** (for cloud cover of at least 2 octas) are plotted in figure 8.

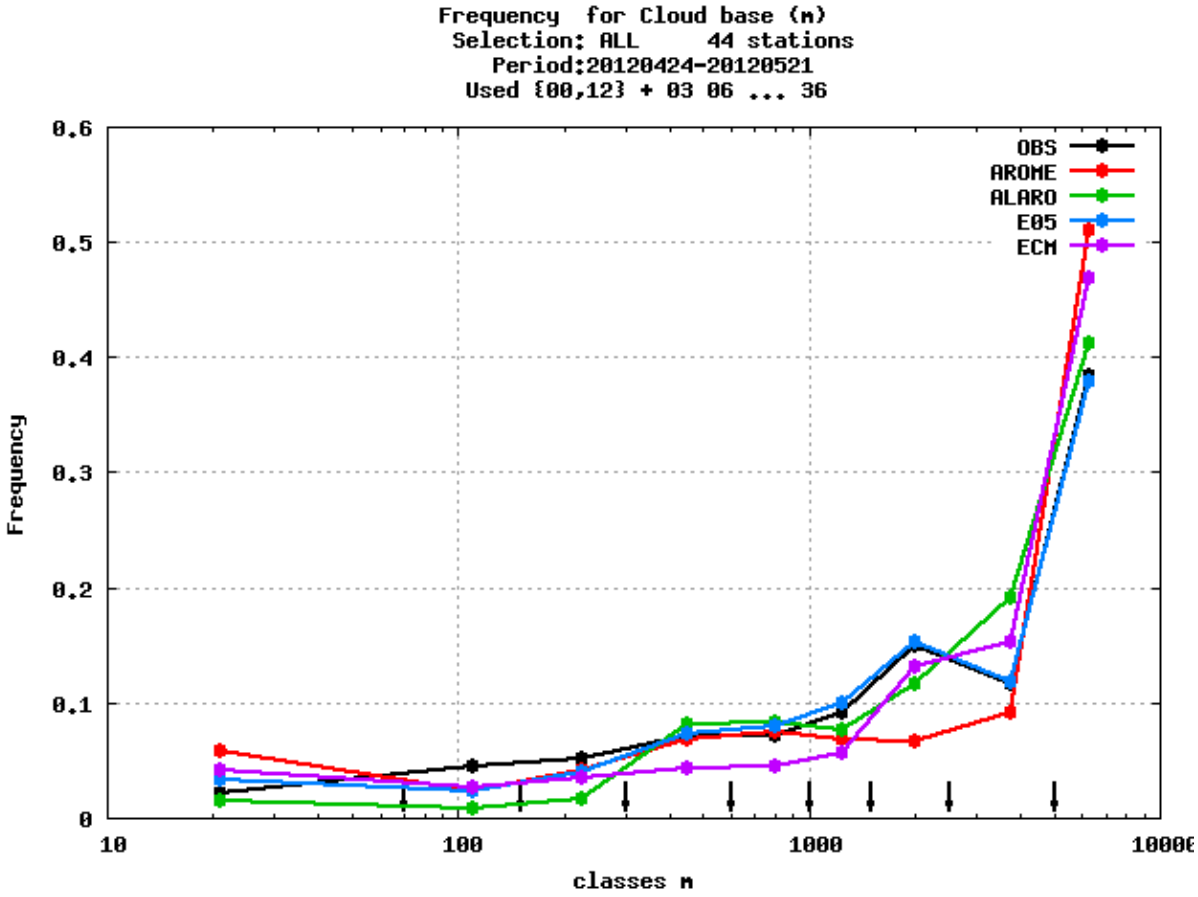

Figure 8: The frequency of cloud base for different heights in meters for AROME (red), ALARO (green) , HIRLAM 7.3 (blue) , ECMWF (violet) and observed (black)

It is clear from Figure 8, that during this period the occurrence of a cloud base of 20 m (in practice "fog") in AROME is overpredicted. The forecast from ECMWF and HIRLAM are slightly more skilful. Only ALARO underpredicts fog. For low stratus (100-300m) all models underestimate the frequency of occurrence. Cloud base above one kilometre is more often observed than forecasted in AROME, but no clear systematic error is seen for HIRLAM.

The skill score result for different thresholds of cloud base is shown in Figure 9. Only ETS is shown, but other scores, such as Kuipers skill score, AI etc give a similar result. See Annex 1 for explanation of the different scores.

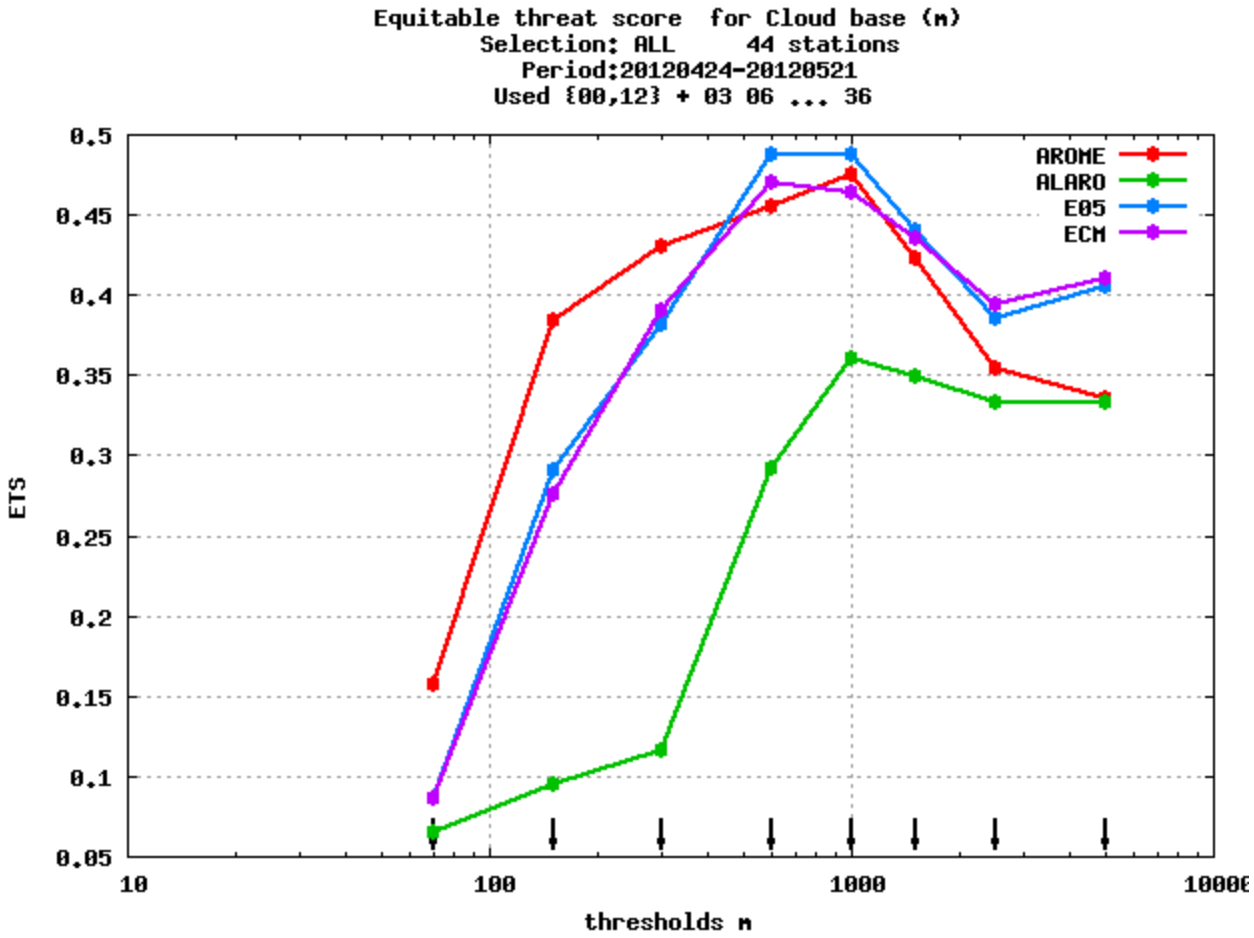

Figure 9: ETS for different thresholds of cloud base. AROME (red), ALARO (green), HIRLAM 7.3 (blue), ECMWF (violet)

AROME has the best score of the four models for the lowest thresholds. HIRLAM and ECMWF have a similar result during this period. The reason for the poor result for ALARO is not known.

The verification results of **cloud cover;** both low cloud, (up to 2.5 km) and all detectable cloud (up to 7.5 km) for the same period as for the cloud base are shown in the following figures. The results for different forecast lengths are in Figure 10.

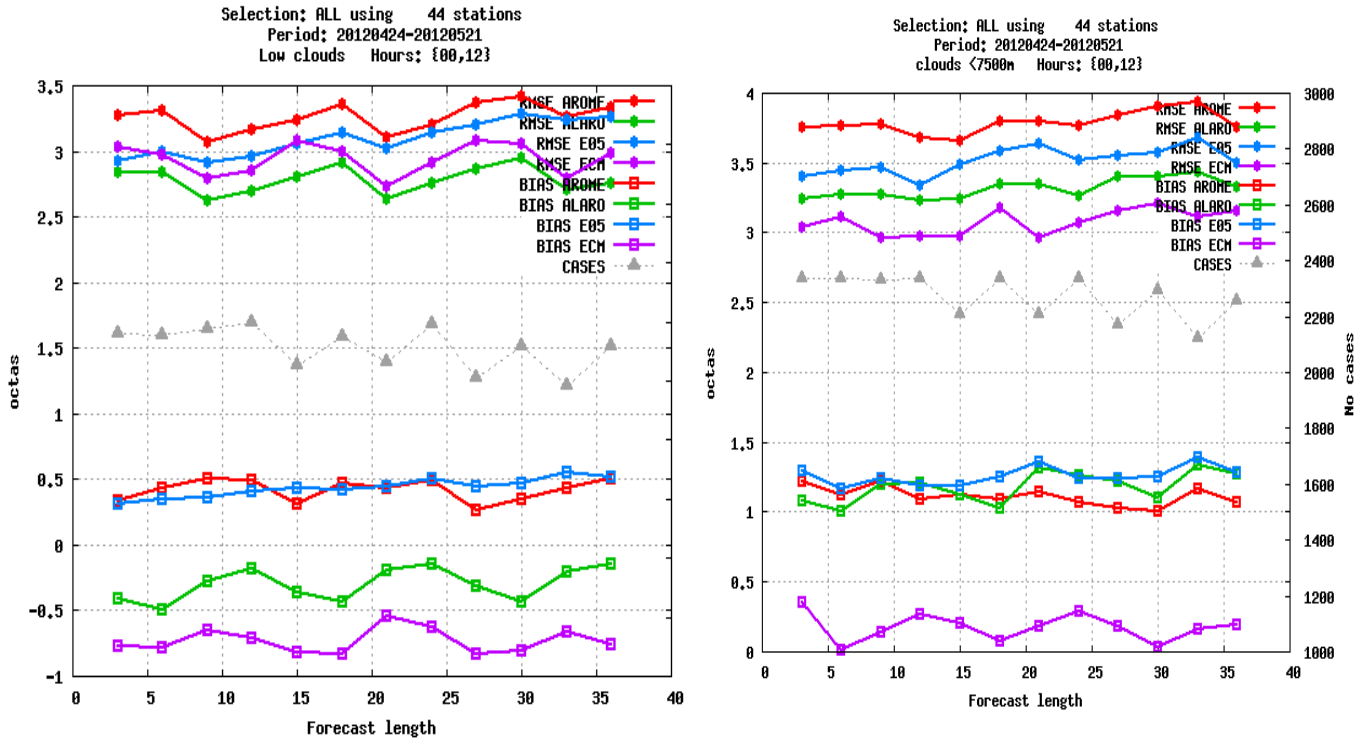

Figure 10: The RMS error and bias for low cloud and all detectable clouds. AROME (red), ALARO (green), HIRLAM 7.3 (blue) and ECMWF (violet)

There are more clouds in the forecast from AROME both regarding low cloud and all detectable cloud compared to the automatic observations. For ECMWF the opposite is seen. Many models have been tuned to fit manual observations which generally have a larger amount of cloud cover due to perspective effects. This effect is most pronounced for clouds with large vertical thickness such as cumulonimbus. Models with no bias relative to manual stations may have a positive bias when verified against automatic stations.

AROME has the largest RMS error, and ECMWF and ALARO have the smallest error. It might seem contradictory that AROME has a relatively large RMS error for low cloud, but has relatively high skill for cloud base (see Figure 14-ETS). One possible reason is the different variability of cloud cover in the models.

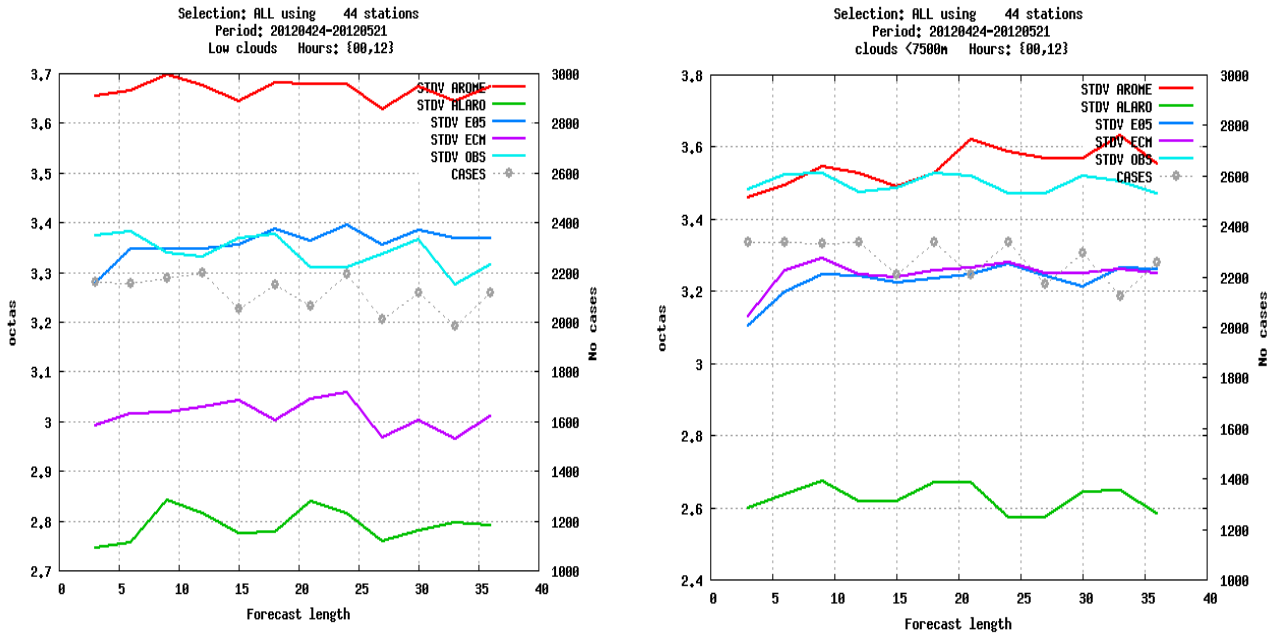

Figure 11: The standard deviation of the forecasts of low cloud (left) and all detectable cloud (right) . ALARO (green), HIRLAM 7.3 (blue), ECMWF (violet) and observation light blue.

A high variability often leads to higher RMS error and AROME has the highest variability of all models. Probably the variability in AROME is too high, since it is higher than for observations. This is seen for both low cloud and all *detectable* cloud (se Figure 11). An automatic station only 'sees' a very small area of the sky, near zenith. It is probably much smaller than a square of 2.5 km which is the size of the grid squares of AROME. A grid square of AROME represents a larger area and thus should have a lower variability than the observations.

An alternative method of examining the same characteristics of the models is to use the frequency bias, see Figure 12.

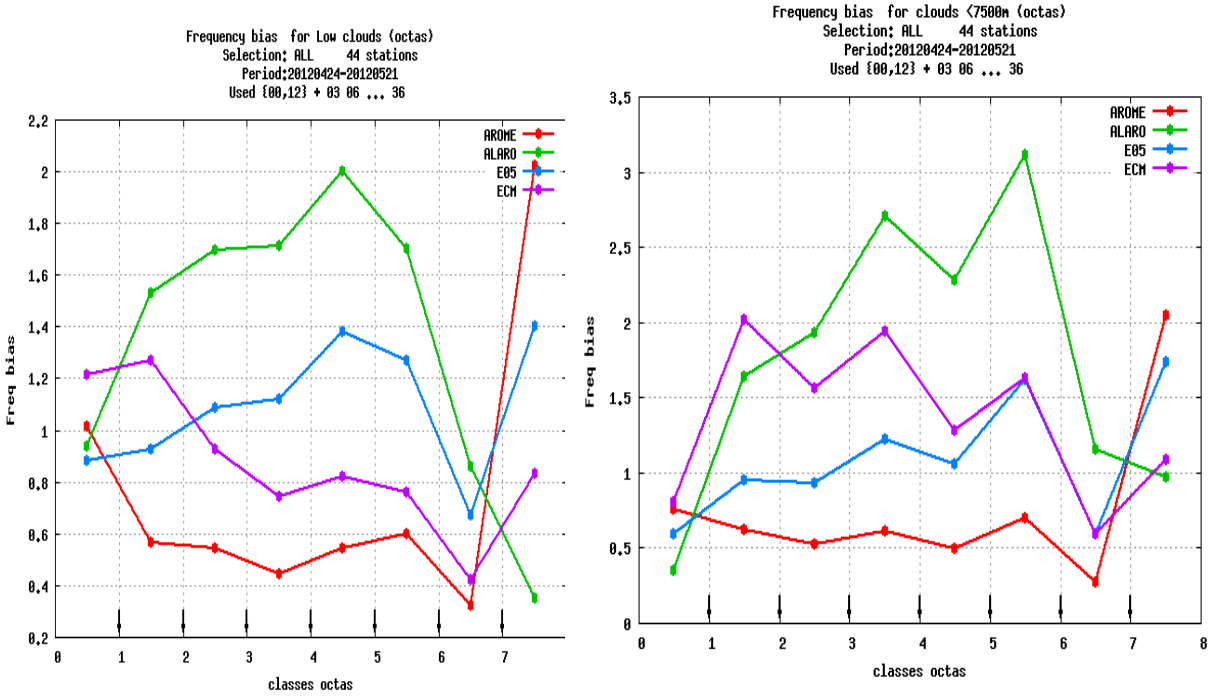

Figure 12: The frequency bias of the forecasts of low cloud (left) and all detectable clouds (right). AROME (red), ALARO (green), HIRLAM 7.3 (blue) and ECMWF (violet)

The frequency bias should ideally be near one. AROME has too few events with octas between 2 and 6, and too many events with octas between 7 and 8, especially for low cloud. ALARO has the opposite characteristics.

The diurnal cycle of cloud cover is seen in Figure 13.

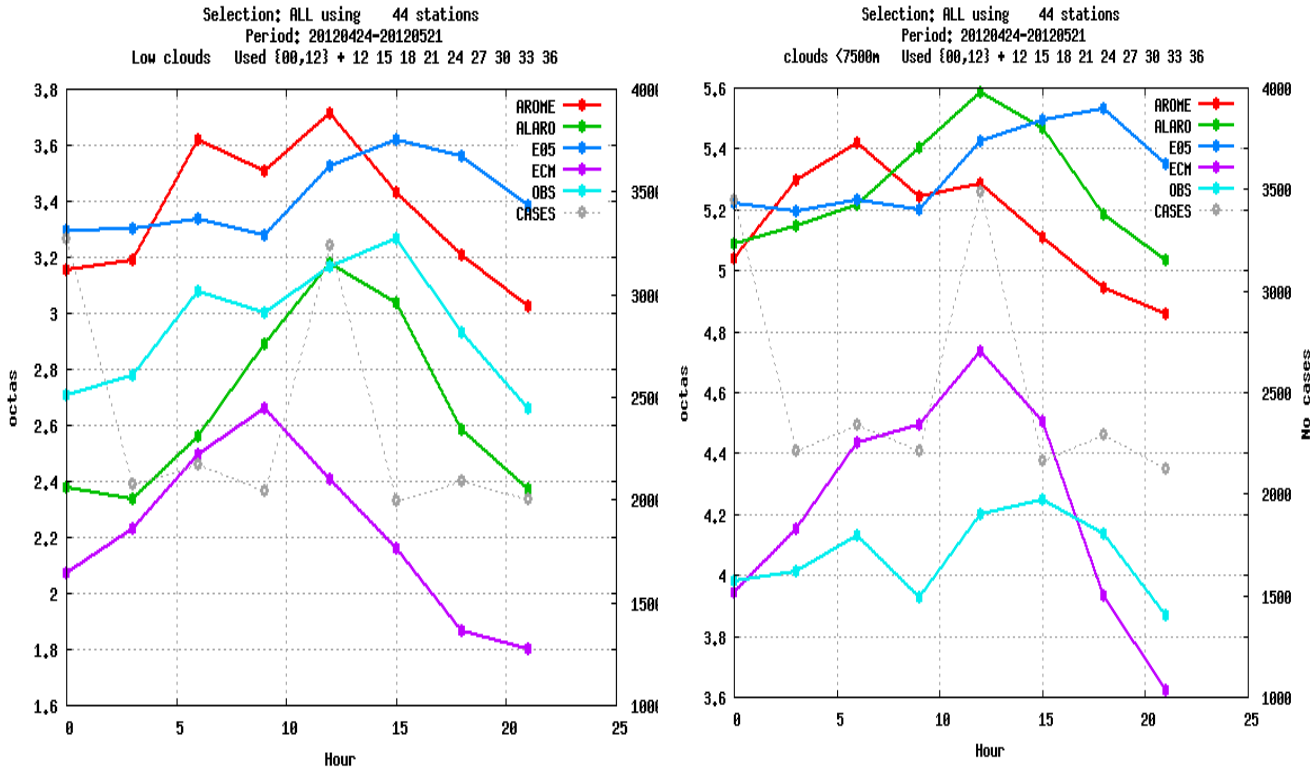

Figure 13: The amount of low cloud (left) and detectable cloud (right) for different times of the day (in UTC). AROME (red), ALARO (green), HIRLAM 7.3 (blue), ECMWF (violet) and observation light blue

The weak diurnal cycle of low cloud and all detectable cloud is well predicted by the AROME model, but on average the amounts are higher than observed. ECMWF over-predicts diurnal cycles for both cloud types.

The different skill scores show somewhat diverging results, but ETS, shown in Figure 14 is representative.

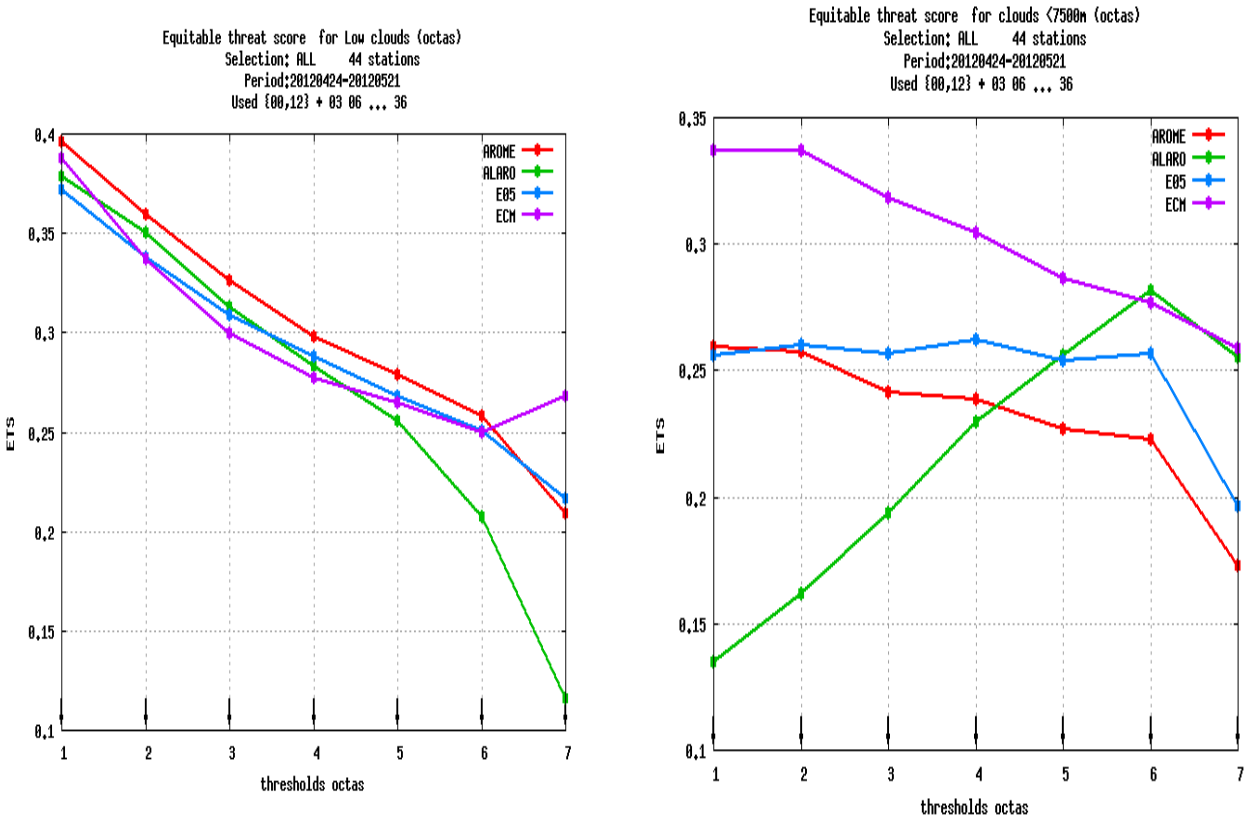

Figure 14: ETS for low cloud (left) and all detectable clouds (right). AROME (red), ALARO (green), HIRLAM 7.3 (blue) and ECMWF (violet)

AROME shows a good verification result compared to the other models for low cloud, and an average result for cloud base less than 7500 m. A model with low variability, such as ALARO, often has a good verification result with respect to RMS error, but the verification result is not that good when different skill scores based on contingency tables (see Annex 1 for explanation) are examined. This is seen from the verification result for this period.

3.2 Winter 2010 (AROME cycle 37h1.1)

This period from November 20 to December 9, 2010 was a cold period in early winter. The quality of forecasting cloud cover, cloud base and fog when there are cold or very cold weather conditions is dependent on the ability of the model to describe the processes in mixed-phase and ice clouds. It is also dependent on how realistic the surface scheme is with respect to snow and ice. These processes should also take forest into consideration, since forest covers a large part of northern Europe.

There are still some weaknesses in the description of these processes in AROME and in the current version of SURFEX (the surface scheme used together with the HARMONIE AROME model).

In the previous example (from chapter 2.3,

Figure 6) there is an over-prediction of the amount of low cloud and the forecast temperatures are not low enough. However the results in chapter 2.3 are from the cycle 36h1.4 of HARMONIE. Some improvements have been made in cycle 37h1.1. The two week period with cold winter conditions (20.11-09.12.2010) verified here is with AROME cycle 37h1.1. The observed and forecast frequencies of cloud base (for cloud cover of at least 2 octas) for this cold period are plotted in Figure 15.

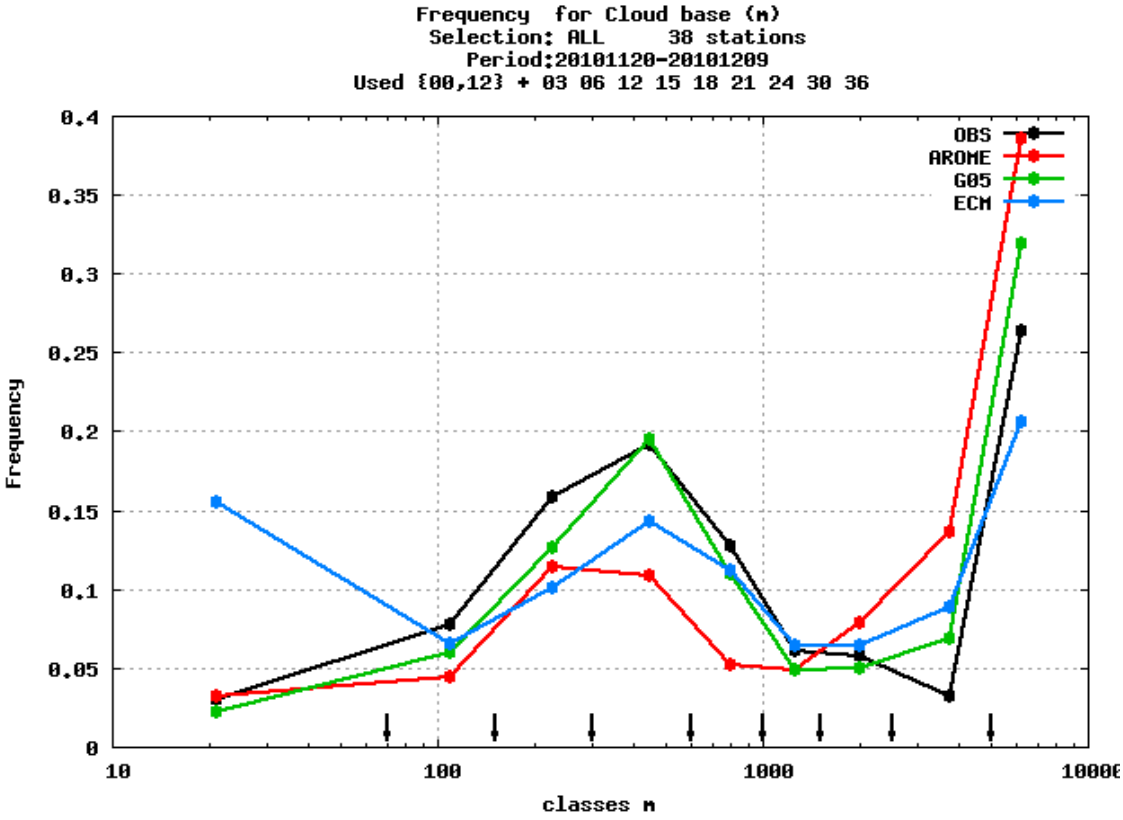

Figure 15: The frequency cloud base of different heights in meters for AROME (red), HIRLAM 5.5km (green) , ECMWF (blue) and observed (black)

The bias of the forecast can be derived as the difference between the forecast frequency and the frequency of the observations as seen in AROME has no significant bias for the lowest cloud base, but too few occurrences of clouds with height between 100 m and 1000 m. The opposite is seen for cloud base for middle level and high clouds. The most striking result for the ECMWF forecast is the large bias for the lowest level. There is a negative bias for ECMWF for cloud base between 100 m and 1000 m. For AROME this negative bias is more pronounced. For higher cloud base there is a slight positive bias for all models. The ECMWF model has been improved since 2010, so this bias is probably reduced in the present version. HIRLAM has no major bias.

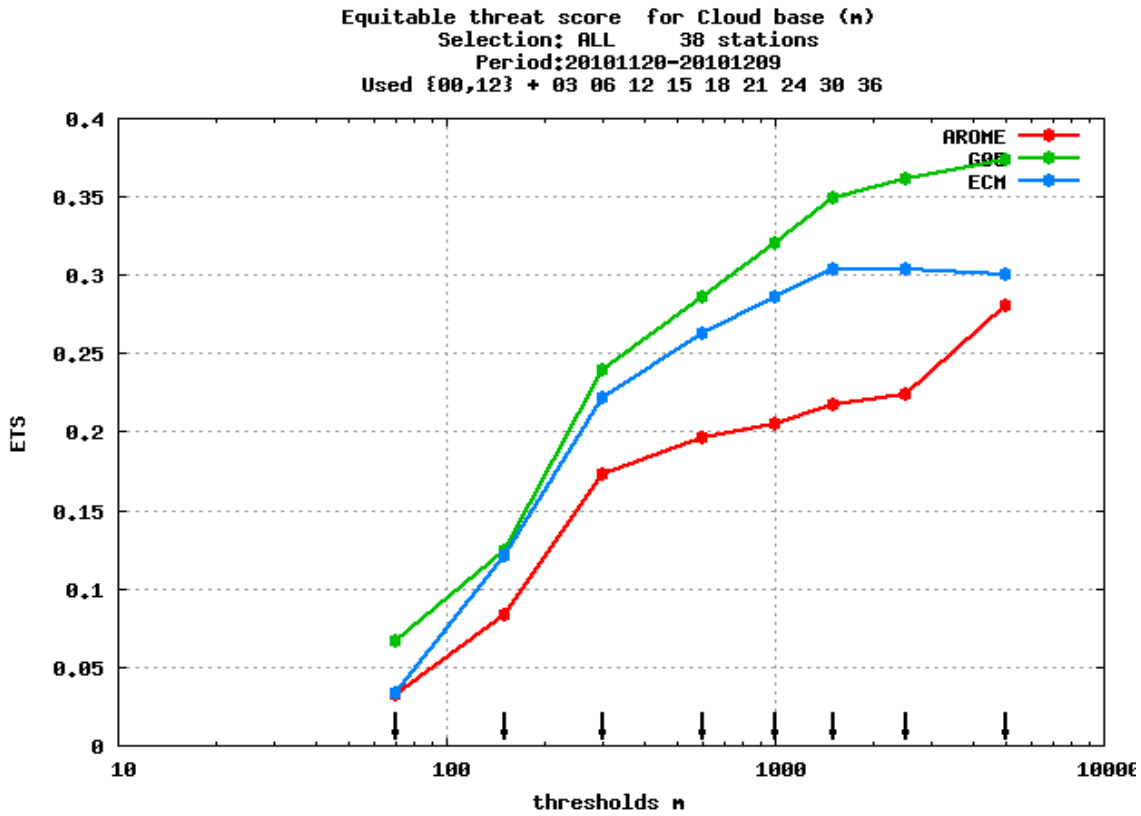

Figure 16: The ETS for cloud base of different heights in meters for AROME (red), HIRLAM-7.1.2 with 5.5 km grid (green), ECMWF (blue) and observed (black)

The result with ETS is seen in figure 17.

The result is in general not as good as for the spring period. One reason may be that the link between an area of condensation in a model and the observed cloud or fog is weaker when it is cold because the cloud may be an ice cloud instead of a water cloud. An ice cloud or ice fog is more optically transparent than a water cloud or fog consisting of water droplets, thus an observed ice cloud may be an area with a lower visibility. This means that treating cloud at the lowest level as fog may be questioned. Another reason for the degradation of the forecast quality could be because the micro-physics in mixed-phase clouds and ice clouds is more complicated than in water clouds.

The verification results of **cloud cover,** both for low cloud, (up to 2.5 km) and all detectable cloud (up to 7.5 km) for the cold period are in the following figures.

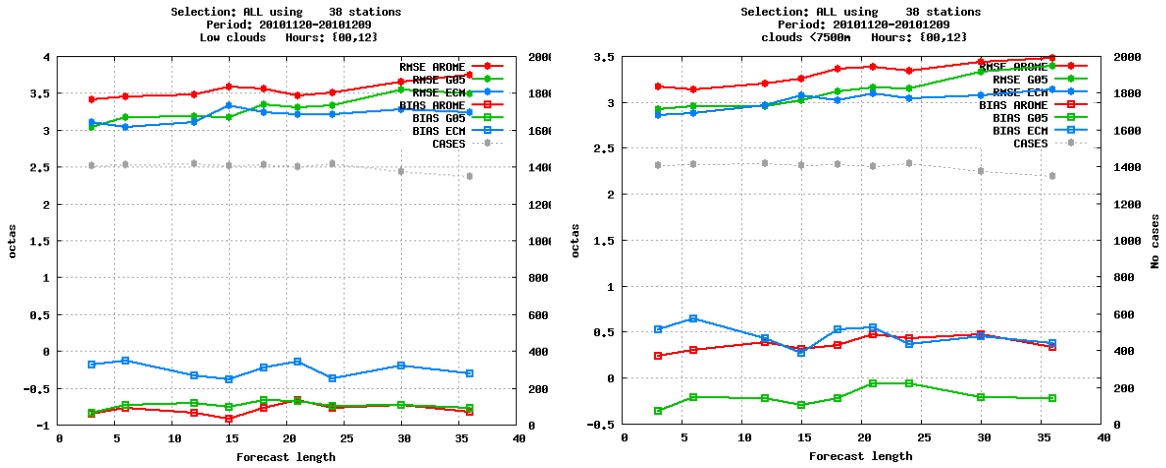

Figure 17: The RMS error and bias for low cloud and for all detectable clouds. AROME (red), HIRLAM 7.1.2(green) and ECMWF (blue).

The RMSE is smallest for ECMWF and largest for AROME. The RMSE may be affected by the variability of the forecast. The standard deviation for the forecasts and observations are seen in Figure *18*.

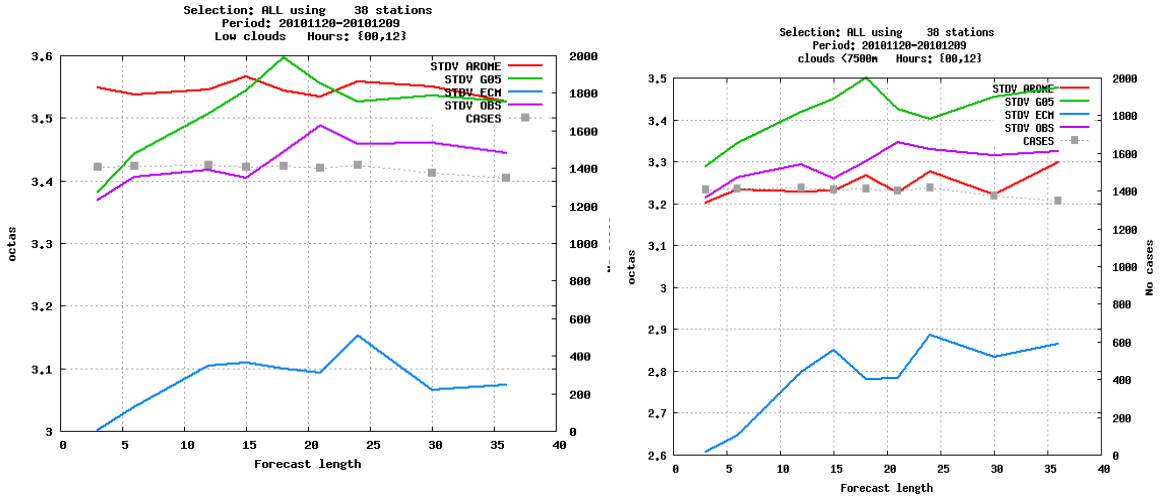

Figure 18: The standard deviation for low cloud and for all detectable clouds. (See explanation of the statistical methods in the last chapter.) AROME (red), HIRLAM 7.1.2(green), ECMWF (blue) and observation in violet

ECMWF has the lowest variability, and this contributes to the small RMS error seen in figure 17. HIRLAM has a similar variability to that which is found in AROME and a bias of the same magnitude, thus the large RMS error for AROME is not explained by the variability and bias. The frequency bias for low cloud and all cloud below 7500 m for different octas are shown in

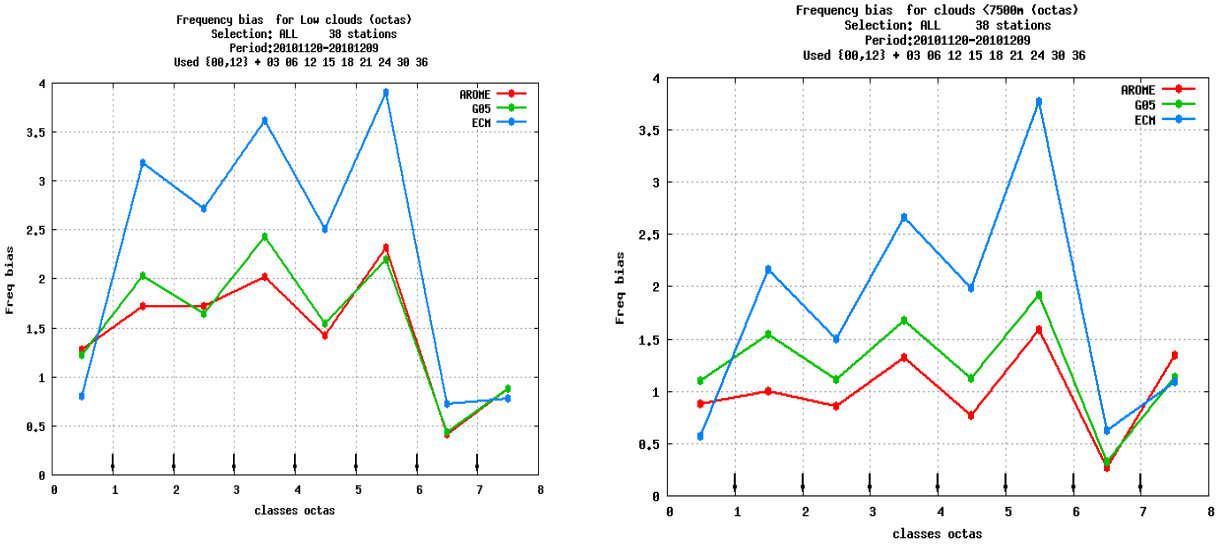

Figure 19: The frequency bias of the forecasts of low cloud (left) and all detectable cloud (right). AROME (red), HIRLAM-7.1.2 (green), ECMWF (blue)

All models, especially ECMWF, predict too much low cloud between 2 and 6 octas. This is also the case for all detectable clouds for ECMWF and to some extent for HIRLAM. AROME has no particular frequency bias for all detectable cloud.

The corresponding ETS values are seen in Figure *20*.

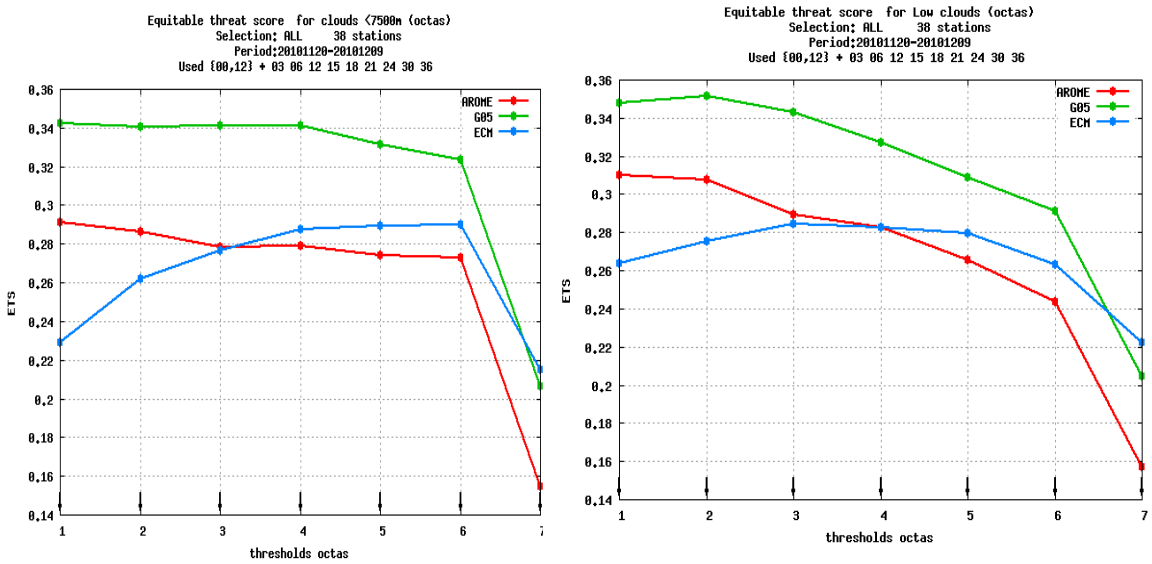

Figure 20: ETS for low cloud (left) and detectable clouds (right). (See explanation of statistical methods in the following chapter.) AROME (red), HIRLAM-7.1.2 (green), ECMWF (blue)

HIRLAM has the highest ETS for both low cloud and detectable cloud. ECMWF and AROME have similar scores. Comparing the spring and winter period it is seen that the ETS values are similar for HIRLAM whereas for ECMWF and AROME the ETS values are higher for the spring period (that means better).

4 Conclusions

The AROME model has been compared with other models with respect to cloud cover and cloud base, including fog. The observations that are used in this study are satellite pictures and automatic stations which are able to detect clouds up to 7.5 km above ground.

- In general, the forecasts of cloud base regarding low cloud are as good as or better than for the other models in this comparison, but there is an over-prediction of low cloud in AROME.
- Also fog is over-predicted.
- There is an under-prediction of high cloud base. It is not possible to say if this is caused by too low amount of high clouds or if high clouds are hidden by low clouds in the model.

AROME has a positive bias of both all detectable cloud and of low cloud. It is not obvious how this should be interpreted. Automatic stations only detect clouds directly above the stations, whereas a much larger part of the sky is included in manual cloud observations. Manual observations in general lead to larger values of the cloud cover due to perspective effects. It is possible that verification against manual observations may give a different result.

The RMS error of low cloud and all detectable cloud is high for AROME compared to other models, but the skill scores are quite good anyway. This is probably due to a larger variation of cloud cover for AROME than for other models. The variation may bee too large.

The only serious problem regarding cloud prediction in AROME is seen for very cold weather. There is an over-prediction of low cloud and fog which affects the 2m-temperature making it too warm. The reason for this problem is not known. A possible reason is an error in the surface, (too large transport of moisture from the ground). Another possible reason is that the vertical diffusion (turbulence) is too weak in this situation. It is also possible that the cloud micro-physics has to be improved such that mixed-phase clouds contain more liquid water.

5 Figures and tables

ANNEX 1

Explanation of common verification scores

In this annex the different verification and evaluation scores are explained in detail. The first part of the annex is copied from 01-2012 METCOOP MEMO (http://metcoop.org/memo chapter 2.10) Abbreviations of the scores are used in the figures and the text in this report.

RMS error (RMSE): $\sqrt{1/N \sum (f(i) - v(i))^2}$.

Means "root mean square error" and is computed as the root of the mean of the forecast, f(i), minus observation, v(i) squared. N is the number of observations.

Characteristics: Measures the correspondence between observations and forecast. Perfect value is zero. Lowering the variability of a forecast may result in a smaller RMS error, without increasing the value of the forecast.

BIAS or systematic error (BIAS): $1/N (\sum (f(i) - v(i))$.

It is computed as the difference between the mean of the forecast and the mean of the observation.

Characteristics: Measures the mean correspondence between observations and forecast. Perfect value is zero. Negative value means an 'under-prediction' of the event, positive value means the opposite.

Standard deviation: $\sqrt{1/N \sum (x(i) - x(mean))^2}$. x may be either a forecast or an observation.

It is the root of the mean of the squared value minus the mean of the value.

Characteristics: Measures the mean variability of the forecast or the observation. The variability of forecasts and observations should normally not differ very much. But exceptions may exist. One example is when a forecast is representing a mean value of grid square with an expected smaller variability than an observation representing a point.

Explanation of skill scores

General definition:
Any forecast verified with statistical measure with the result S may be compared with the result found by using a reference forecast S(ref). This reference forecast could be any forecast based on the same statistics; for instance a random forecast, a climatological forecast or the result from another model.
The skill score is then defined as:
Skill-score = (S - S(ref)) / (S(perfect) - S(ref)) .
S(perfect) is the best possible result that may be obtained in the study. For example it is zero for the RMS error.

One is the best possible result. This is when S equals S(perfect). The skill score is zero if it has the same value as the reference forecast S(ref). Negative values indicate negative skill.

Supplementary scores based on use of contingency tables:

Variables used in contingency tables:

The simplest contingency table consists of only two different observations and forecasts:

	Obs. Severe events	Obs. Non-severe events	Number of forecasts
Forecast severe events	A	b	a+b
Forecast non-severe events	C	d	c+d
Number of observations	a+c	b+d	a+b+c+d=N

A perfect model should have a + d = N and b = c = 0.

From this table many different types of scores may be derived:

Frequency BIAS (FB): (a+b)/(a+c) for severe events and (c+d)/(b+d) for non-severe events.

Characteristics: Measures the BIAS or systematic error of the forecast. No BIAS gives the value one (perfect value), a positive BIAS gives a value above one and a negative BIAS gives a value below one.

F**alse alarm rate (FAR):** b/(b+d)

Characteristics: Measures the number of "alarms" of severe weather compared to the number of the event with no severe weather. Perfect value is zero.

False alarm ratio: b/(a+b)

Characteristics: Measures the number of "alarms" of severe weather compared to the number of the forecast of severe weather. Perfect value is zero.

Probability of detection / Hit Rate (HR): a/(a+c)

Characteristics: Measures the number of correct forecasts of severe weather compared to the number of observations of severe weather. Perfect value is one.

T**reat score:** a/(a+b+c) = a /(N - d)

Characteristics: Measures the number of correct forecasts of severe weather compared to the total number in the sample that do not contain correct forecasts of no severe weather. Perfect value is one.

Different skill scores with some common characteristics:

1: Perfect value is one.

2: A random forecast gives the value zero. Here, a random forecast means a forecast with the same forecast frequency as the tested one, but its values are randomly distributed among the sample. It has the following values for a,b,c and d:

a(random) = (a+b)(a+c)/N , b(random) = (a+b)(b+d)/N , c(random)= (c+d)(a+c)/N and d(random)= (c+d)(b+d)/N

3: A negative value indicates negative skill, but for some of the scores it indicates that the forecast has a 'negative signal', which means that the forecast may have a value if forecasts of severe weather and non-severe weather are replaced with each other.

Equitable treat score (ETS): (a – (a+b)(a+c)/N)/ (a+b+c – (a+b)(a+c)/N).

(Or (a – a(random)) /(a+b+c – a(random)))

Special characteristics:

Good: No large tendency of favouring forecasts with a large positive or negative BIAS.

Poor: No clear relation with the value of the forecasts with respect to cost/loss relations.

Kupiers skill score (KSS):

(Probability of detection) - (False alarm rate) = a/(a+c) – b/(b+d)

Good: Has a clear relation with the value of the forecasts with respect to cost-loss relations.

It represents the economical value of the forecast for an 'optimal' user. It disfavours a forecast with a very large positive or negative bias.

Poor: Favouring a forecast with a moderate positive or negative bias for more extreme event. An example is large amounts of rain, occurring infrequently. Then the 'optimal' user has an unlikely low cost-loss relation, with a very low protection cost. In case of a false alarm, KSS only accounts for this low protection cost. This leads to a higher KSS when heavy rain is forecasted more often than observed.

AI (area index):

(ad-bc)/((b + d)(a + c)) + c / (b + d) ln(N c /((a+c)(c+d))) + b / (a + c) ln(N b /((b+d)(a+b)))

Good: Has a clear relation with the value of the forecasts with respect to the cost-loss relations. It represents the economical value of the forecast for a normally large group of 'optimal' users. Has no large tendency of favouring forecasts with a large positive or negative bias, since it also accounts for the size of the group that is 'optimal' users. In a case such as the one mentioned above, this group is very limited.

Poor: Complicated and not very well known.

In order to a get a value similar to the other scores, one may have to use the square of AI.

Symmetric Extreme Dependency Score (SEDS):

((ln((a+b)/N) + ln((a+c)/N)) + ln(a/N)) - 1

Extremal Dependency Index (EDI):

(ln(b/(b+d)) - ln(a/(a+c))) / (ln(b/(b+d)) + ln(a/(a+c)))

Symmetric Extremal Dependency Index (SEDI):

(ln(b/(b+d)) - ln(a/(a+c)) -ln(1-(b/(b+d))) + ln(1-(a/(a+c))))/(ln(b/(b+d)) + ln(a/(a+c)) +ln(1-(b/(b+d))) + ln(1-(a/(a+c))))

Those three scores have some characteristics in common:

Good: The result does not deteriorate when the observed frequency of the events considered becomes rare as many other scores do. This means that it is less difficult to compare the results of forecasts from different climate regimes. Also the result does not vary so much with long term variation of the observed frequency of the observed phenomena. For example a year with lower occurrences of gale should give mainly the same result as a windier year, if the forecast quality is unchanged.

Poor: Those new scores are less known and complicated.

The best score of these three seems to be the first one, Symmetric Extreme Dependency Score (SEDS) since it has no large tendency of favouring a forecast with a large positive or negative bias. The other two are problematic in that sense.

The cost-loss ratio

The cost-loss ratio (**C/L**) is often used for economical evaluation of forecast value. It is ratio of the cost for protection (**C**) of a severe weather event to the net loss of value (**L**) if the severe weather occurs without protection. The contingency table above becomes (in economical terms):

	Obs. Severe events	Obs. Non-severe events	Forecast related costs
Forecast severe events	a **C**	b **C**	(a + b) **C**
Forecast non-severe events	c **L**	0	c **L**
Observation related costs	a **C** +c **L**	b **C**	(a + b) **C** + c **L**

Using current forecast will give an economical outcome of E(c) = (a + b) **C** + c **L** (as in table)

In other cases somewhat different from the table:

A prefect forecast will give an economical outcome of E(perf) = (a+c) **C**

Newer protecting will give an economical outcome of E(np) = (a+c) **L**

Always protecting will give an economical outcome of E(ap) =(a+b+c+d) **C**

A skill score based on always protecting then becomes (**X = C/L**) :

S(ap) = [E(c) – E(ap)] / [E(perf) – E(ap)] = (c – (c+d) **X**) / (b + d) ,

and based on newer protecting,

S(np) = [E(c) – E(np)] / [E(perf) – E(np)] = (a + b – b /(1-**X**)) / (a + c) .

It can be shown that an 'optimal user' is a user for which the the cost-loss ratio is equal to the frequency of severe weather, (a+c)/N. Then S(ap) = S(np) = Kuipers skill score.

In Figure *21* there is an idealized example in which a = 30 , b= 6, c =10 and d=54.

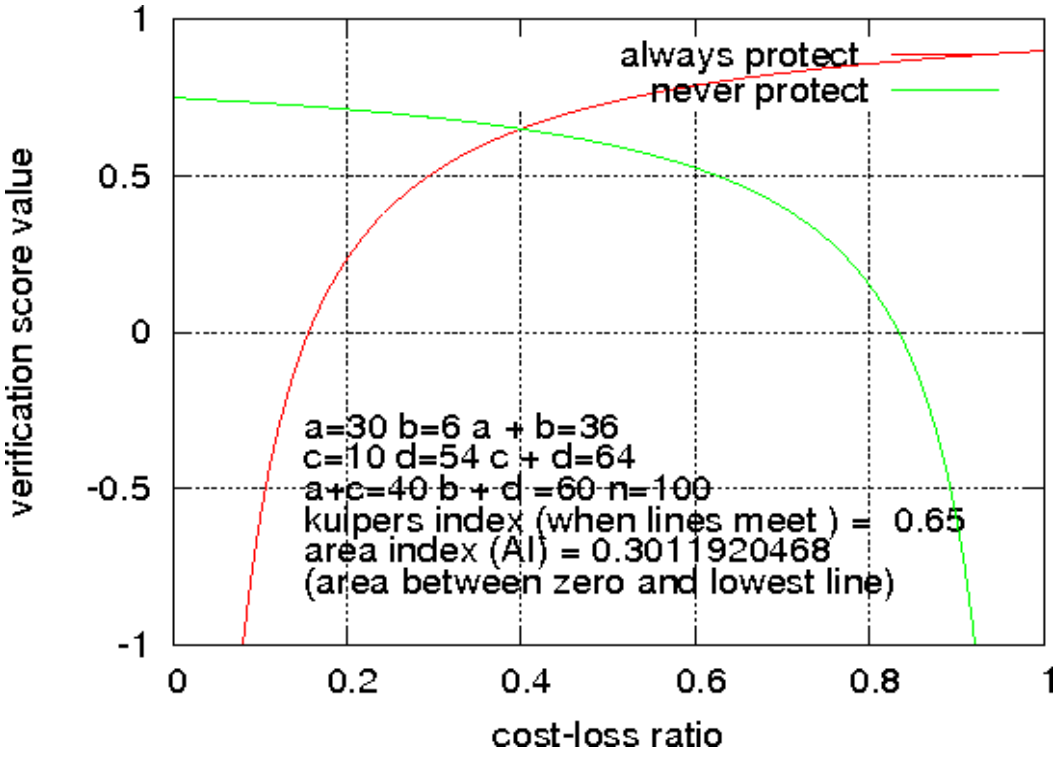

Figure 21: Red curve is the S(ap) curve (always protect) and green is S(np) (never protect) for different cost-loss ratios. a,b,c and d is as mentioned in figure.

Lines meet at the cost-loss ratio 0.4, which is the cost loss ratio for the 'optimal user' in this case. The Kuipers skill score is then 0.65. AI, or area index is the 'positive' area between the lowest of the two scores S(ap) and S(np). In this case it is about 0.30.

Norwegian Meteorological institute
Postboks 43 Blindern, NO 0313 OSLO
Phone: +47 22 96 30 00 Telefax: +47 22 96 30 50

Swedish Meteorological and Hydrological Institute
SE 601 76 NORRKÖPING
Phone +46 11-495 80 00 Telefax +46 11-495 80 01

ISSN: 1893-7519