Norwegian Meteorological Institute met.no

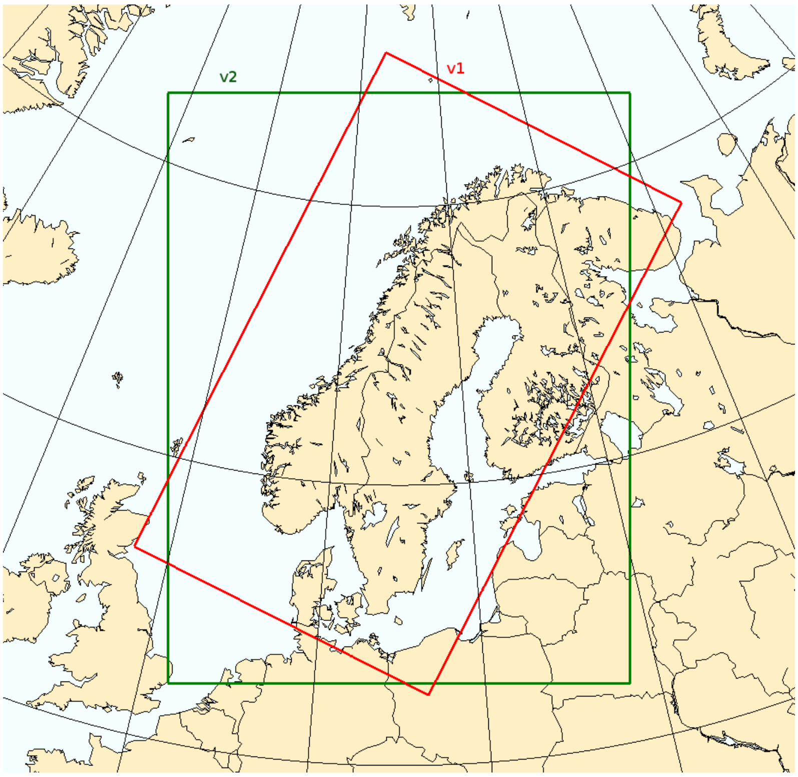
SMHI

METCOOP MEMO No. 02, 2012

Verification study II

Supplementary verification of HARMONIE AROME

Morten A. Køltzow, Karl-Ivar Ivarsson, Dag Bjørge, Solfrid Agersten



Front:
Model domains for the HARMONIE AROME model used in this study.

MetCoOp

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

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

Date: 2012-12-20

ISSN: 1893-7519

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

METCOOP MEMO (Technical Memorandum) No 02, 2012

Verification study II

Supplementary verification of HARMONIE AROME

Morten A. Køltzow, Karl-Ivar Ivarsson, Dag Bjørge, Solfrid Agersten

Summary

This report is a supplement to the previous report 01/2012 METCOOP MEMO “Verification study. HARMONIE AROME compared with HIRLAM, UM and ECMWF”.

Additional verification score, for 24 hr precipitation, for the four test periods used in the first report is given. Precipitation quality measured with Fraction Brier Skill Score shows that AROME is in general among the best models for all precipitation thresholds. In addition, the quality of all models increases with increasing spatial scales. The results also show that HARMONIE AROME do add more value to the ECMWF forecasts for Norway, than for Sweden.

Comparing HARMONIE AROME with a version of the Unified Model on 4 km horizontal resolution from met.no and from UK Met Office reveals the conclusions from the first verification report are still valid; with the exception that UM4 are at a similar or slightly higher quality with respect to mean sea level pressure.

An overview of the differences found between HARMONIE AROME and HARMONIE ALARO are given. In general, the HARMONIE AROME configuration shows slightly smaller errors for MSLP and 10 m wind speed compared with HARMONIE ALARO. With respect to precipitation ALARO shows better results except for very small precipitation amounts. However the number of precipitation events in the models (and especially in ALARO) is greater than that seen in the observations. The HARMONIE AROME and ALARO forecasts of total cloud cover show very different characteristics.

A comparison of two different versions of HARMONIE AROME (cycle 36 and cycle 37) is given. The main difference introduced by the upgrade from cy36 to cy37 for HARMONIE AROME is a decrease in the wind speed, and thereby also a decrease in forecast quality, especially in the mountains.

Finally the comparison between HARMONIE AROME on two different domains (see front page) revealed only minor differences with the exception of a positive impact on precipitation by using a slightly larger integration domain.

Table of contents

1 Introduction

MetCoOp (Meteorological Co-operation on NWP) is a project where the Norwegian Meteorological Institute (met.no) and Swedish Meteorological and Hydrological Institute (SMHI) co-operate in order to have a common production of numerical weather prediction. The goal is to produce and deliver the best short range numerical weather forecasts for a common domain. This verification report is an addition to the first verification report "Verification study. HARMONIE AROME compared with HIRLAM, UM and ECMWF", 01/2012 METCOOP MEMO (Køltzow et.al [3]) which serves as a recommendation for which numerical weather prediction model system to use in the co-operation between SMHI and met.no. This report is a supplement to the previous report and therefore we recommend the reader to familiarize themselves with the contents of 01/2012 METCOOP MEMO [3].

The first chapter in this report gives additional verification scores, for 24 hr precipitation, for the four test periods used in the first report. Afterwards the results of a comparison between HARMONIE AROME and the Unified Model (UM) run by UK MetOffice, and UM run by met.no are given as a supplement to the comparison with UM in the first report.

Chapter four gives an overview of the differences found between HARMONIE AROME and HARMONIE ALARO. In chapter 5 a comparison of two different versions of HARMONIE AROME is given. Finally, in Chapter 6, the differences in the HARMONIE AROME forecasts running on two different domains (see front page) are outlined.

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

2 Verification of 24-hours precipitation using Fractions Brier Skill Score (FBSS)

This chapter is written to as an addition to the verification results of 01/2012 METCOOP MEMO [3]. To forecast precipitation in amount, time and location is difficult, but very important. It is expected that a high-resolution non-hydrostatic model has the ability to resolve the convection and possibly be more accurate in the amount and location of the precipitation. The ability to predict large precipitation amounts for a limited area is not easily shown using the most common verification methods and scores; therefore this chapter is added to show the model skills in this important aspect. If the model smoothes the precipitation over a larger area, as we are used to in the global models, the RMSE can be quite small. This is not necessarily a measure of accurate results for example cases of heavy rainfall over a small area. The spatial distribution of precipitation is by nature not smooth.

2.1 Background

A potential part of the added value for high resolution models is connected to an improved description of spatial variability of the precipitation. This is not necessarily reflected in point by point verification measures, but the ability of a model to forecast precipitation amount at different scales is one important aspect of the quality of high resolution models. Two commonly used methods for measuring this ability are Fractions Briers Skill Score (FBSS) and SAL (Scale, Amplitude, Location). Only verification using FBSS is considered here.

Radar data has often been used as “observations”, for verification with FBSS, but since this data is still somewhat unreliable, a network of rain gauges from climate stations is used in this study. The observation network is not dense enough to catch the finest structures of the precipitation field. Structures of the 'real' precipitation field down to about 30 km (0.3 degrees) will be studied. This value is not definite since the network density of rain gauges varies with geographical location.

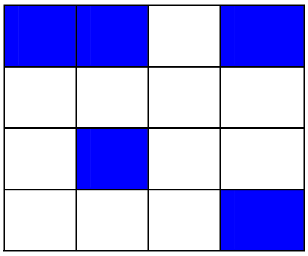
The study is limited to forecasts issued at 00 UTC and valid from 06 UTC to 06 UTC the following day. This means a forecast length of 6 - 30 hours.

2.2 Description of the verification score, FBSS

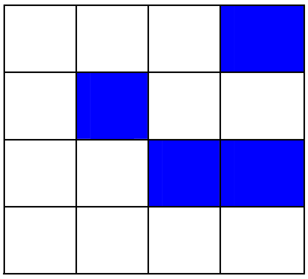
The main principle of the Fractions Briers Skill Score (FBSS) is to check if the forecast has any useful information about the spatial variation of a field by comparing the mean values of forecasts and observations for different sizes of areas. This is done in three steps:

1. Convert the forecast and the observation into one or zero, dependent on if they exceed a predefined threshold (in the given area). The same grid is used for all models. The value from the forecast is calculated by use of bilinear interpolation to every observation point. Squares with less than a prescribed number of observations are not used. There must be at least 3 observations in a square. A consequence of this is that fewer observations are used when squares used are small.

Example:



Forecast, fraction 6/16 (p)



Observed, fraction 4/16 (o)

2. Calculate the mean values of the zeroes and ones. These values are treated as probabilities (frequencies) of exceeding the threshold.

3. The mean square error (commonly called the Brier score (BS), or in this case fractions Brier score, FBS) of the probabilities is calculated.

$$FBS=\frac{1}{N}\sum_{i=1}^{S} M_i(p_i-o_i)^2$$

N is the total number of observations used, S is the number of squares used and M is the number of observations in each square. Finally, p and o are the forecast fraction and the observed fraction respectively. (4/16 and 6/16 in the examples above)

The skill score FBSS is the FBS for a specific model forecast compared with the FBS for a reference model.

$$FBSS=\frac{FBS-FBS_{reference}}{0-FBS_{reference}}=1-\frac{FBS}{FBS_{reference}}$$

For more information about this method, see Roberts and Lean (2008) [4].

2.3 Description of the figures

The reference forecast in this study is ECMWF (~16 km grid). In the figures, FBSS is shown (from -1 to 1) on the vertical axis.

FBSS > 0 means that the test forecast is better than the ECMWF forecast. FBSS < 0 means that the test forecast is worse than the ECMWF forecast. The horizontal axis shows the area as the size of the square in degrees (1 degree is about 111 km).

The forecast models used in this verification are the same as previously described in the prior verification study, see 01/2012 METCOOP MEMO [3]. In addition results for ALARO are included.

Explanation of the legend:
ARB= AROME (2.5 km grid)
A05= ALARO (5.5 km grid),
G05= HIRLAM 7.1.2 (5.5 km grid)
UM4 = Unified model with 4 km grid.

When the lines in the diagram are smooth, random effects are not so likely.

The figures show the ability to forecast precipitation amounts above a given threshold for a given area. The text above the graphs specifies the different thresholds per 24 hour.

2.4 Verification results

This chapter will summarize the results of the verification of 24 hr precipitation using the FBSS verification score for the different periods, and for all periods together.

2.4.1 August 2011

This was a period with some exceptionally large amounts of rainfall over both Norway and Sweden, For more information see 01/2012 METCOOP MEMO [3] chapter 3.4.

The number of used observations when the smallest squares (0.3 deg or 33 km) are used is in table 1. The corresponding numbers of observed precipitation amounts above a certain threshold for each country are given in table 1. There are some more observations for larger squares, so the table shows the minimum number of observations used. The results for Fractions Brier Skill Score (FBSS) for this period are seen in Figure 1 (Norway) and Figure 2 (Sweden).

Table 1.

Threshold	Norway (total 2395 obs.)	Sweden (3796 obs.)
0.1 mm	1395	2031
2 mm	868	1112
5 mm	532	712
10 mm	260	428
20 mm	87	209
35 mm	36	68

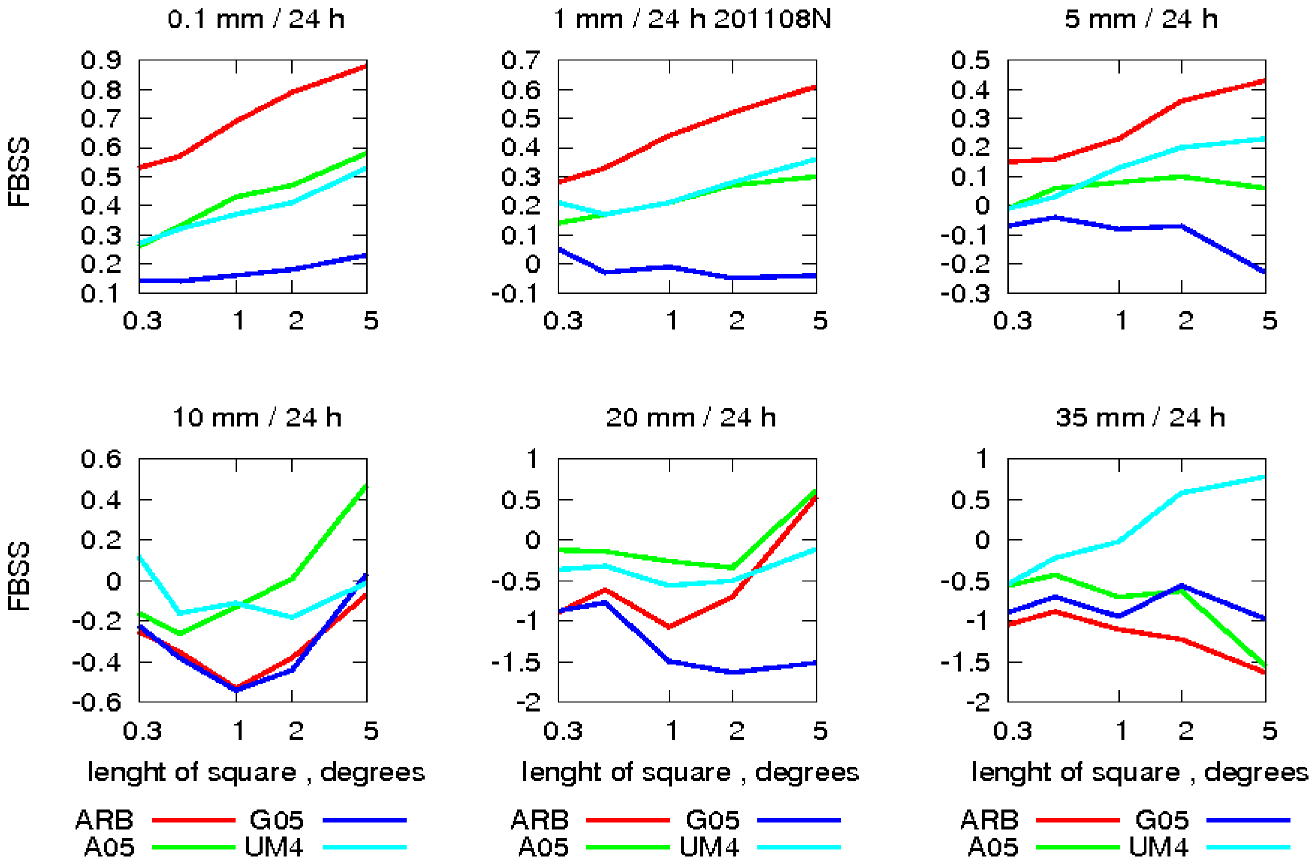


Figure 1: FBSS for different forecasts and thresholds of precipitation over Norway for August 12-23 2011. For explanation, see chapter 2.3.

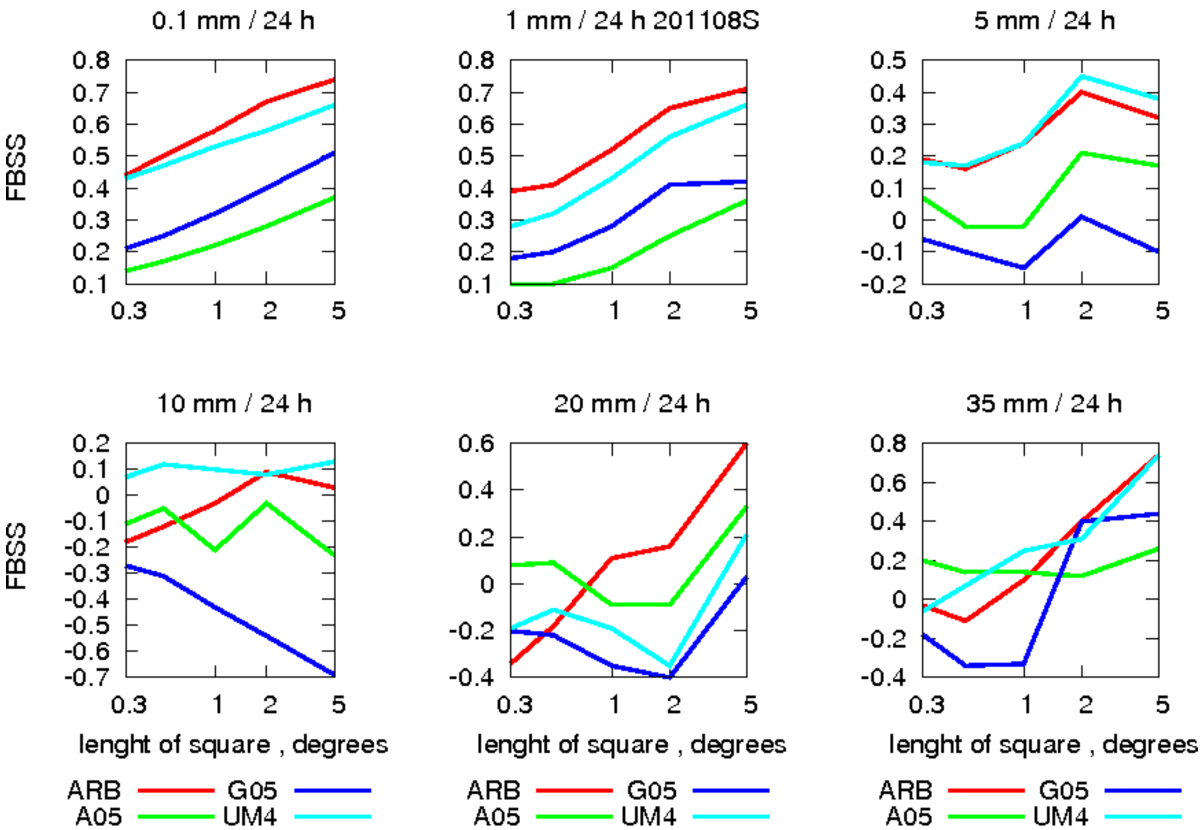


Figure 2: FBSS for different forecasts and thresholds of precipitation over Sweden for August 12-23 2011. For explanation, see chapter 2.3.

The figures show that AROME was the best model for the lowest precipitation thresholds over Norway and Sweden. The result is more variable for higher thresholds. Probably random effects play a role here, since the number of cases with high precipitation is rather few as seen in table 1. In many cases ECMWF is the best model for the higher thresholds.

2.4.2 December 2011 – January 2012

This period was dominated by strong westerly or south-westerly winds. This means that most of the precipitation was over Norway. For more information see 01/2012 METCOOP MEMO [3] chapter 3.5.

The results for the two countries are seen in Figure 3 (Norway) and Figure 4 (Sweden).

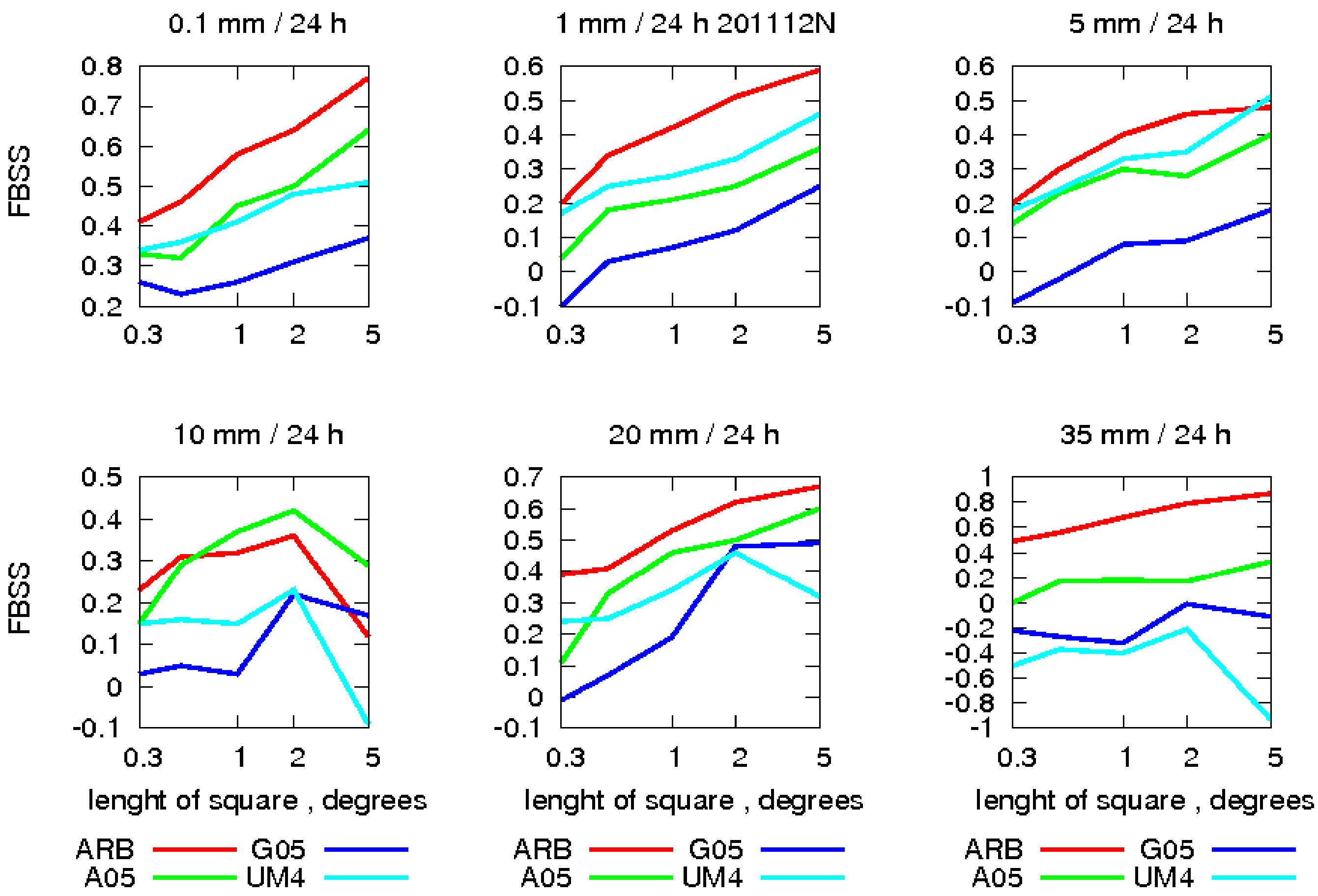


Figure 3: FBSS for different forecasts and thresholds of precipitation over Norway for December 20 2011 to January 7 2012. For explanation, see chapter 2.3.

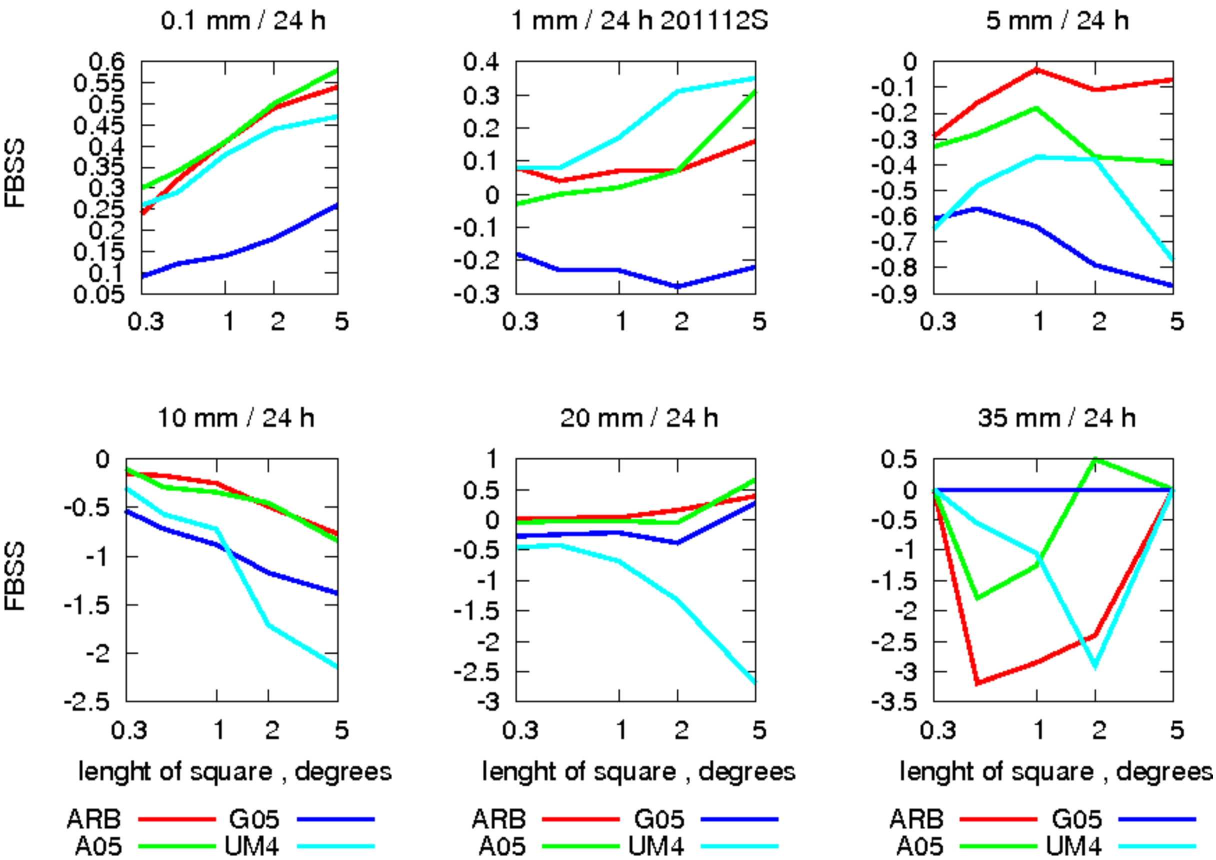


Figure 4: FBSS for different forecasts and thresholds of precipitation over Sweden for December 20 2011 to January 7 2012. For explanation, see chapter 2.3.

AROME has the best results for Norway. HIRLAM has the worst result, but is often better than ECMWF.

ECMWF is the best model over Sweden for higher thresholds. For the small precipitation amounts (0.1 mm and 1 mm/24 hr) AROME, ALARO and UM4 are very alike. The drier conditions over Sweden cause random effects for larger precipitation amounts than for Norway, see table 2 below.

The number of used observations when the smallest squares (0.3 deg or 33 km) are used is in table 2. It shows the same type of numbers as table 1, but for December 2011 – January 2012.

Table 2.

Treshold	Norway (total 3826)	Sweden (total 5501)
0.1 mm	2899	3529
2 mm	2059	1809
5 mm	1474	955
10 mm	928	378
20 mm	390	53
35 mm	125	0

Zero occasions of 35 mm for Sweden are only the case when the smallest square-size is used. For 0.5 degrees and larger it is 4 occasions.

2.4.3 March 2012

The weather in this period was also dominated by winds from south-west. For more information, see 01/2012 METCOOP MEMO [3] chapter 3.6.
The results of Fractional Brier Skill Score for Norway and Sweden are seen in Figure 5 and 6 respectively. The number of precipitation exceeding the different thresholds used for this period is in table 3. (Squares of 0.3 degrees.) There are more cases when larger squares are examined. For the largest squares, 5 degrees (about 550 km) there are 3 cases with more than 35 mm, 19 with more than 20 mm and 121 with more than 10 mm for Sweden. For 0.5 degrees (55 km), the corresponding numbers are 1,7 and 71 respectively.

Table 3.

Treshold	Norway (total 3198)	Sweden (total 4522)
0.1 mm	1917	1308
2 mm	1261	340
5 mm	947	59
10 mm	616	3
20 mm	281	0
35 mm	91	0

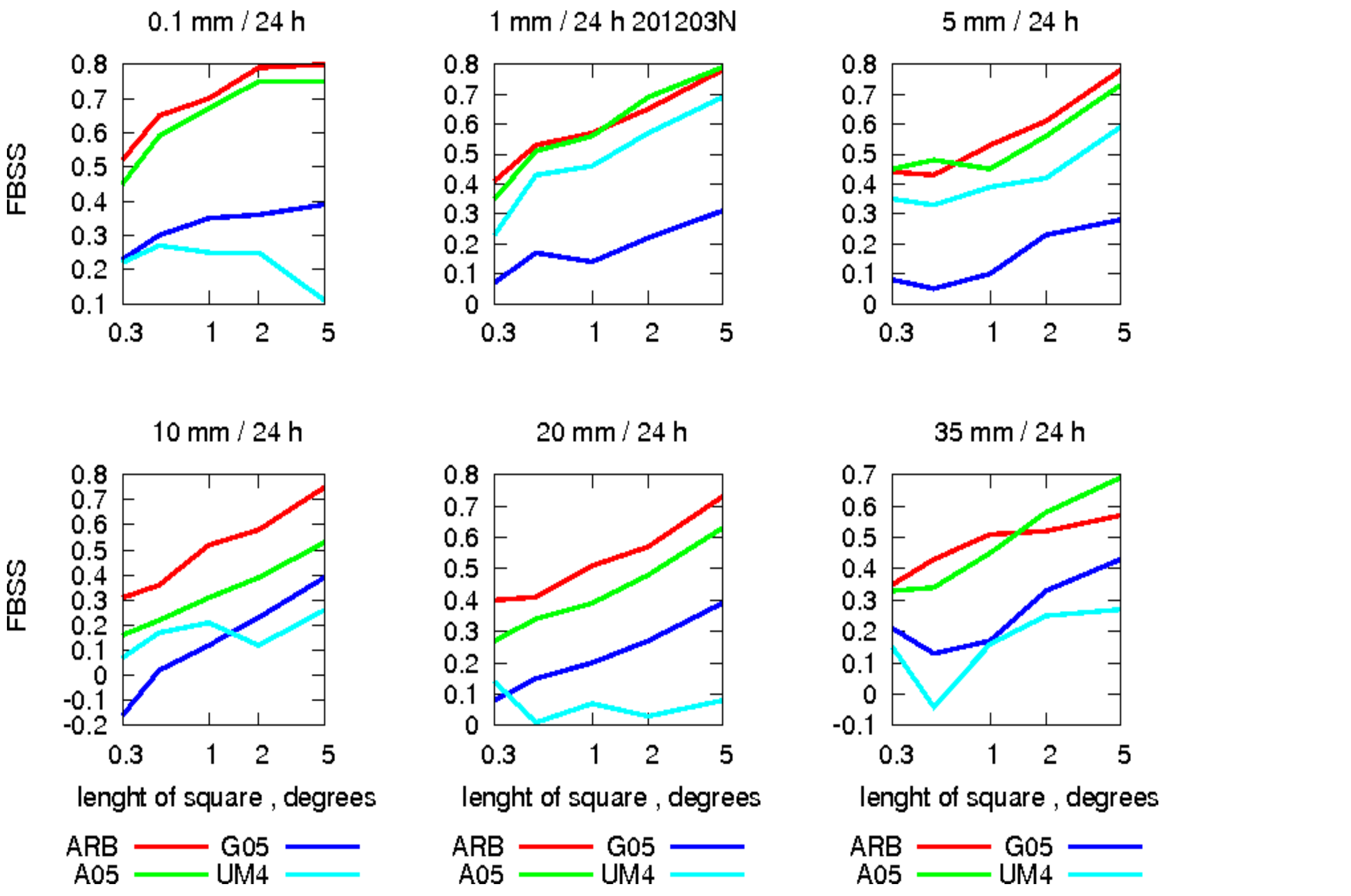


Figure 5: FBSS for different forecasts and thresholds of precipitation over Norway for March 3-23 2012. For explanation, see chapter 2.3.

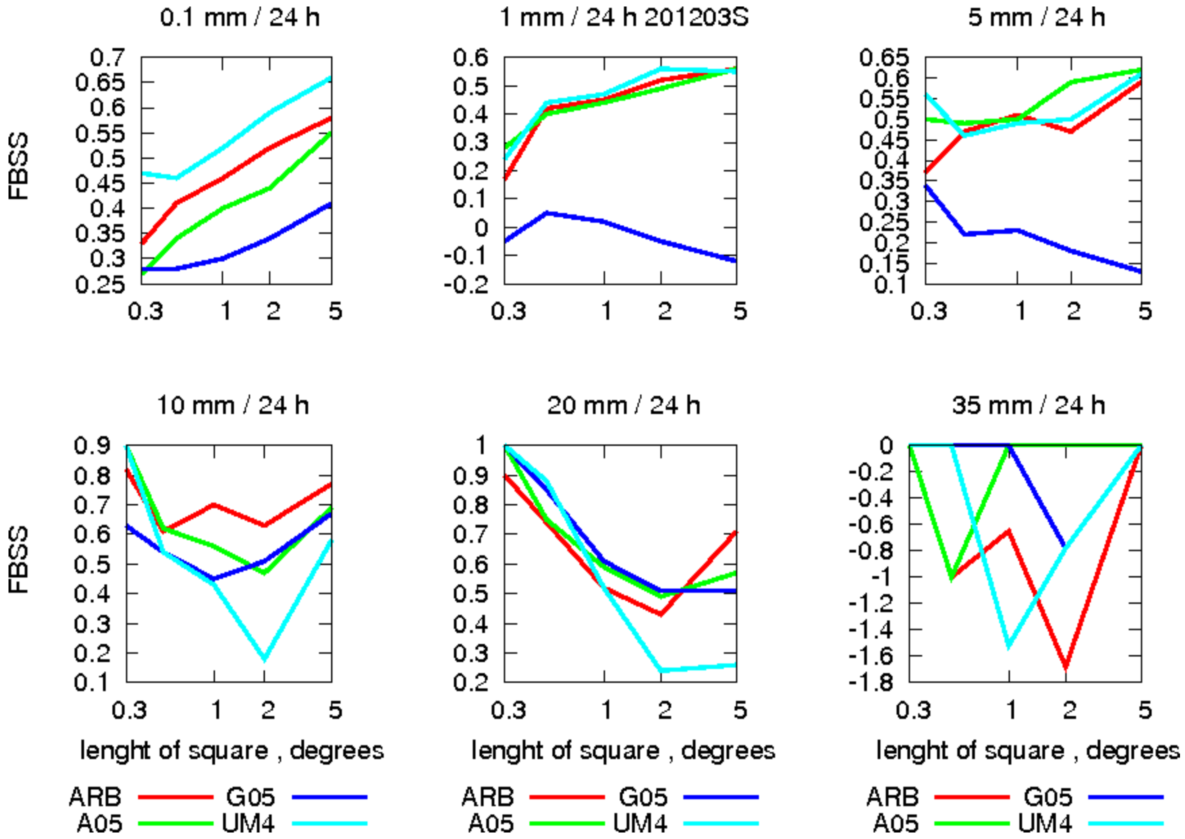


Figure 6: FBSS for different forecasts and thresholds of precipitation over Sweden for March 3-23 2012. For explanation, see chapter 2.3.

The results resemble to some degree that for December-January, AROME is often the best of the models over Norway, and always scores better than ECMWF.

The results differ from the December-January period in that most models have a better score than ECMWF. There are very few cases with more than 35 mm over Sweden (zero for 0.3 degrees) so this result is very unreliable.

2.4.4 May 2012

There were variable weather conditions during the May period. For more information, see 01/2012 METCOOP MEMO [3] chapter 3.7.
The results for Norway and Sweden are shown in Figure 7 and Figure 8 respectively and the number of precipitation exceeding the threshold used in this verification for this period is in table 4.

Table 4.

Treshold	Norway (total 5336)	Sweden (total 8023)
0.1 mm	2572	3013
2 mm	1351	1508
5 mm	806	829
10 mm	416	356
20 mm	125	58
35 mm	50	5

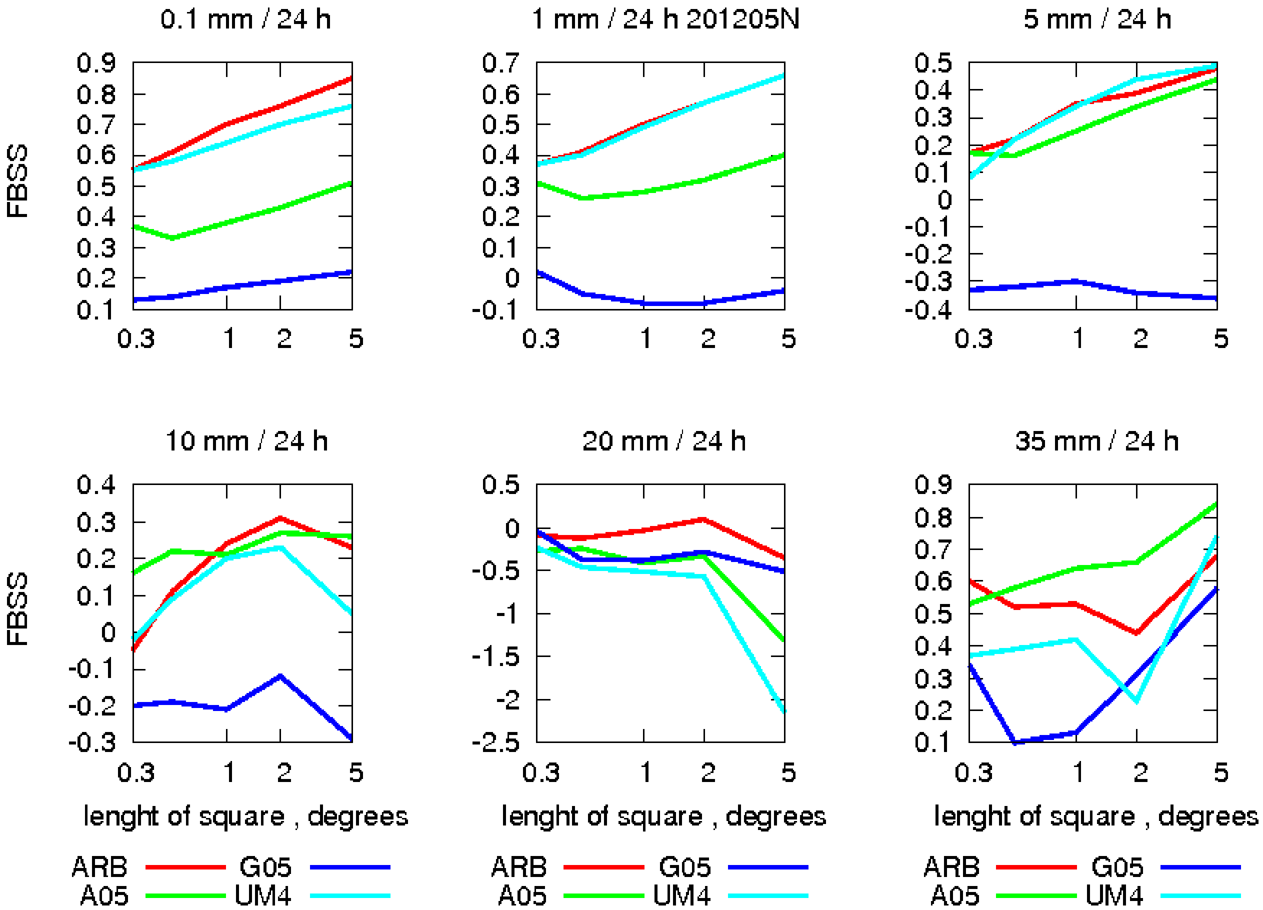


Figure 7: FBSS for different forecasts and thresholds of precipitation over Norway for May 2012. For explanation, see chapter 2.3.

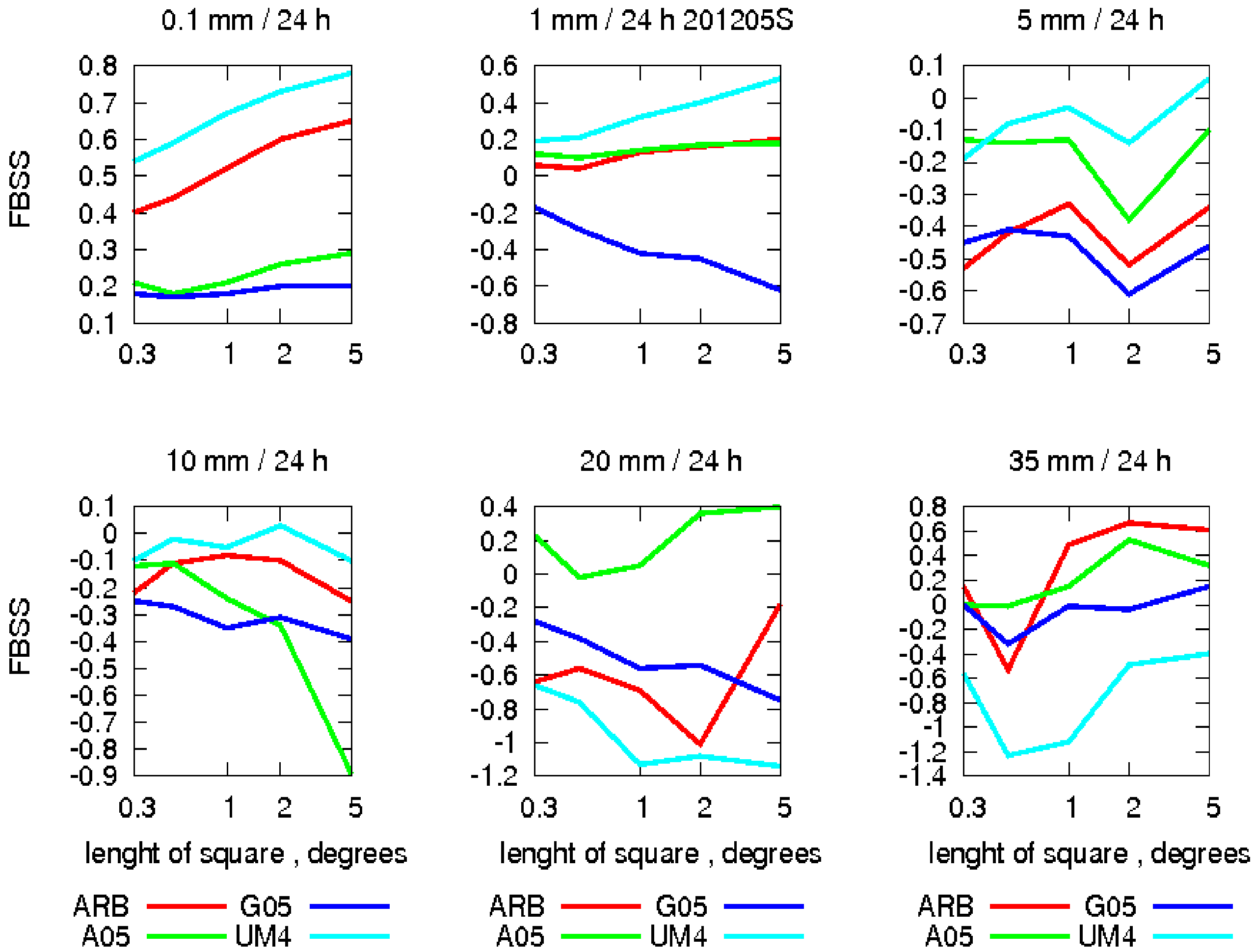


Figure 8: FBSS for different forecasts and thresholds of precipitation over Sweden for May 2012. For explanation, see 2.3.

AROME and UM4 are the two best for Norway for low precipitation thresholds. For higher amounts AROME is among the better.

Over Sweden UM4 is the best up to 10 mm. For 20 mm and 35 mm the results are variable and are probably affected by random effects. Few occurrences of heavy rainfall probably affect the results.

2.4.5 All periods

It is clear from the previous figures that the results vary for the different periods, from case to case. This may be weather related, due to few observations or short periods. Following are the combined results for all periods.

The number of precipitation exceeding the different thresholds for 0.3 degrees used for all periods together is in table 5. There are only 73 cases with more than 35 mm over Sweden for squares of a size of 0.3 degrees. For 0.5 degrees or larger it is 157 to 188. This means that one has to consider the statistical uncertainty for more than 35 mm over Sweden also when all results are put together, at least regarding the result for 0.3 degrees.

Table 5.

Treshold	Norway (total 14746)	Sweden (total 21842)
0.1 mm	8780	9881
2 mm	5538	4769
5 mm	3758	2555
10 mm	2219	1165
20 mm	883	320
35 mm	302	73

In Figure 9 the result for Norway for all periods together is shown, and for Sweden in Figure 10.

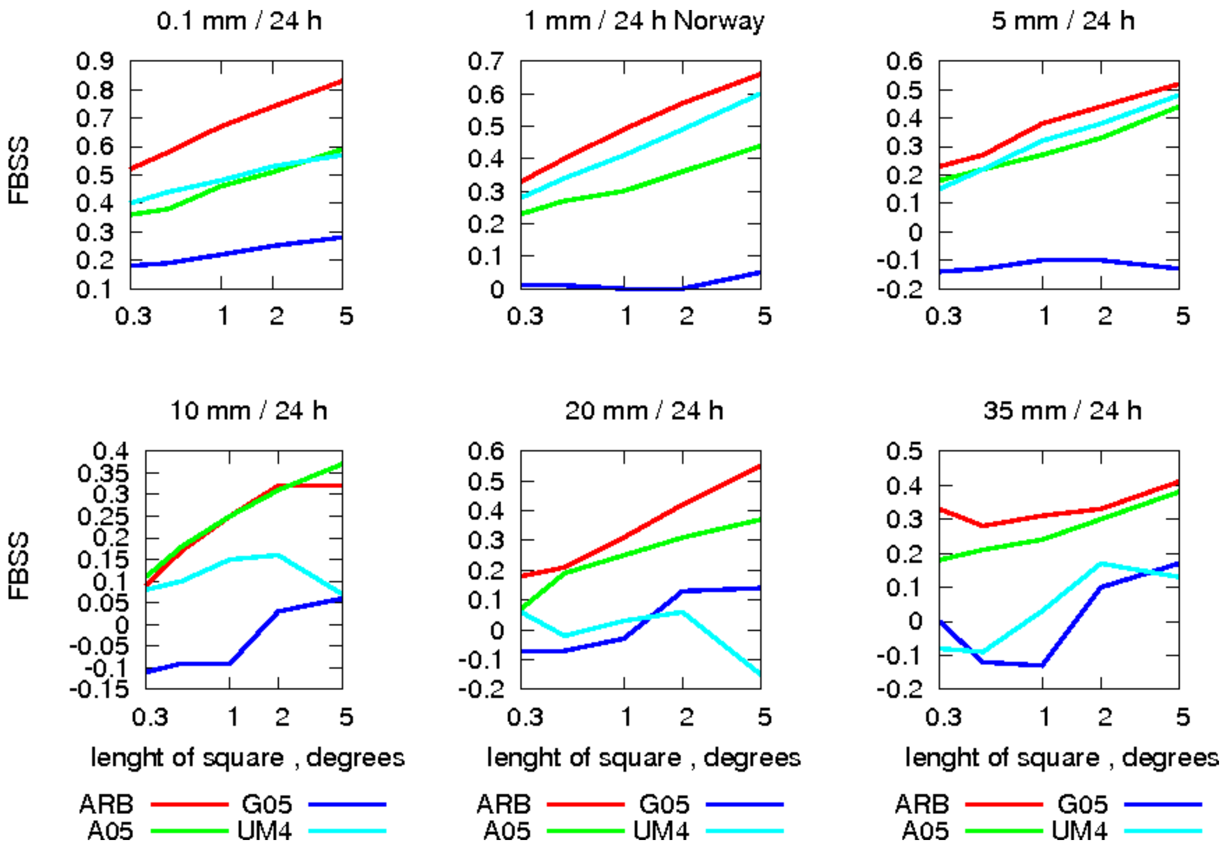


Figure 9: FBSS for different forecasts and thresholds of precipitation over Norway for all cases together. For explanation, see chapter 2.3.

AROME is the best model for all detectable scales and thresholds over Norway, except for 10 mm. The second best is either ALARO or UM4. HIRLAM is only better than ECMWF for 0.1 mm. Positive FBSS for AROME for all thresholds and all spatial scale implies that the use of AROME does add value to the forecast from the coarse resolution global ECMWF model.

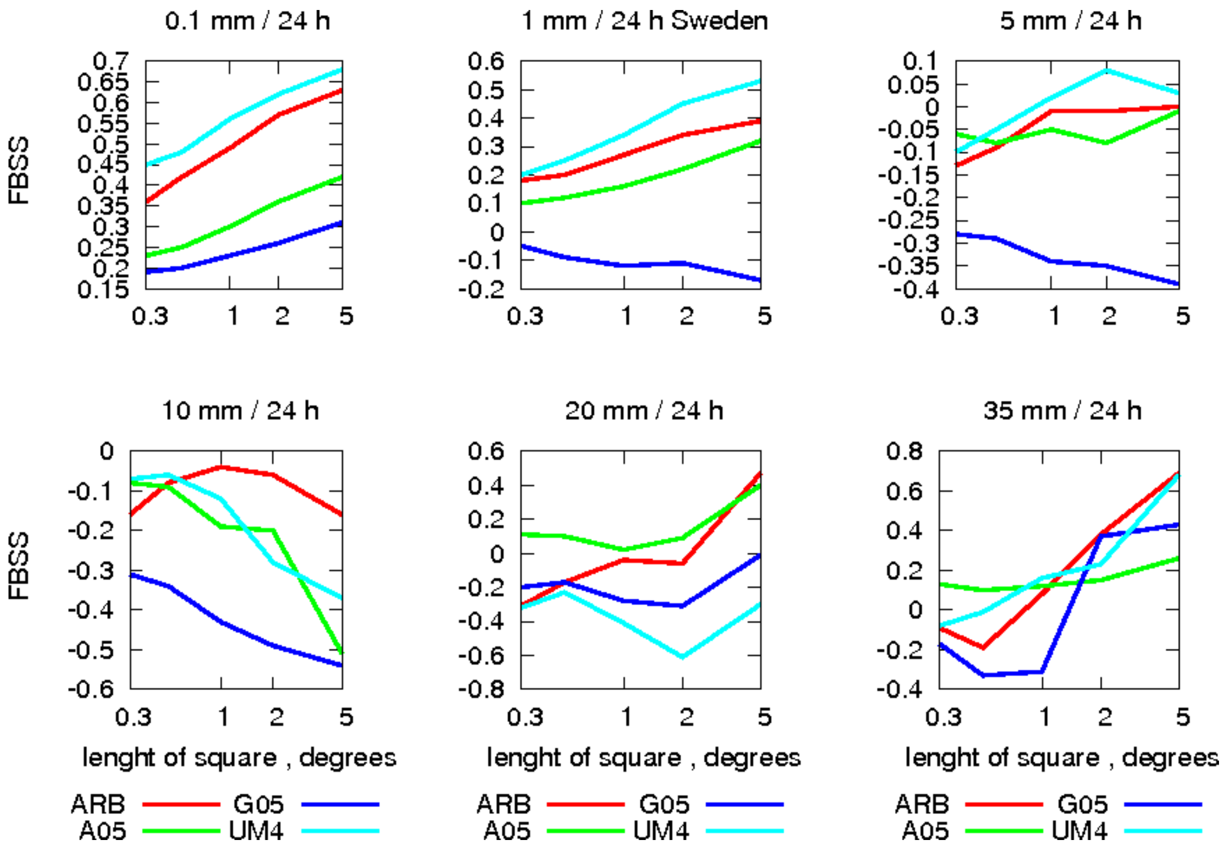


Figure 10: FBSS for different forecasts and thresholds of precipitation over Sweden for all cases together. For explanation, see chapter 2.3.

The result for Sweden is somewhat different than for Norway. For small amounts the best result is found for UM4, but the opposite is seen for the largest amounts. HIRLAM is only better than ECMWF for 0.1 mm. For 20 mm and 35 mm there are rather mixed results. AROME is generally one of the best models for these periods (best for 10 mm), and adds value compared to ECMWF for small precipitation amounts.

It is important to recognise that the results indicate at which scales the limited area models add value to the ECMWF model, but do not tell at which scales the model results are suitable for use. One way of examining this is to compare with a climatological forecast. This is seen in Figure 17 where all cases and both countries are used. The 'true' climatological frequencies are not known, so the 'sample climate' is used here. A 'sample climate' forecast is when the relative frequency of occurrence of the particular event **in the present data-set** is used as a 'forecast'. Those frequencies are to the right in table 6.

Table 6.

Treshold	Norway + Sweden (total 36674)	Relative frequency (used as 'sample climatology')
0.1 mm	18692	0.5097
2 mm	10326	0.2816
5 mm	6324	0.1724
10 mm	3391	0.0925
20 mm	1203	0.0328
35 mm	373	0.0102

Example: 'sample climatology' for 10 mm is then 33891/ 36674 = 0.0925 Thus, the forecasts in this test are compared with a forecast that always 'predicts' that 9.25% of a square of 33 km has more than 10 mm of rain.

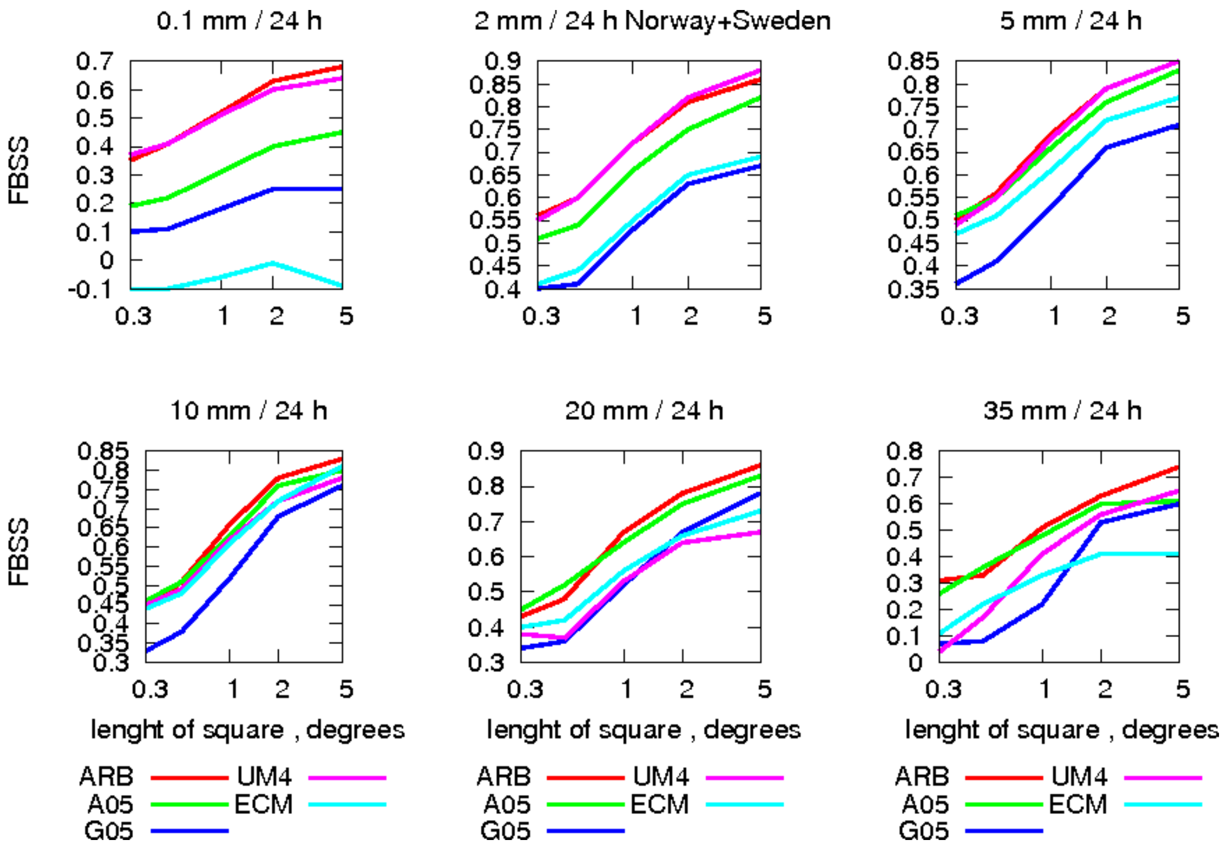


Figure 11: FBSS for different forecasts and thresholds of precipitation for all periods together and both countries. Reference forecast is 'sample climate.'

ECMWF has negative skill for the 0.1 mm threshold, which indicates that there seems to be no useful information of the spatial distribution of 'no rain' contrary 'any rain'. AROME and UM4 are the two best for this threshold. ECMWF has much higher quality for higher thresholds than 0.1 mm and is nearly as good as AROME and ALARO for 10 mm. All models have useful information for all scales studied here for 2 mm or higher thresholds. But for 35 mm the added value for HIRLAM and UM4 is low for squares of 0.3 degrees.

3 HARMONIE AROME compared to the Unified Model (UM)

3.1 Background

In the MetCoOp project it is important to show that quality of the chosen version of the HARMONIE model system has a similar or better quality than other model systems for Sweden and Norway. In the first MetCoOp report (01/2012 METCOOP MEMO [3]) the HARMONIE AROME model system (cy36h1.4) was compared with model results available for operational forecasts at met.no and SMHI. The models were the high resolution version (deterministic) ECMWF model (~16 km horizontal resolution), a HIRLAM version employed at SMHI on 5 km horizontal resolution and a version of the Unified Model with 4 km horizontal resolution (UM4) employed at met.no. The latter model system has over several years proven its strengths and weaknesses in operational use at met.no. However, there has been awareness that the met.no version of the UM4 model system has not been optimally tuned and forced. At met.no the UM4 simulations are a dynamical downscaling of HIRLAM simulations at 8 km horizontal resolution (HIRLAM8). It is especially challenging to interpolate soil variables between two models with different representation of soil processes (i.e. Kristiansen et al. 2012 [1]). For two of the test periods used in the first MetCoOp report (March and May 2012) the UK Met Office has provided us with results from their implementation of the Unified Model with 4 km horizontal resolution on a domain as similar as possible to the Hires1 domain. The main differences between the met.no (hereafter named UM4) and the UK Met Office (hereafter named MO) set-ups are:

Nesting strategy:

- MO is nested in the UK global model and runs surface assimilation.
- UM4 is nested in the global ECMWF model via met.no HIRLAM 8 km and runs without surface assimilation.

Different versions of UM:

- MO is based on vn7.9
- UM4 is based on the older vn7.7.
- There are differences in the schemes for radiation, boundary layer and soil temperature.
- There are slightly different domains

To ensure that the comparison of HARMONIE AROME (hereafter named AM_Hires1) and the Unified Model in the first MetCoOp report has a general validity we include in this report a comparison of AM_Hires1 and the two UM versions (UM4 and MO) from the two available test periods.

3.2 Results

The presented results are averaged over March (5th-23th) 2012 and the May (1th-31th) 2012 periods. For more information about the periods see 01/2012 METCOOP MEMO [3], chapter 3.6 and 3.7. The periods have also been investigated separately and differences between March and May will be commented on.

3.2.1 Verification results of mean sea level pressure (MSLP)

For MSLP the MO shows a better result than UM4 with respect to RMSE and BIAS. AM_Hires1 has smaller errors than UM4, the MO has equal or slightly smaller errors than AM_Hires1. For MSLP there might therefore be a shift in the conclusions from the first verification report; AM_Hires1, based on these periods is not necessarily better than the Unified Model with respect to MSLP.

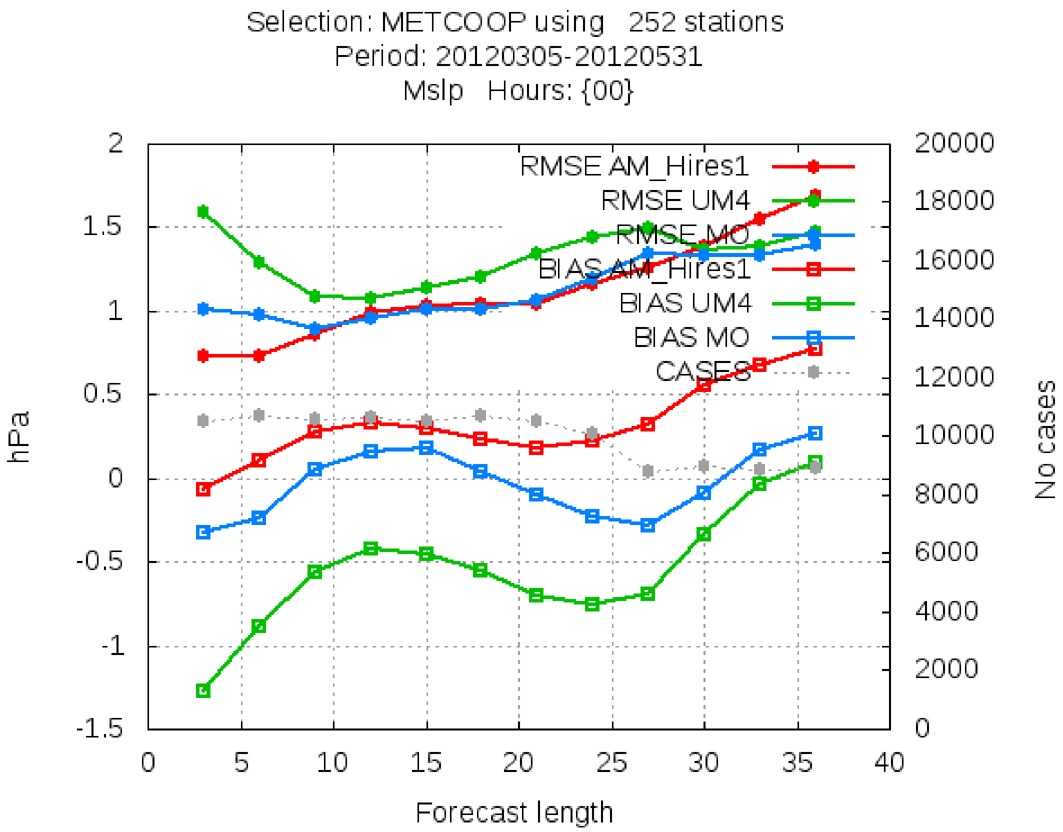


Figure 12: MSLP, BIAS and RMSE as a function of lead time for AM_Hires1 (red), UM4 (green) and MO (blue) averaged over the March and May periods.

3.2.2 Verification results of 10 m wind speed

For high wind speeds UM4 give higher ETS score than MO due to a higher hit rate (despite a higher false alarm ratio). AM_Hires1 scores for all wind speeds are equal or better than MO. All models underestimate the observed frequency of high wind speeds. However, AM_Hires1 and UM4 are in better agreement with the observations than MO. In May, UM4 is equal or better (higher ETS) than MO for all wind speeds, while in March UM4 is better than MO for high wind speeds, but shows a lower ETS score for small wind speeds (not shown). The conclusion that AM_Hires1 for the investigated periods is equal or better than the Unified Model with respect to high wind speeds is still valid.

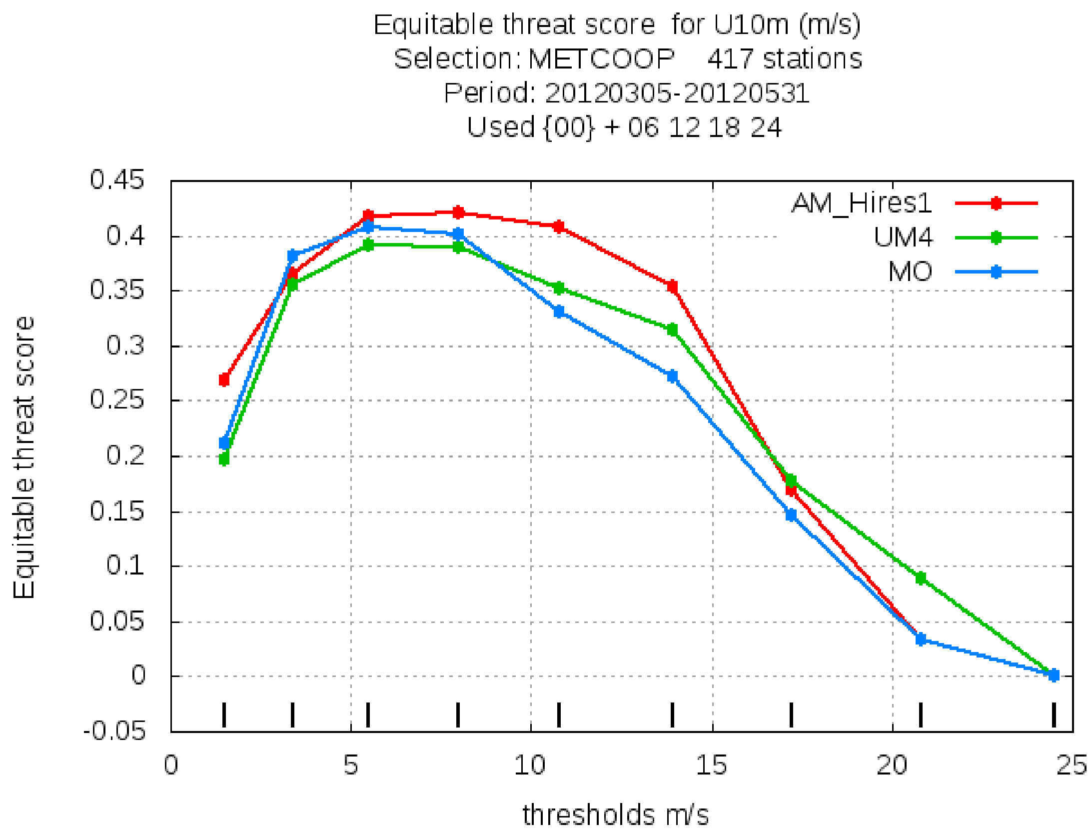


Figure 13: ETS for different thresholds for AM_Hires1 (red), UM4 (green) and MO (blue) averaged over the March and May periods.

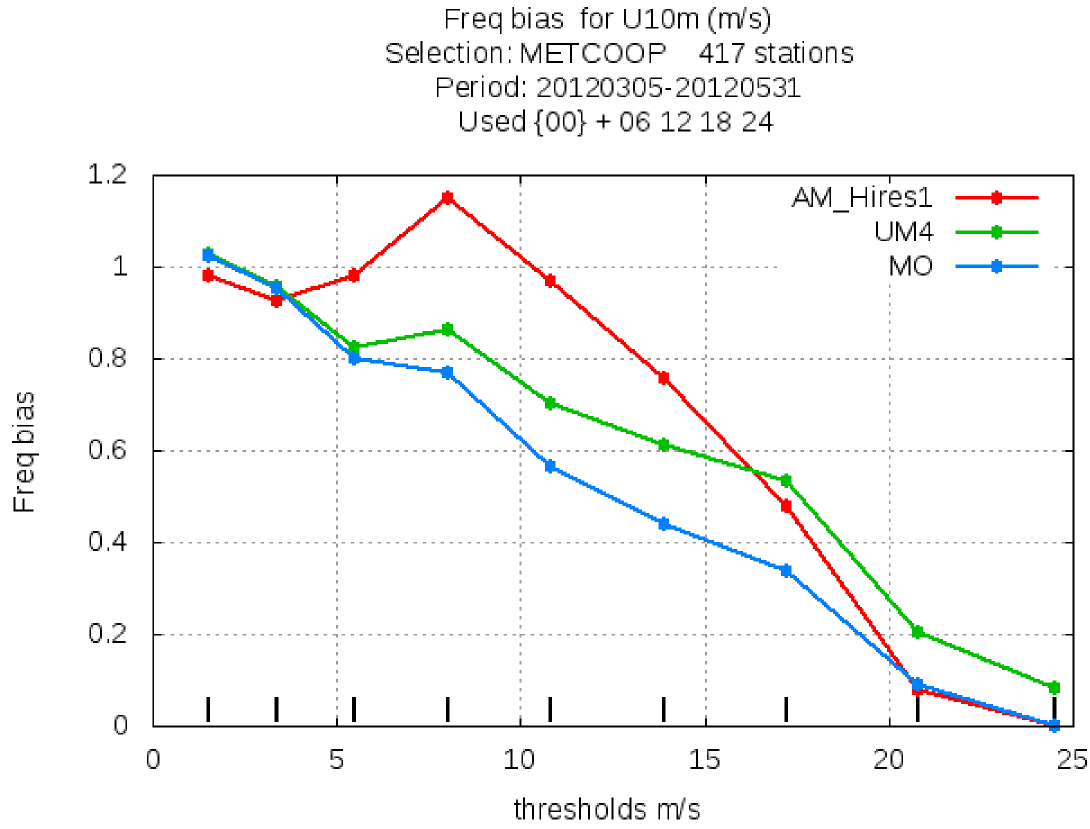


Figure 14: Frequency bias for different thresholds for AM_Hires1 (red), UM4 (green) and MO (blue)

3.2.3 Verification results of 2-metre temperature (T2m)

In Figure 15 the BIAS and RMSE for T2m is shown. The model temperatures used in the verifications are not height corrected to station height. A slightly higher error due to a coarser grid is therefore expected in the 4 km models compared with HARMONIE on 2.5 km horizontal resolution.

The near surface temperature is better in MO than in UM4 (clearly better in March and slightly better in May). However, AM_Hires1 has smaller RMSE than MO and the conclusion that AM_Hires1 for the investigated periods is better than the Unified Model with respect to near surface temperature is still valid.

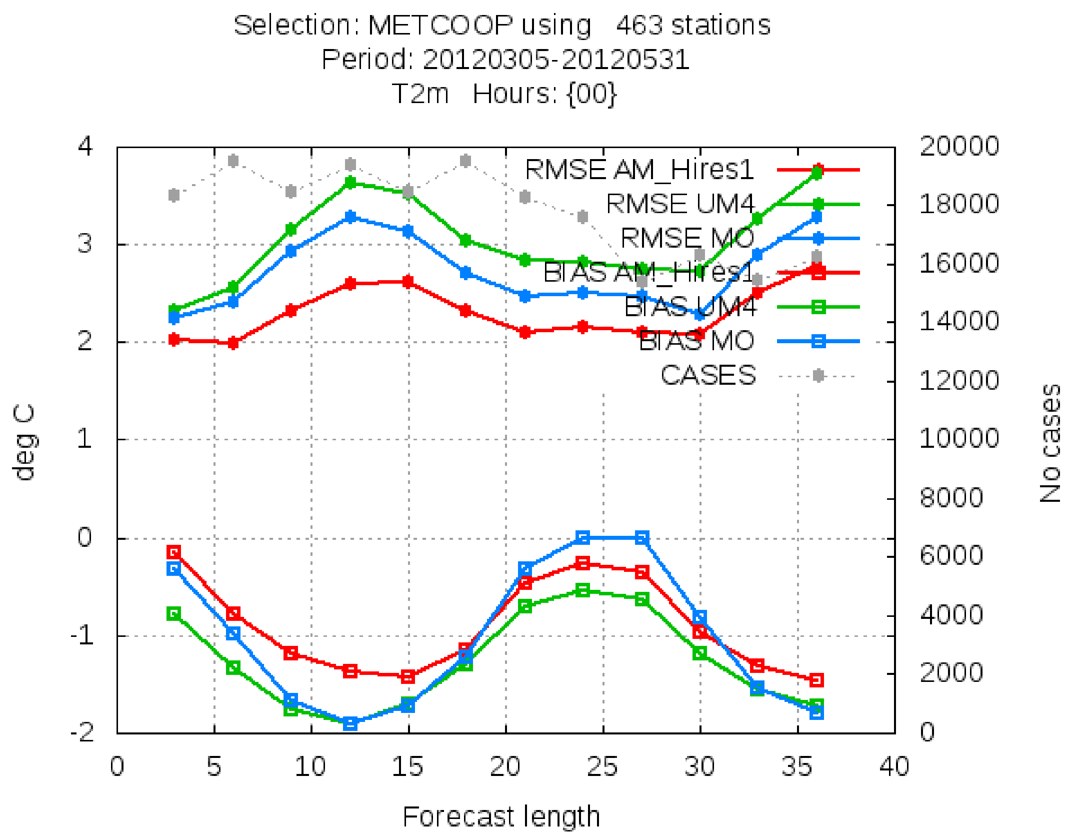


Figure 15: Near surface temperature, BIAS and RMSE as a function of lead time for AM_Hires1 (red), UM4 (green) and MO (blue) averaged over the March and May periods.

3.2.4 Verification results of 12 hr accumulated precipitation.

A comparison between the three models reveals some differences between Norway and Sweden. Averaged over Norwegian stations (Figure 16) UM4 shows higher ETS than MO for all precipitation amounts (up to 10 mm/12 hr). MO has higher hit rate and false alarm ratio than UM4 (not shown). These features are present both in March and May. The ETS score for AM_Hires1 lies between UM4 and MO. AM_Hires1 is nearly equal to UM4 on small amounts (<0.5 mm/12 hr), slightly worse on medium amounts (0.5 - 5 mm/12 hr) and has a slightly higher score on 10 mm/12 hr.

In Sweden, both Unified Model versions are on average very similar in quality: MO has a higher ETS (Figure 17) than UM4 in March and a lower ETS than UM4 in May. The Unified Model versions are better than AM_Hires1 up to 2 mm/12 hr. For larger precipitation amounts there are large spread in the results due to few precipitation events. MO and AM_Hires1 show higher *hit rates* than UM4. UM4 has the best false alarm ratio and AM_Hires1 has the highest false alarm ratio (not shown).

The comparison does not change the general conclusion on precipitation forecasts in the first verification report. However it is difficult to draw firm conclusions based on these short periods.

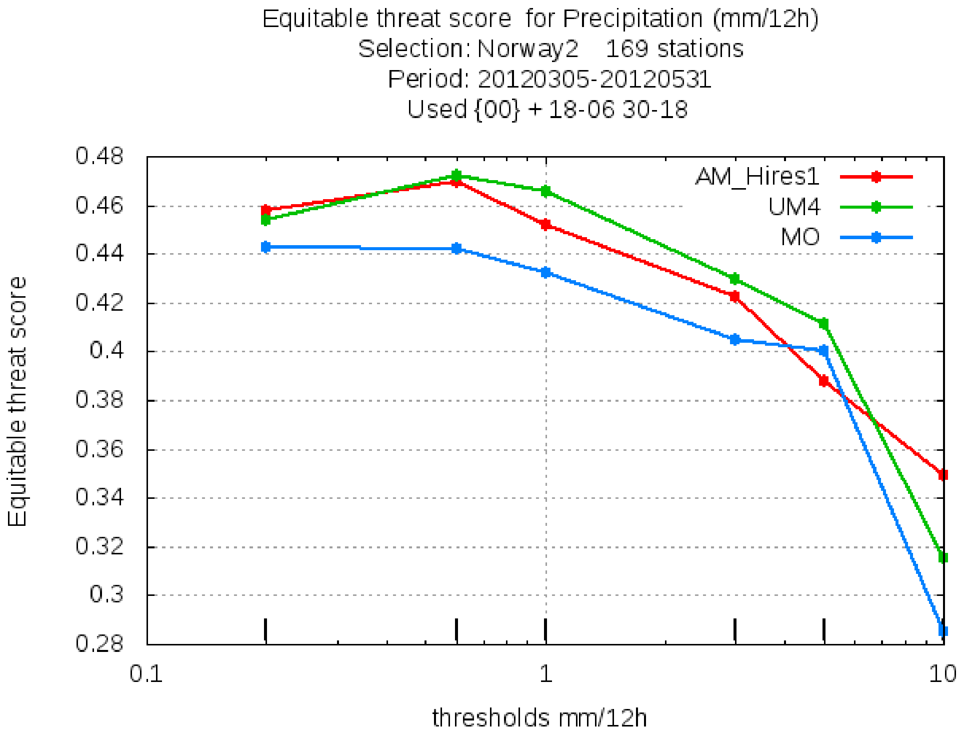


Figure 16: 12 hr accumulated precipitation, ETS as a function of precipitation amounts in forecasts for AM_Hires1 (red), UM4 (green) and MO (blue) averaged over the March and May periods for Norway

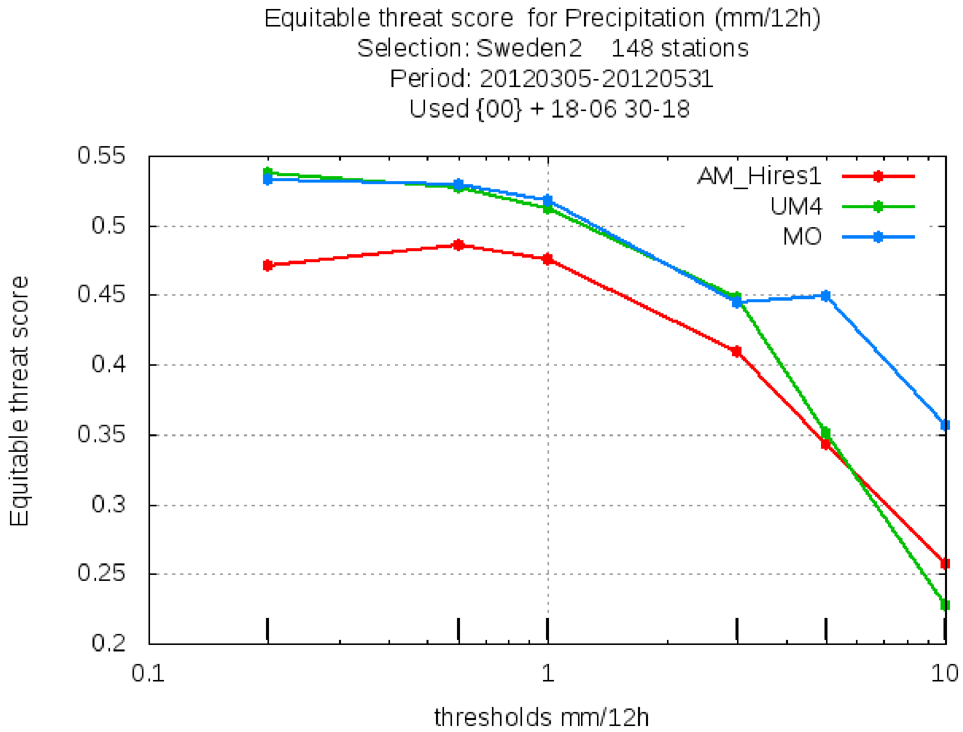


Figure 17: 12 hr accumulated precipitation, ETS as a function of precipitation amounts in forecasts for AM_Hires1 (red), UM4 (green) and MO (blue) averaged over the March and May periods for Sweden.

3.2.5 Verification results of total cloud cover (TCC)

For total cloud cover, MO verifies slightly better than UM4. UM4 shows slightly better result than AM_Hires1 in terms of ETS, for Norway (Figure 18) and Sweden (Figure 19). This comparison does not change the general conclusions about the quality of AM_Hires1 cloud cover.

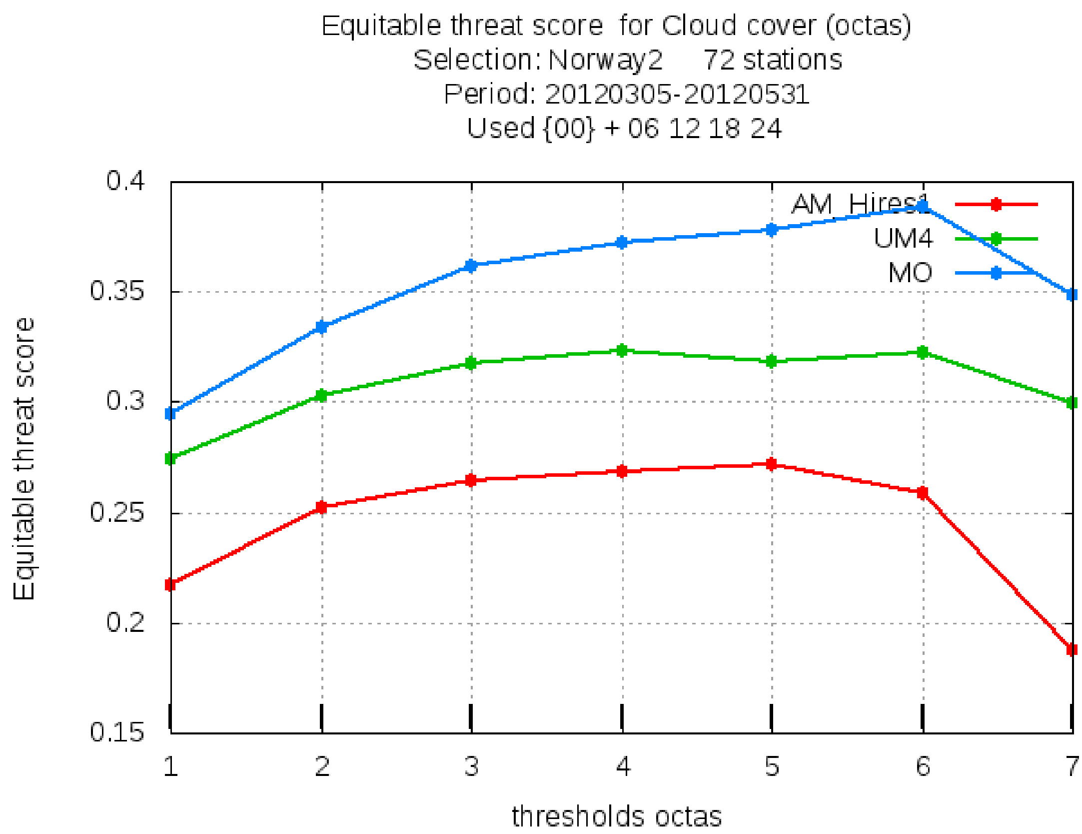


Figure 18: Total Cloud Cover, ETS as a function of total cloud cover forecasts for AM_Hires1 (red), UM4 (green) and MO (blue) averaged over the March and May periods for the Norwegian observations (mostly manual stations).

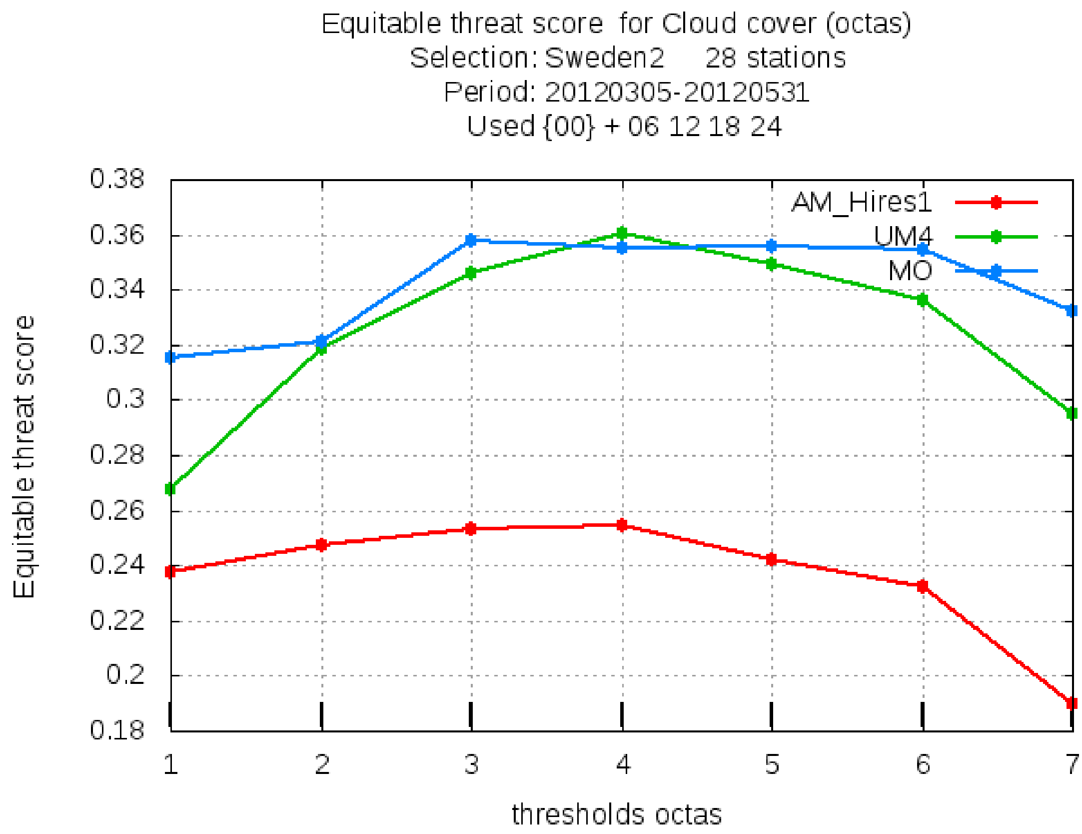


Figure 19: Total Cloud Cover, ETS as a function of total cloud cover forecasts for AM_Hires1 (red), UM4 (green) and MO (blue) averaged over the March and May periods for the Swedish observations (mostly automatic stations).

3.2.6 Summary

For the investigated periods the MO verifies better than UM4 with respect to pressure, temperature, total cloud cover and precipitation in Sweden, and worse with respect to 10 m wind speed and precipitation in Norway. However, the periods are short and it is difficult to draw conclusions.

With the summarized findings in the previous paragraph there are no major changes in the conclusions from the first MetCoOp report with respect to differences in quality between AM_Hires1 and the Unified Model except for pressure. For pressure it is reasonable to believe that the Unified Model in general verifies equal or slightly better than AROME.

4 HARMONIE AROME compared to HARMONIE ALARO

4.1 Background

The main objective for the MetCoOp project is to set up an operational NWP system for Norway and Sweden based on the AROME configuration of the high resolution HARMONIE model system. However, in the HARMONIE model system two main model configurations are available, the AROME configuration designed for horizontal resolution of 2.5 km and the ALARO configuration designed for 5.5 km horizontal resolution. The main differences between AROME and ALARO are mainly related to the description of the physical processes (http://hirlam.org/index.php?option=com_content&view=article&id=65&Itemid=102 , the resolution and often to the employed integration domain. A brief inter comparison (based on cycle 36h1.4) between these two configurations is included in this chapter.

4.2 Results

The verification has been done for three periods (August 2011, December 2011- January 2012 and March 2012). There are differences in the forecasts from the two configurations, but only some of them are described here. When similar findings are present in more than one of the investigated periods only results from one of the periods are used to highlight the results.

4.2.1 Verification results of mean sea level pressure (MSLP).

The AROME configuration show in general slightly smaller RMSE than ALARO with respect to MSLP (Figure 20). There are several possible reasons for this behaviour, but the origin of these cannot be distinguished using the present experimental set-up. A better model description of AROME may be one explanation, another may be the differences in domain size and thereby the relative influence of the lateral boundary conditions (from the ECMWF model). Even though the general RMSE is larger in ALARO, the RMSE for the 26th of December 2011 (the Dagmar storm) is smaller in ALARO compared with AROME (Figure 21). This is a similar behaviour to that which was found in the comparison between the small and the large AROME integration domain, see chapter 6.

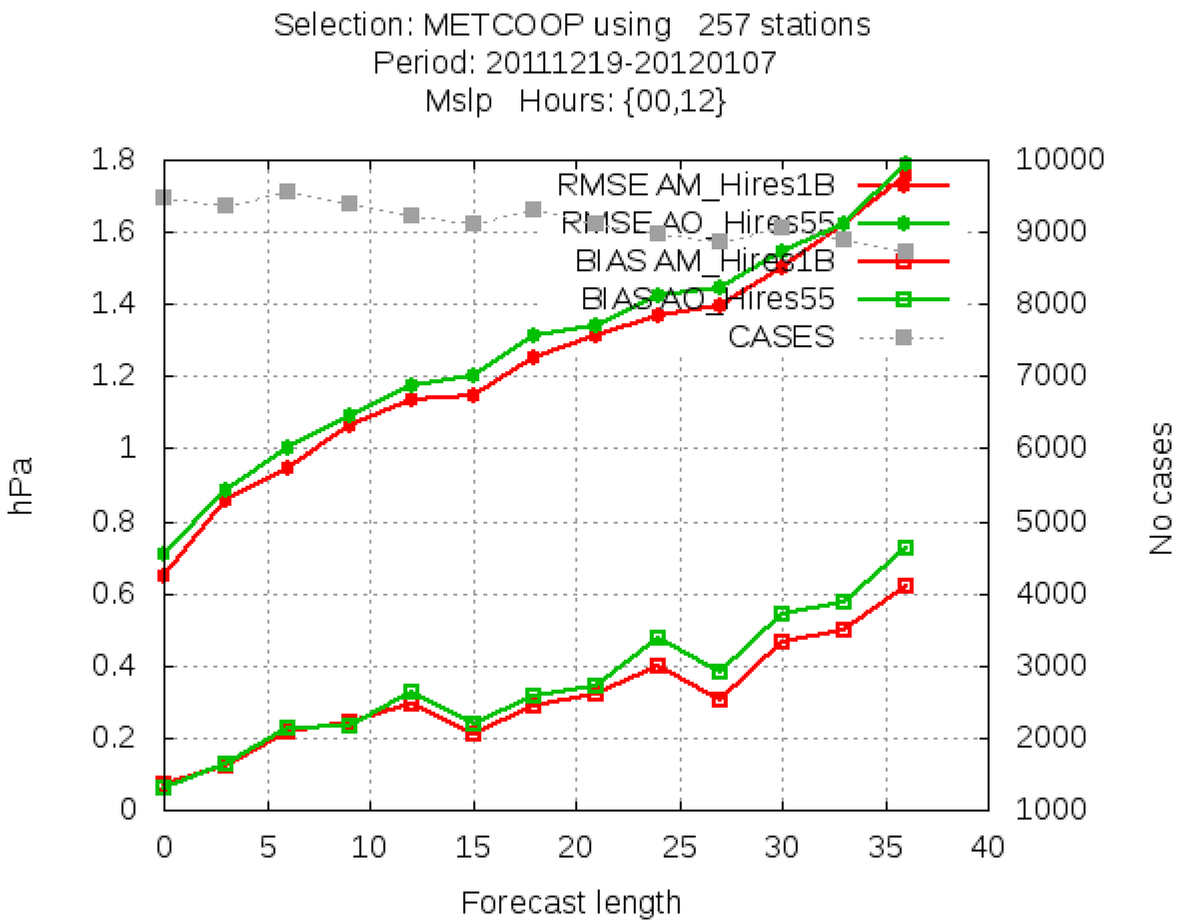


Figure 20: RMSE and BIAS for MSLP from AROME (red) and ALARO (green) as a function of lead time for the winter period.

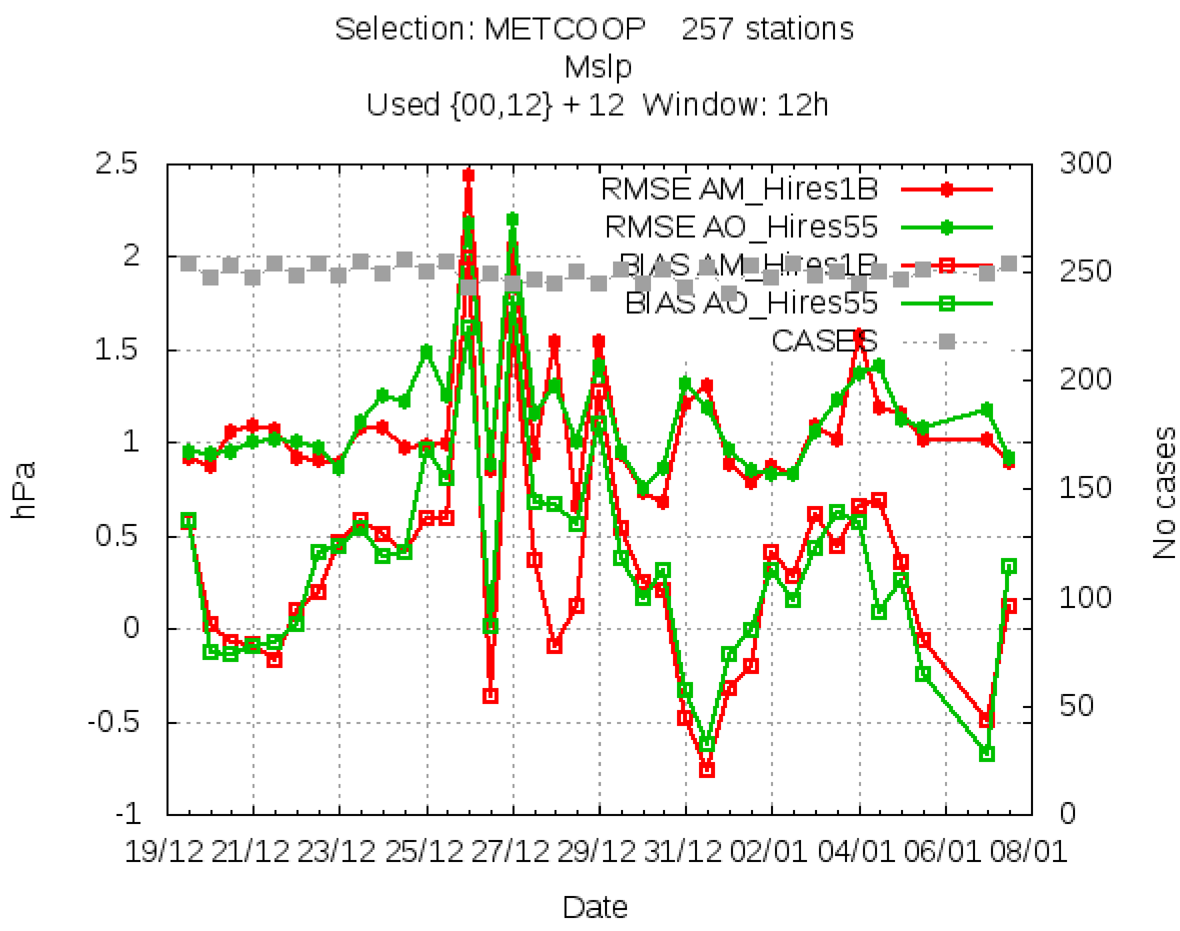


Figure 21: RMSE and BIAS for MSLP from AROME (red) and ALARO (green) as a function of day (12 hr lead time) in the winter period.

4.2.2 Verification results of 10 m wind speed

On average the wind speed in AROME is higher than ALARO. This lead to higher hit rates, but also higher false alarm ratios. The main effect is positive for AROME which is reflected in a higher ETS score. In general this is more pronounced for Sweden than Norway (Figure 23 and Figure 22).

In the mountains there is a very clear difference between the two model configurations. Averaged over 11 mountain stations the wind speeds are approximately 2 m/s higher in AROME than ALARO and in much better agreement with observations (Figure 24)

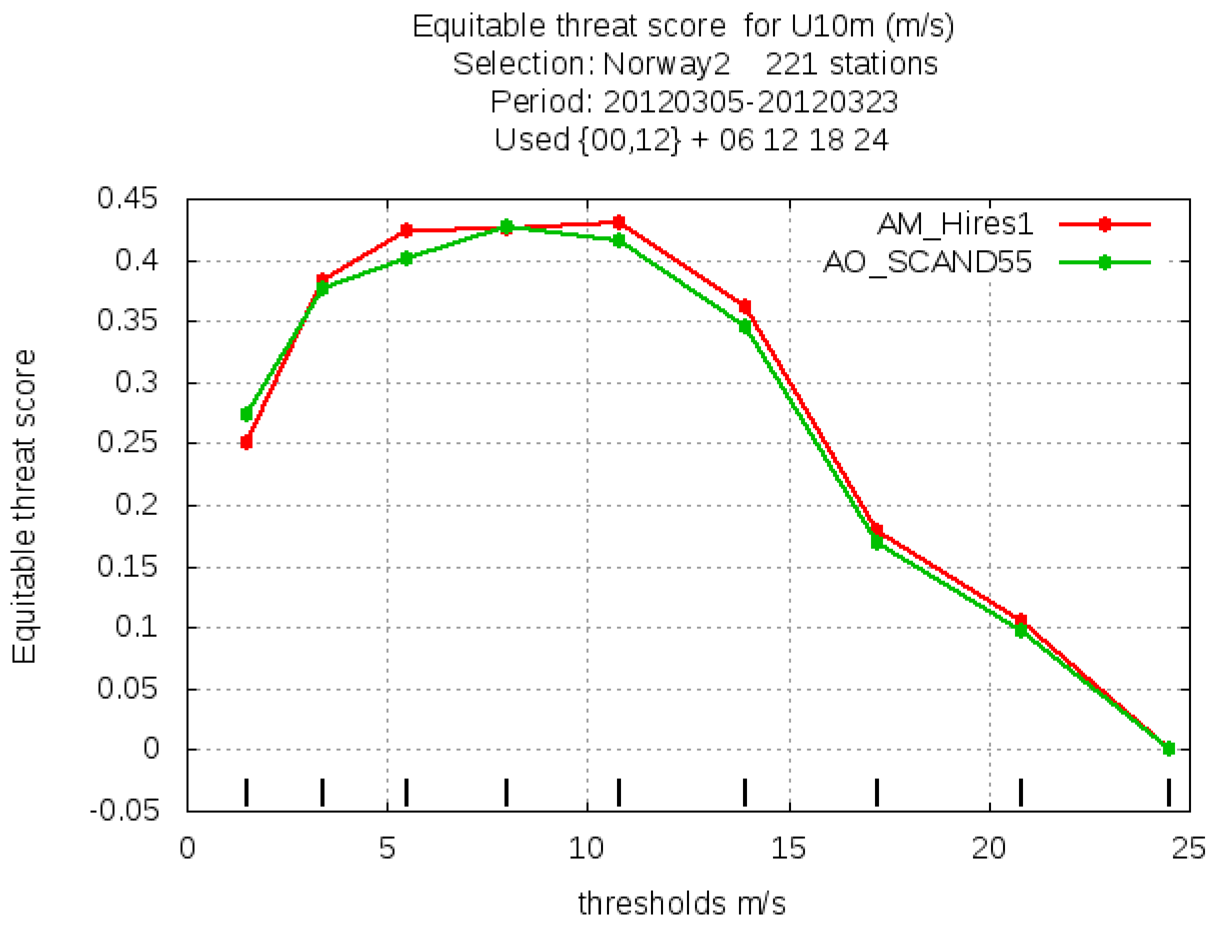


Figure 22: ETS of wind speed at Norwegian stations from AROME (red) and ALARO (green) for the March period.

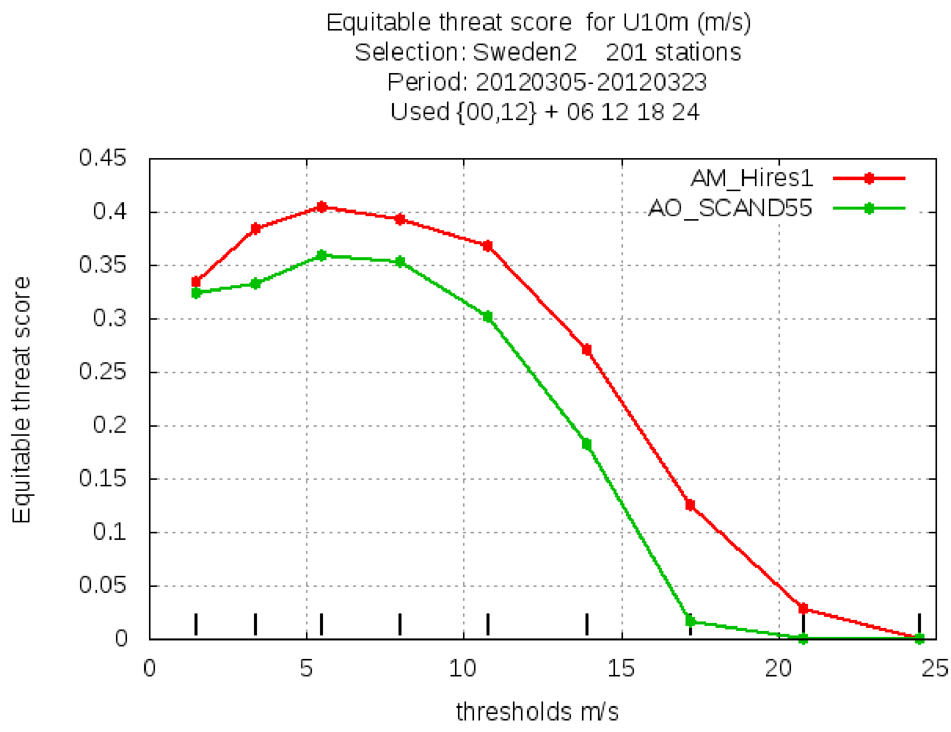


Figure 23: ETS of wind speed at Swedish stations from AROME (red) and ALARO (green) for the March period.

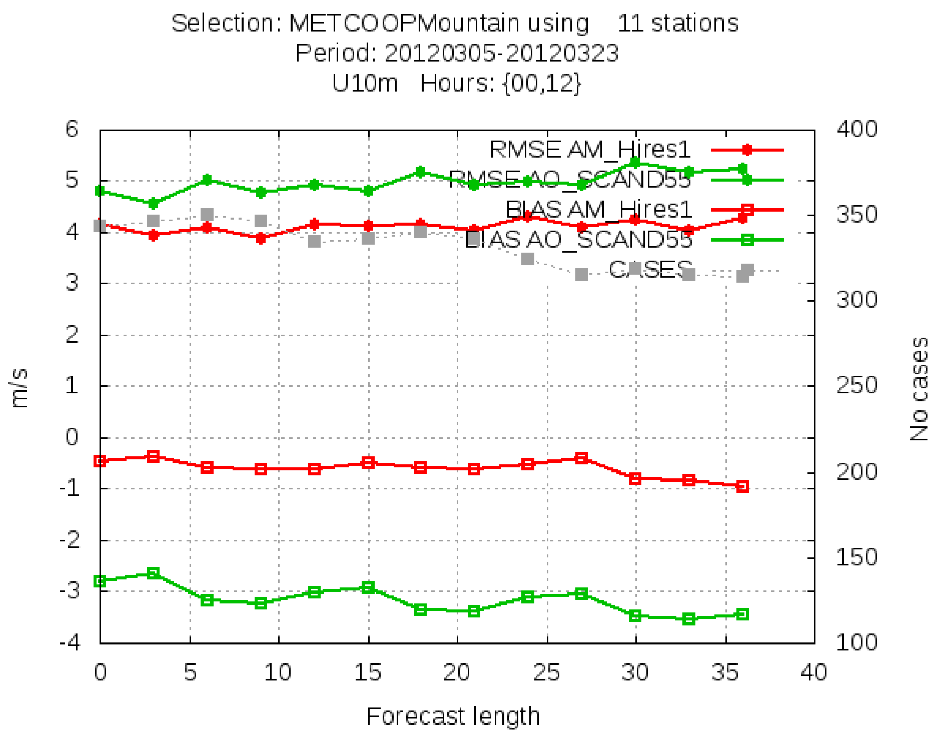


Figure 24: RMSE and BIAS of wind speed as a function of lead time at 11 mountain stations from AROME (red) and ALARO (green) for the March period.

4.2.3 Verification results of 2-metre temperature (T2m)

In Figure 25 RMSE and BIAS for height adjusted T2m (0.6 °C/100 m from model height to station height) are given for the August period. The RMSE score of the two HARMONIE configurations are quite similar, but AROME has in general higher temperatures.

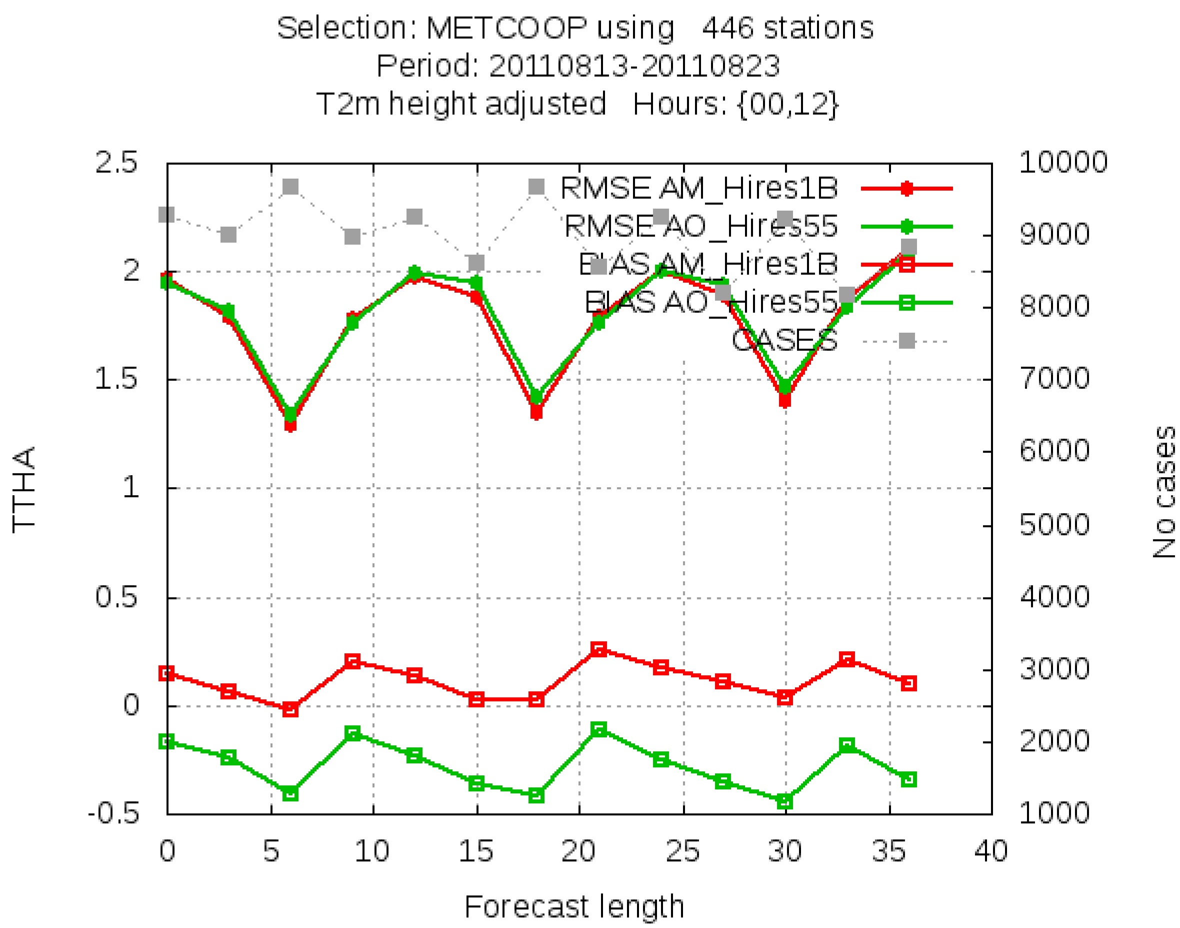


Figure 25: RMSE and BIAS of T2m as a function of lead time averaged over all Norwegian and Swedish stations from AROME (red) and ALARO (green) for the August period.

4.2.4 Verification results of 12 hr accumulated precipitation.

There are more occurrences of precipitation in ALARO than in AROME. There is an overestimation of precipitation events in both models. This is especially pronounced in ALARO (Figure 26) but the numbers vary for the different periods. In general terms, AROME shows a higher ETS score (Figure 27) for small precipitation amounts (less than 1 mm/12 hr, note that this threshold differ slightly between periods). The difference between AROME and ALARO is reflected in a lower hit rate (Figure 28) and false alarm ratio (Figure 29) for AROME compared with ALARO (with the exception for 10 mm/12 hr where the opposite is true). Figures are for the August period, but all three periods show similar results.

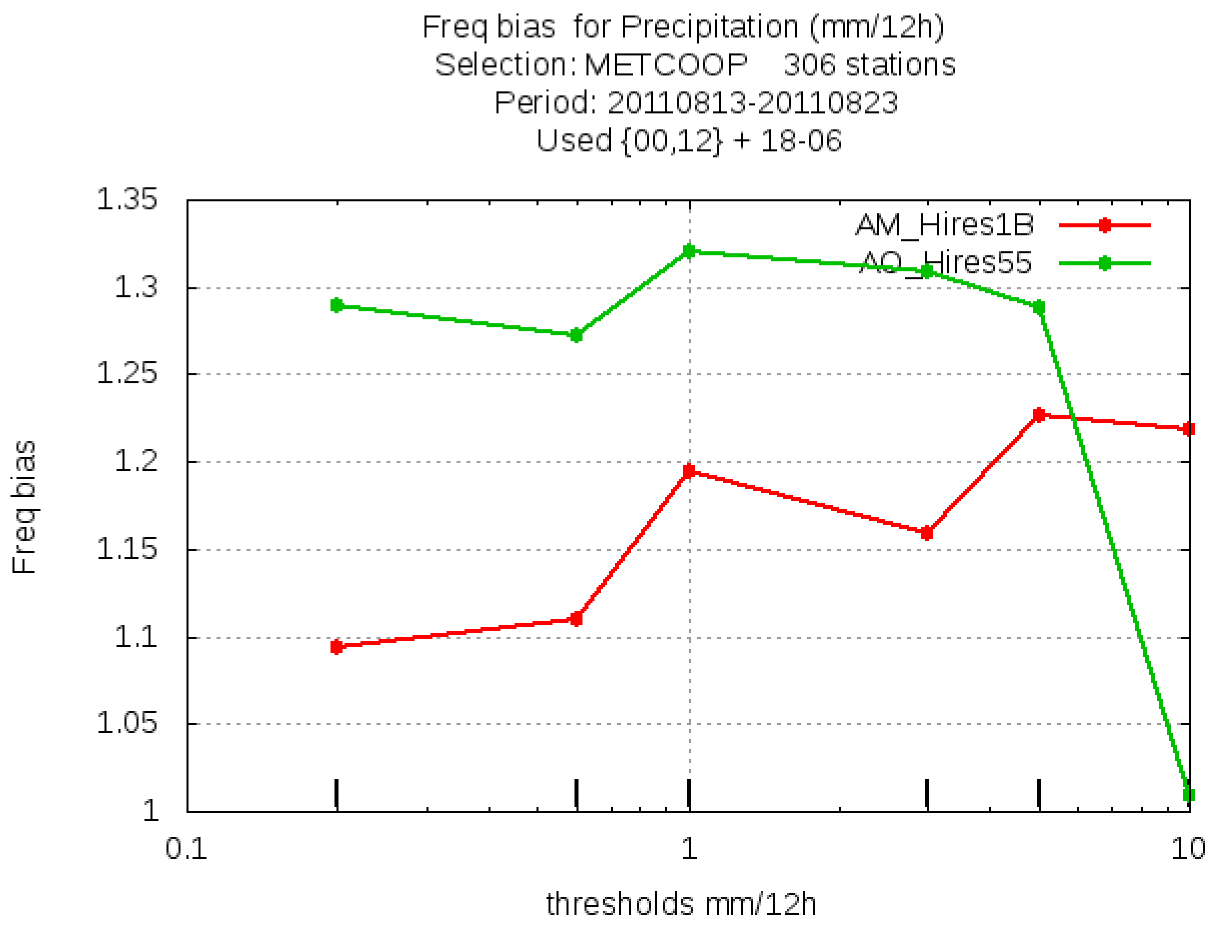


Figure 26: Frequency bias of 12 hr accumulated precipitation from AROME (red) and ALARO (green) for 71 Norwegian stations for the August period.

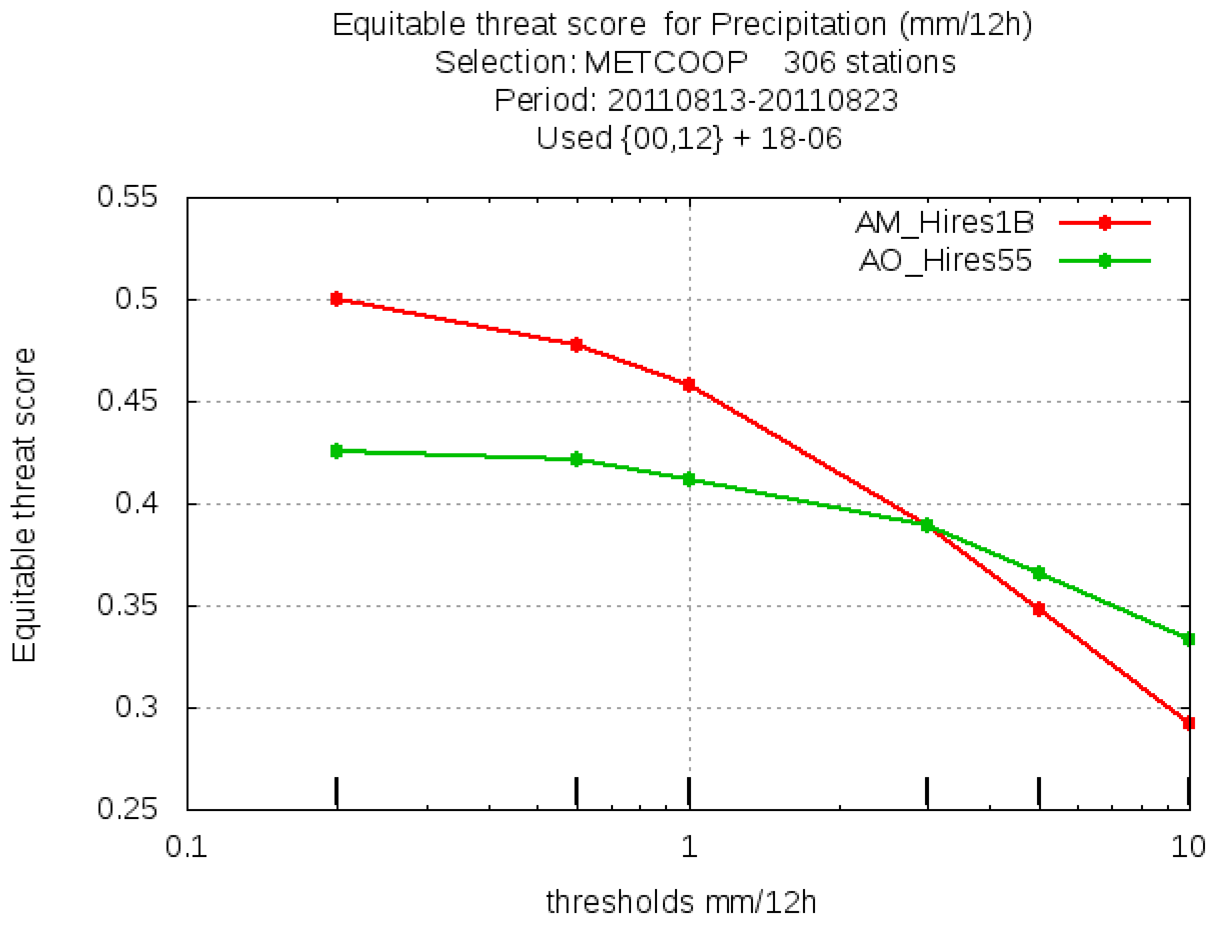


Figure 27: ETS for 12 hr accumulated precipitation from AROME (red) and ALARO (green) for 71 Norwegian stations for the August period.

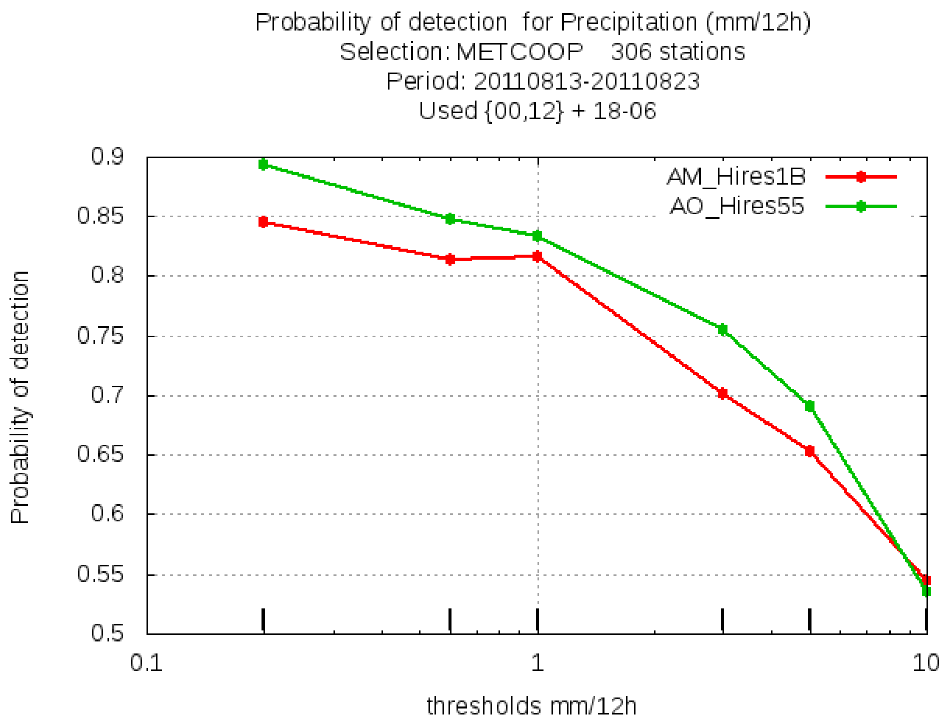


Figure 28: Probability of detection (HR) of 12 hr accumulated precipitation from AROME (red) and ALARO (green) for 71 Norwegian stations for the August period.

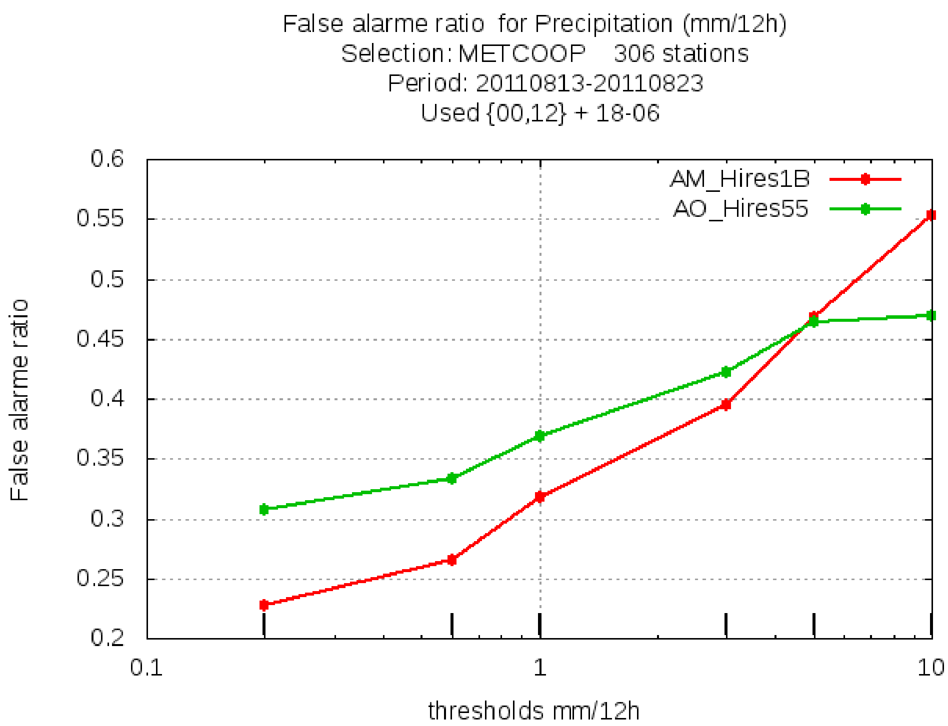


Figure 29: False Alarm Ratio (FAR) of 12 hr accumulated precipitation from AROME (red) and ALARO (green) for 71 Norwegian stations for the August period.

4.2.5 Verification results of total cloud cover (TCC)

The characteristics of the total cloud cover fraction in AROME and ALARO are quite different. Some of these differences in the standard deviation are illustrated in Figure 30. There is higher variability in the AROME cloud cover than in ALARO cloud cover. ALARO has a similar variability to that seen for the Norwegian (manual) observations. In addition, the frequency bias (Figure 31) plot reveals another related behaviour; ALARO has too many occurrences of partly cloudy conditions, while AROME shows the opposite behaviour.

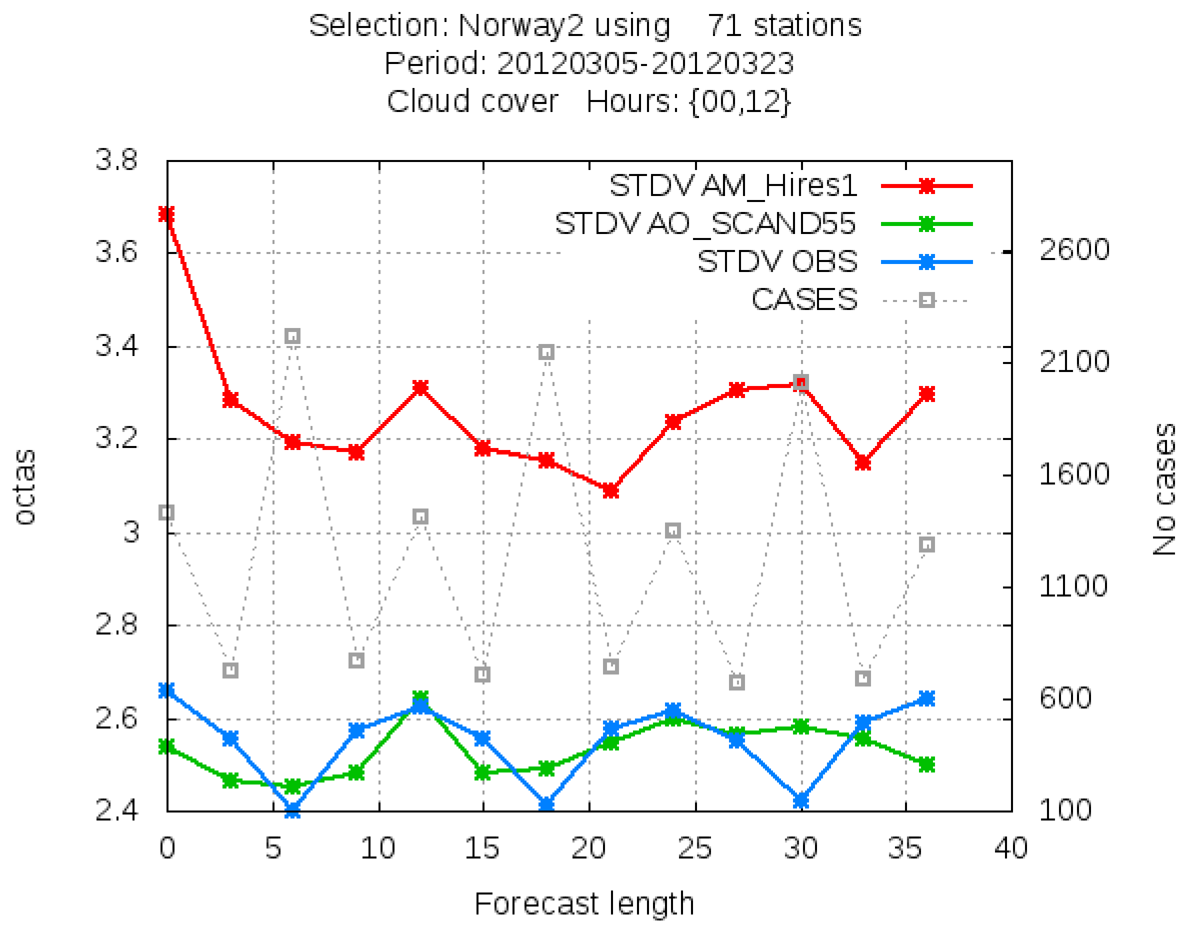


Figure 30: Standard deviation of total cloud cover from AROME (red), ALARO (green) and Norwegian observations, 71 stations (blue) for the March period.

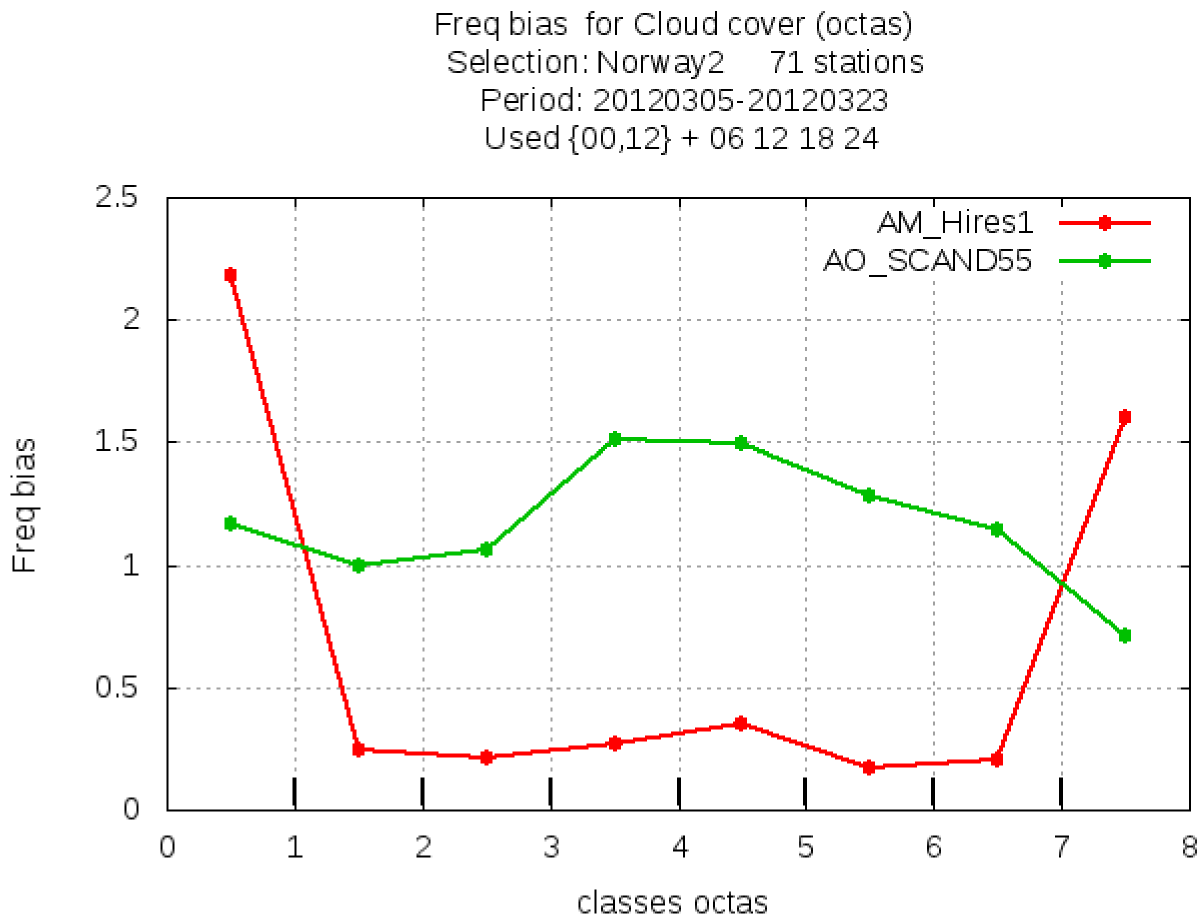


Figure 31: Frequency bias of total cloud cover from AROME (red) and ALARO (green) for 71 Norwegian stations for the March period.

4.2.6 Summary

Based on the results of the three periods evaluated here, it is difficult to come to a firm conclusion on the quality of the two HARMONIE model setups. It is also difficult to find the specific cause of the differences; they may be due to the varying descriptions of physical processes or due to differing model set-ups (e.g. resolution and/or integration domain).

In general, the AROME configuration shows slightly smaller errors for MSLP. The error for 10 m wind speed is on average smaller in AROME than in ALARO. With respect to precipitation ALARO shows better results except for very small precipitation amounts. However the number of precipitation events in the models (and especially in ALARO) is greater than that seen in the observations. The temperature forecasts are quite similar with respect to quality although there is a slight variation between periods. The AROME and ALARO forecasts of total cloud cover have very different characteristics.

5 Comparison of HARMONIE AROME Cycle 36 and Cycle 37

5.1 Background

A new version of the HARMONIE system is called a cycle (shortened cy). Such updates are prior to their release tested both technically and with respect to forecast abilities; to ensure that new model versions improve on the previous ones. However, it is not possible to test all meteorological conditions and all model set-ups (i.e. model domain size and location, forcing data etc.). It is therefore necessary to investigate each cycle with respect to these issues for individual set-ups. In the first METCOOP verification report (01/2012 METCOOP MEMO [3]), cycle 36 (cy36) was evaluated. During June 2012 cycle 37 (cy37) was released. For more information about cy37, see https://hirlam.org/trac/wiki/ReleaseNotes/harmonie-37h1.1 (only available for HIRLAM/ALADIN partners). Cy37 has been run on the August 2011 and December 2011/January 2012 periods (see 01/2012 METCOOP MEMO [3] for description) for the small domain described in chapter 6. A comparison of the results from these two cycles for these two periods will be presented here.

5.2 Results

5.2.1 Verification results of mean sea level pressure (MSLP)

In winter, cy37 give a slightly larger bias and RMSE for MSLP (Figure 32), while the opposite is true for the summer case (Figure 33). The differences between cy36 and cy37 are only significant (90% level) in summer.

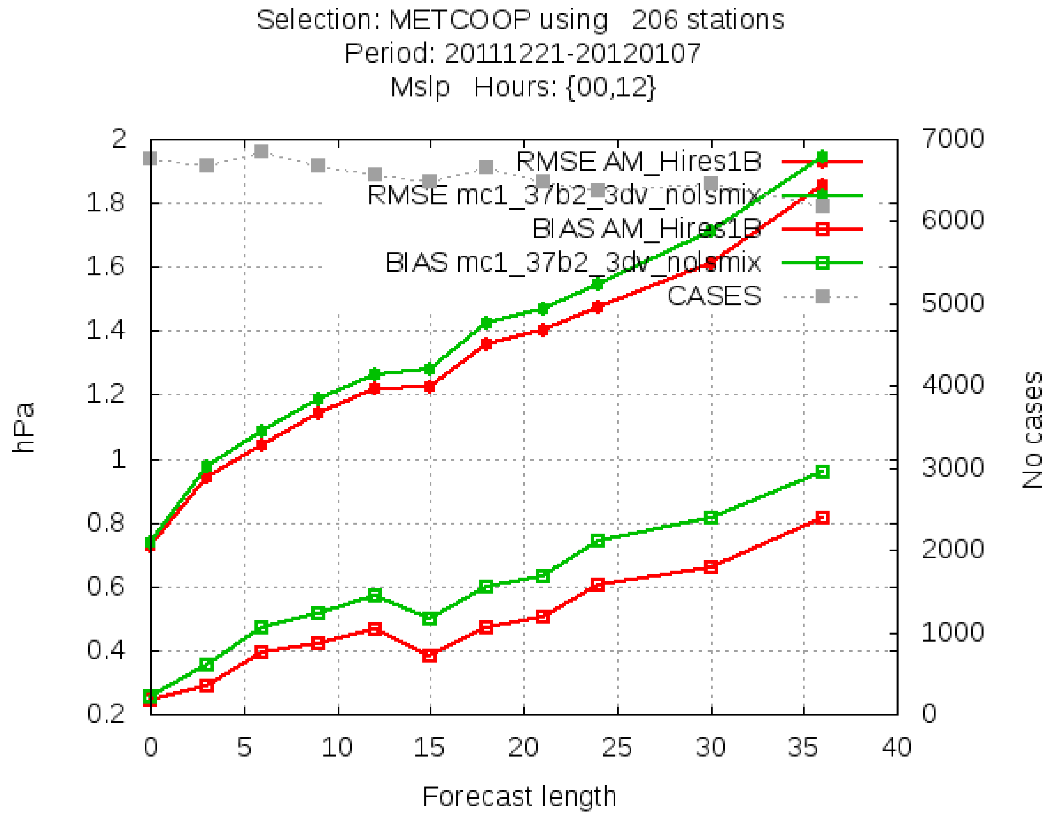


Figure 32: RMSE and bias winter period with cy37 (green) and cy36 (red).

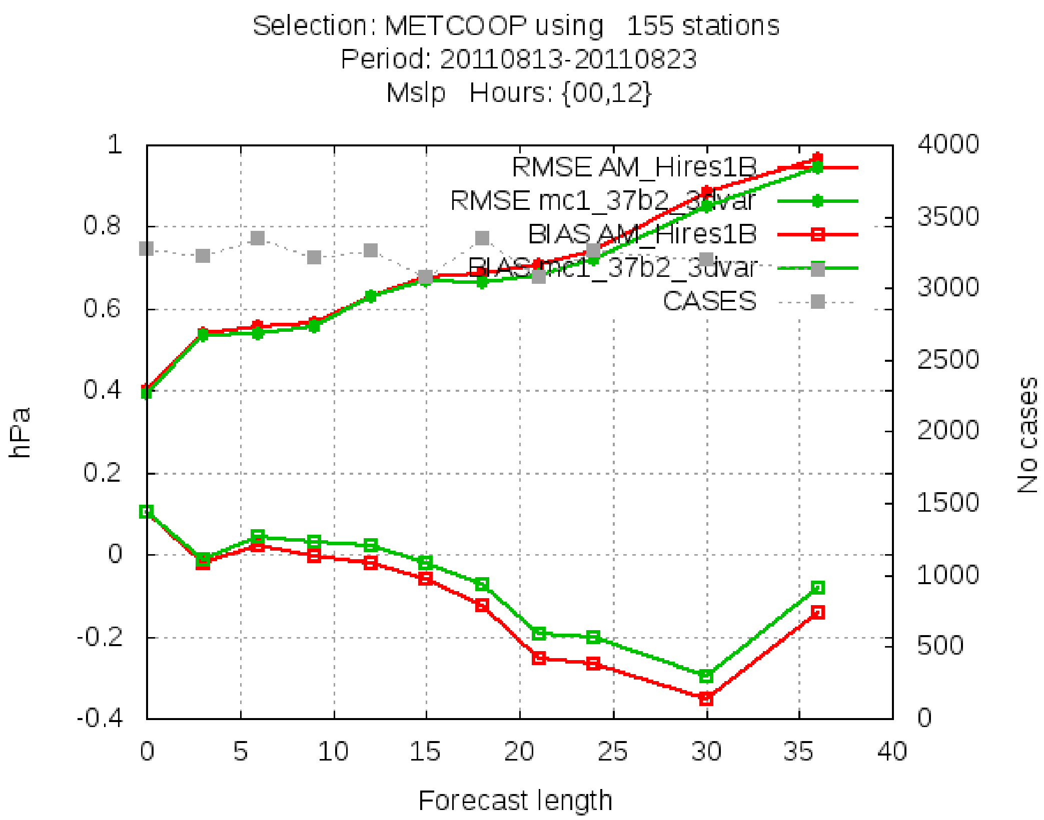


Figure 33: RMSE and bias summer period with cy37 (green) and cy36 (red).

5.2.2 Verification results of 10 m wind speed

For 10 m wind speed there is less wind in cy37 which, in the winter period, is clearly seen in the frequency bias over all Norwegian and Swedish stations (Figure 34) and especially for mountain stations (Figure 35). The ETS are slightly less in cy37 compared to cy36 averaged over all stations (not shown) and significantly less for the mountain stations (Figure 36). As seen in the figures for Hit rate and False Alarm ratio for the mountain stations, the lower wind speed in cy37 leads to fewer false alarms (*Figure 37*), but also to fewer hits (Figure 38). The differences between cy36 and cy37 are most pronounced in the winter period (shown in the figures), but are also present in the summer period (not shown). While the wind speed was very well modelled in the mountains in cy36 this is clearly worse modelled in cy37. The main reason for this change in behaviour is the changed formulation of surface drag between the cycles. Some preliminary tests with cy37 with the surface drag formulation from cy36 give higher and more realistic wind speeds, but also a positive bias.

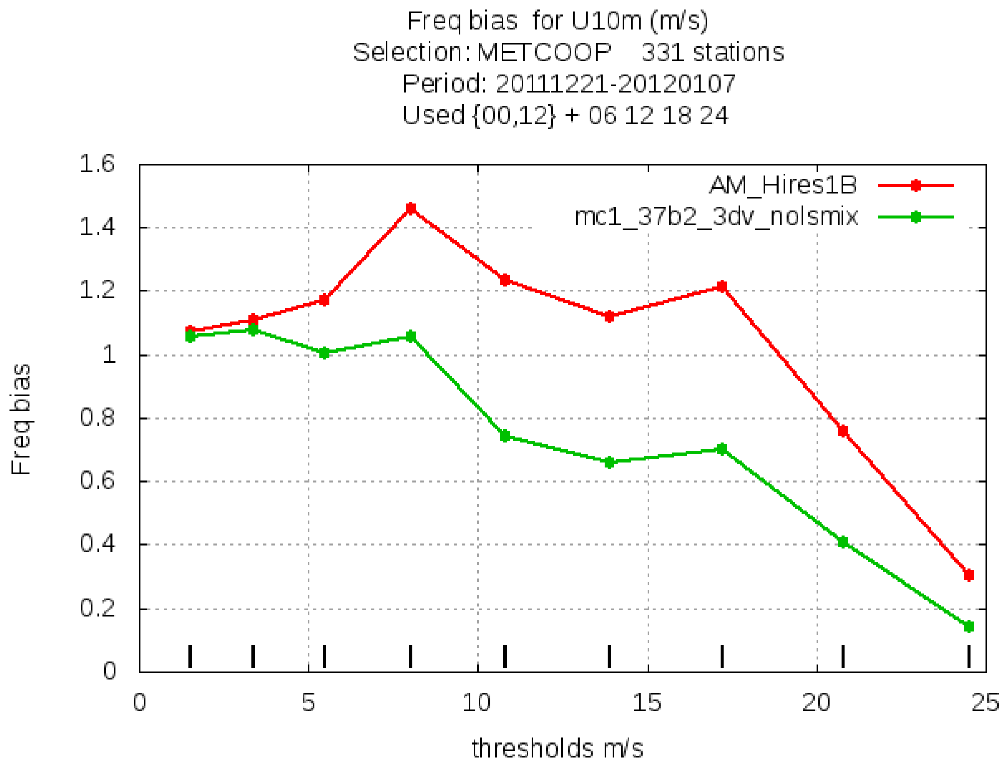


Figure 34: Frequency bias of wind speed for the winter period for all Norwegian and Swedish stations. Cy36 in red and Cy37 in green.

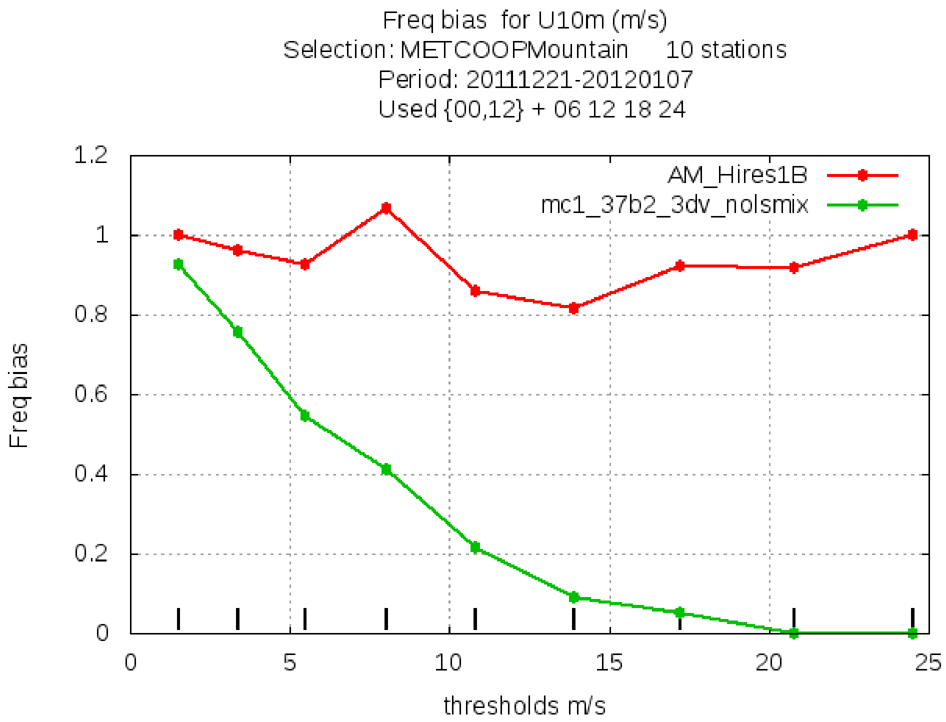


Figure 35: Frequency bias of wind speed for the winter period for mountain stations. Cy36 in red and Cy37 in green.

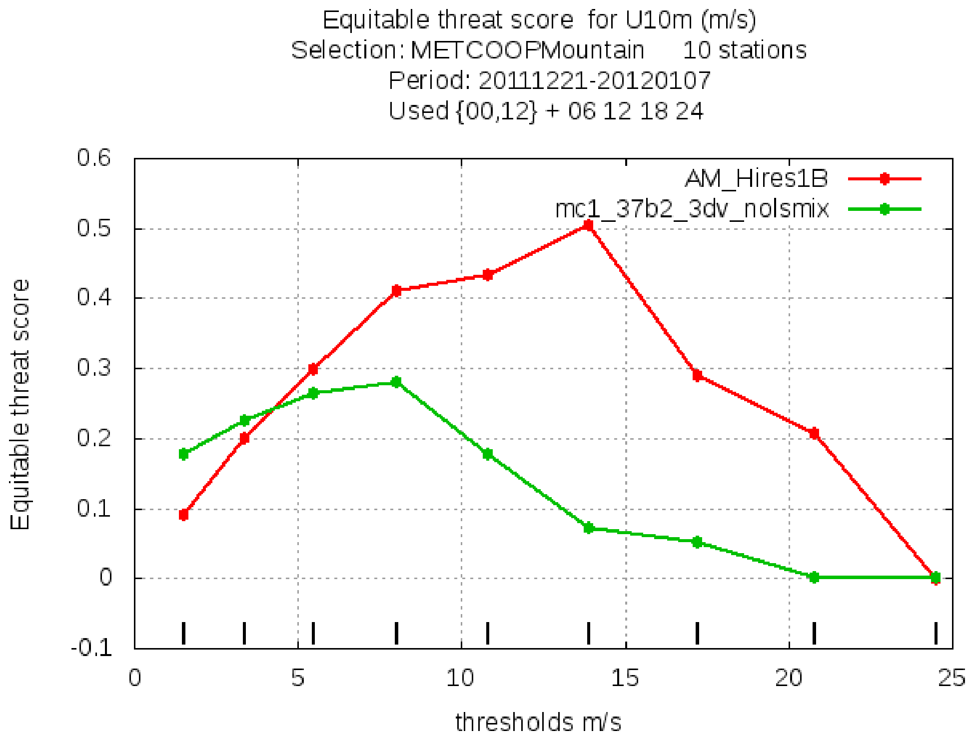


Figure 36: ETS of wind speed for the winter period for mountain stations. Cy36 in red and Cy37 in green.

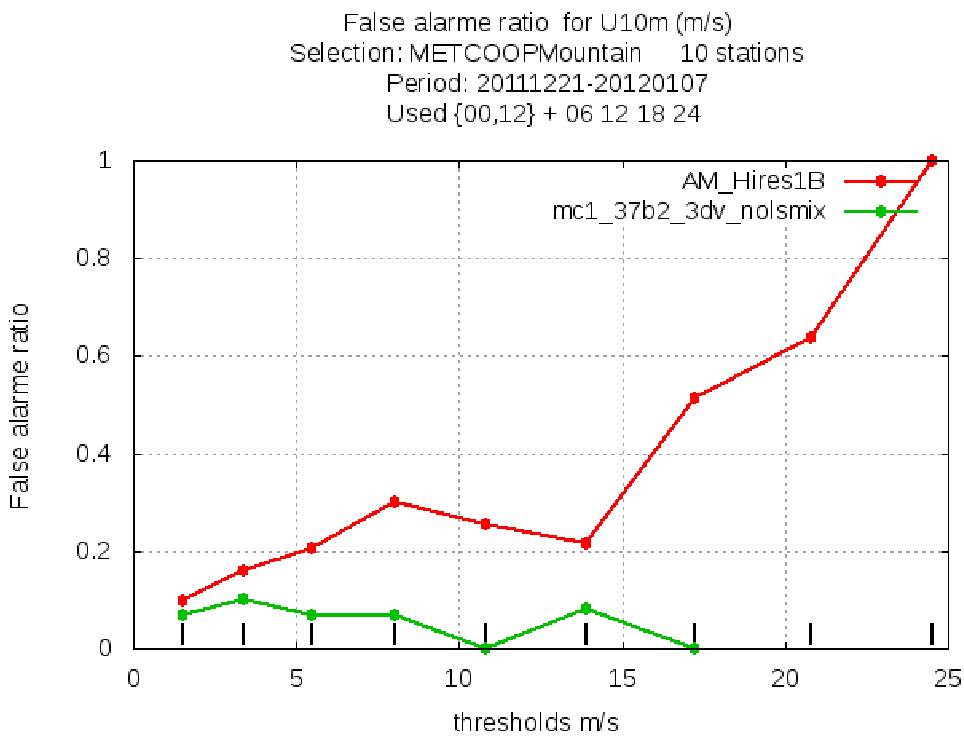


Figure 37: False alarm ratio of wind speed for the winter period for mountain stations. Cy36 in red and Cy37 in green.

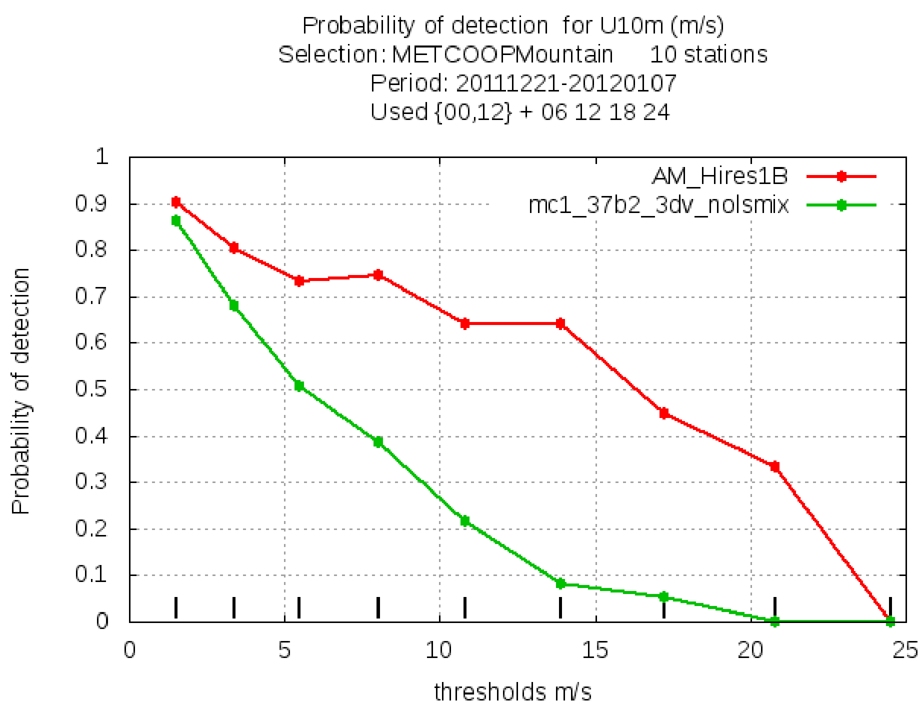


Figure 38: Hit rate of wind speed for the winter period for mountain stations. Cy36 in red and Cy37 in green.

5.2.3 Verification results of 2-metre temperature (T2m)

The RMSEs for T2m for cy36 and cy37 are quite similar, but the forecast quality compared to each other depend on period and region. In Figure 39, RMSE and bias are shown averaged over all Norwegian and Swedish stations for the winter period. In the figure it is seen that cy37 is slightly colder than cy36. This is also found for the summer period (not shown). For this winter period a neutral bias for cy37 and a small positive bias for cy36 are seen. However, for the summer period, cy37 shows a small negative bias while cy36 shows a neutral bias. The periods of verification are very short; it is therefore difficult to come to conclusions, and especially with respect to systematic errors.

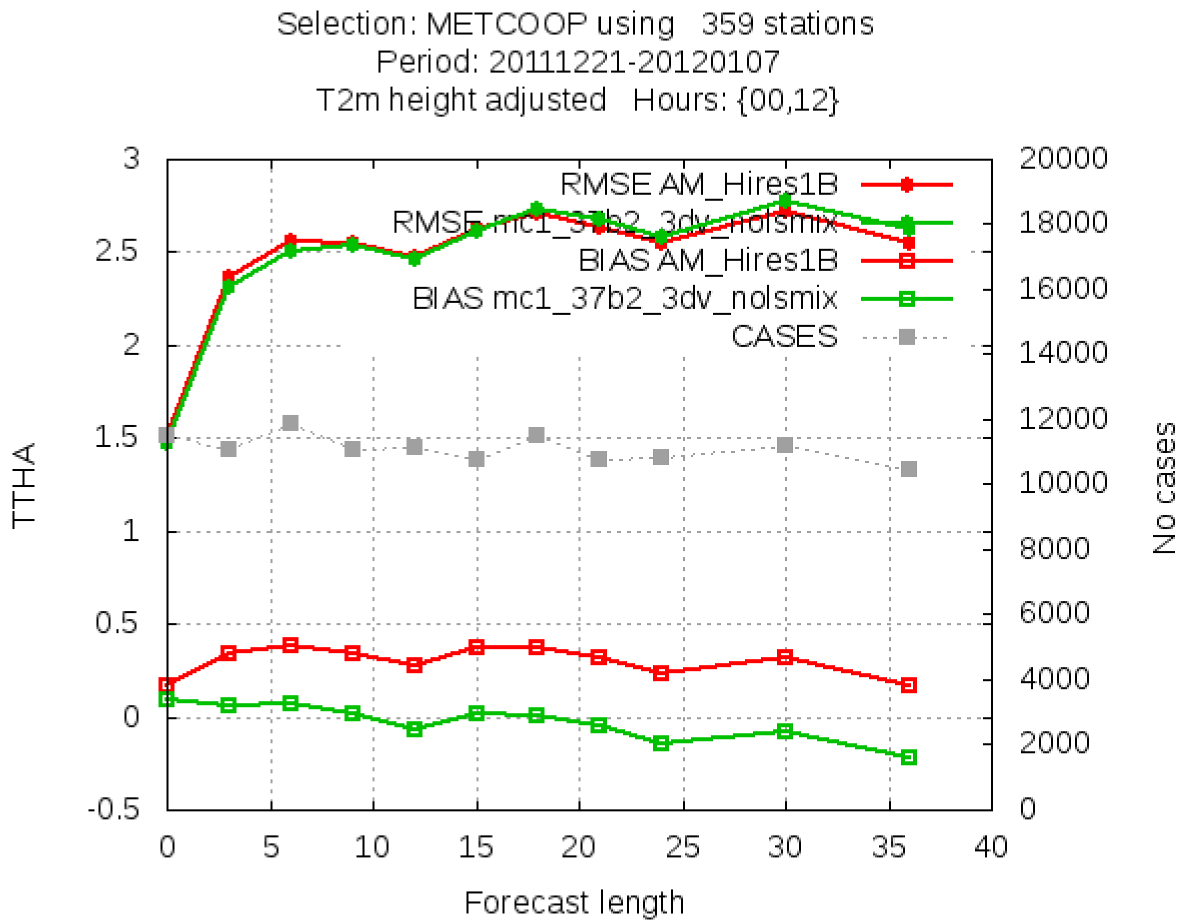


Figure 39: RMSE and bias for temperature (T2m) for the winter period for all Norwegian and Swedish stations. Cy36 in red and Cy37 in green.

5.2.4 Verification results of 12 hr accumulated precipitation and total cloud cover

For precipitation and total cloud cover there are only minor differences in the behaviour of cy36 and cy37 so this will not be shown here.

5.2.5 Summary

This study is based on two relatively short periods. The main difference introduced by the upgrade from cy36 to cy37 is a decrease in the wind speed, and thereby also a decrease in forecast quality, and especially in the mountains. This decrease in wind speed leads to a reduction in the quality of the forecast. Otherwise there are relatively small differences between cy36 and cy37. For more information about cy37, see https://hirlam.org/trac/wiki (only available for HIRLAM/ALADIN partners).

6 Comparison of “Small domain”(v_1) versus “large domain”(v_2)

6.1 Background

A pragmatic approach is often chosen in the set-up of Numerical Weather Prediction (NWP) Limited Area Models (LAMs). This may result in LAMs employed on a domain size as large as possible given available computer resources, demands on resolution (necessary to describe relevant physical processes reasonably) and the urge for early delivery of forecasts. However, it is well known that model results from LAMs, for many reasons, depend on the domain. The MetCoOp Hires_1 domain is the first version of the high resolution domain covering both Sweden and Norway (v1 in red, Figure 40). After a discussion with data assimilation developers and considering the needs for operational coverage the Hires_2 domain was introduced (v2 in green, Figure 40).

Here, the forecast quality produced with Hires_1 (hereafter referred to as the small domain) and the Hires_2 (hereafter referred to as the large domain) domains are compared for cy36h1.4 for the December/January test period. Only surface analysis is applied and no upper-air assimilation is included since background error statistics for the new domain were not available when the experiments were performed.

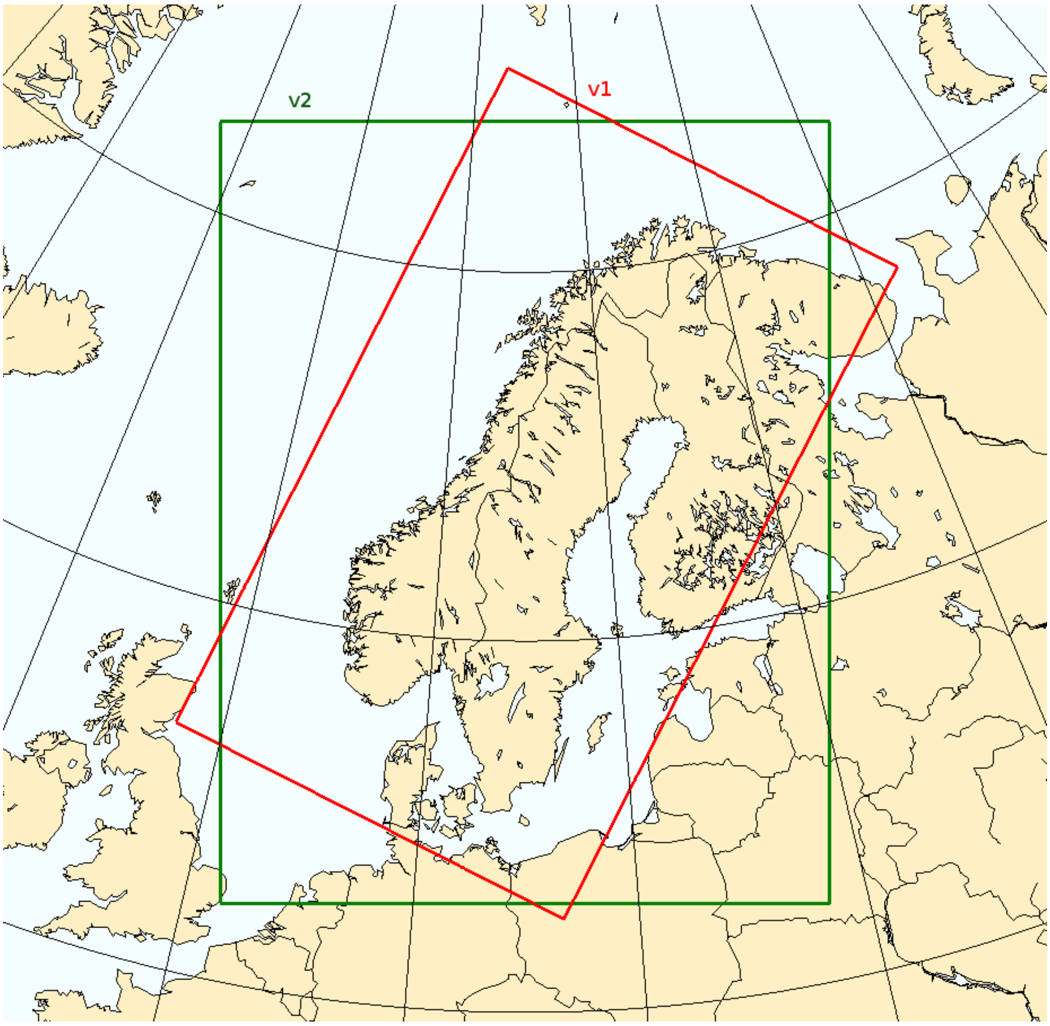


Figure 40: Overview of the” small” domain in red, and the “large” domain in green.

6.2 Results

Based on statistics averaged over the whole period and all Norwegian and Swedish stations the forecast quality of the small and the large domain are very similar. However, closer inspection reveals some differences.

6.2.1 Verification results of mean sea level pressure (MSLP)

The only significant (90% level) differences are found in MSLP on the forecast for day two. The large domain gives a slightly larger MSLP bias and RMSE than the small domain (Figure 41). This is probably due to less influence by the lateral boundaries provided from ECMWF, which are assumed to be of high quality with respect to the large scale phenomena. However, despite a slightly larger average RMSE, the RMSE of the large domain are smaller on particular days (for Swedish stations, Figure 42). These particular days are associated with strong and fast evolving synoptic systems. A hypothesis is that this behaviour is related to the lateral boundary data treatment and update frequency (i.e. fast moving systems will be poorly sampled with a 3 hourly update frequency of the lateral boundaries and associated problems are more pronounced close to the lateral boundaries. Increasing the lateral update frequency of the model from 3 hr to 1 hr will be tested in a separate experiment for this particular period.

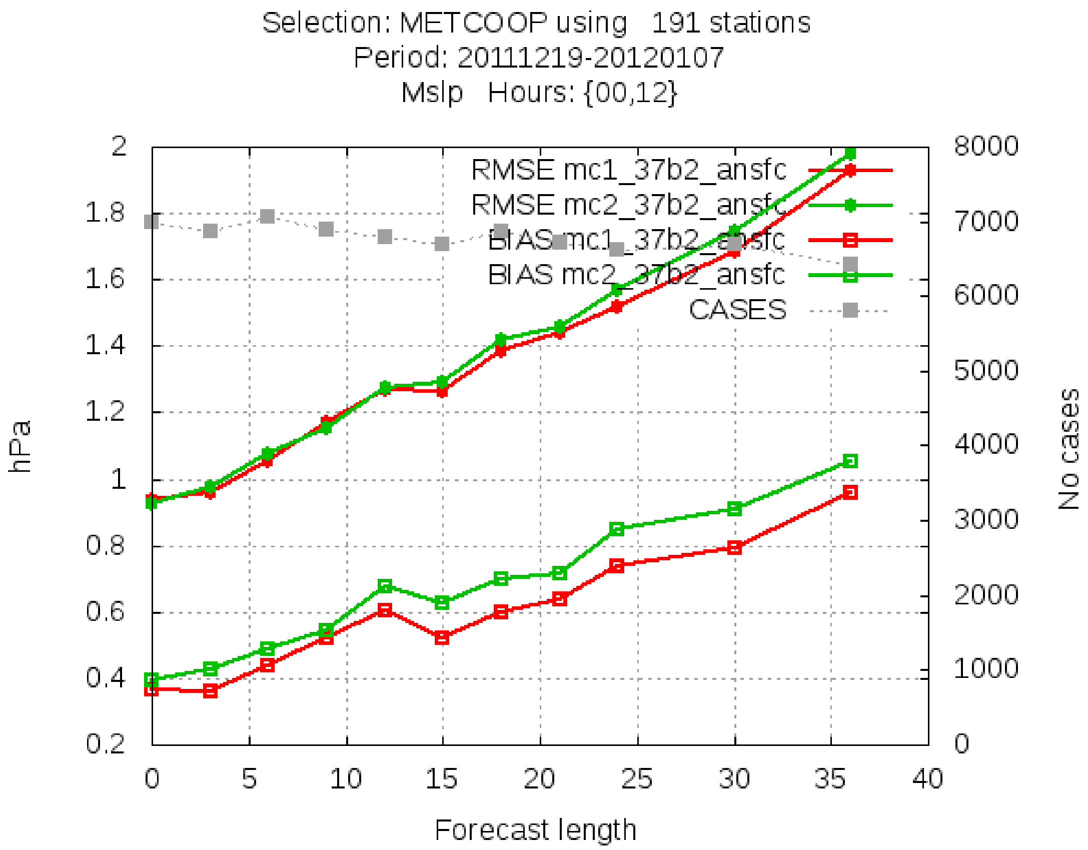


Figure 41: MSLP BIAS and RMSE as function of lead time for the large domain (green), and the small domain (red) for the December- January 2012 period

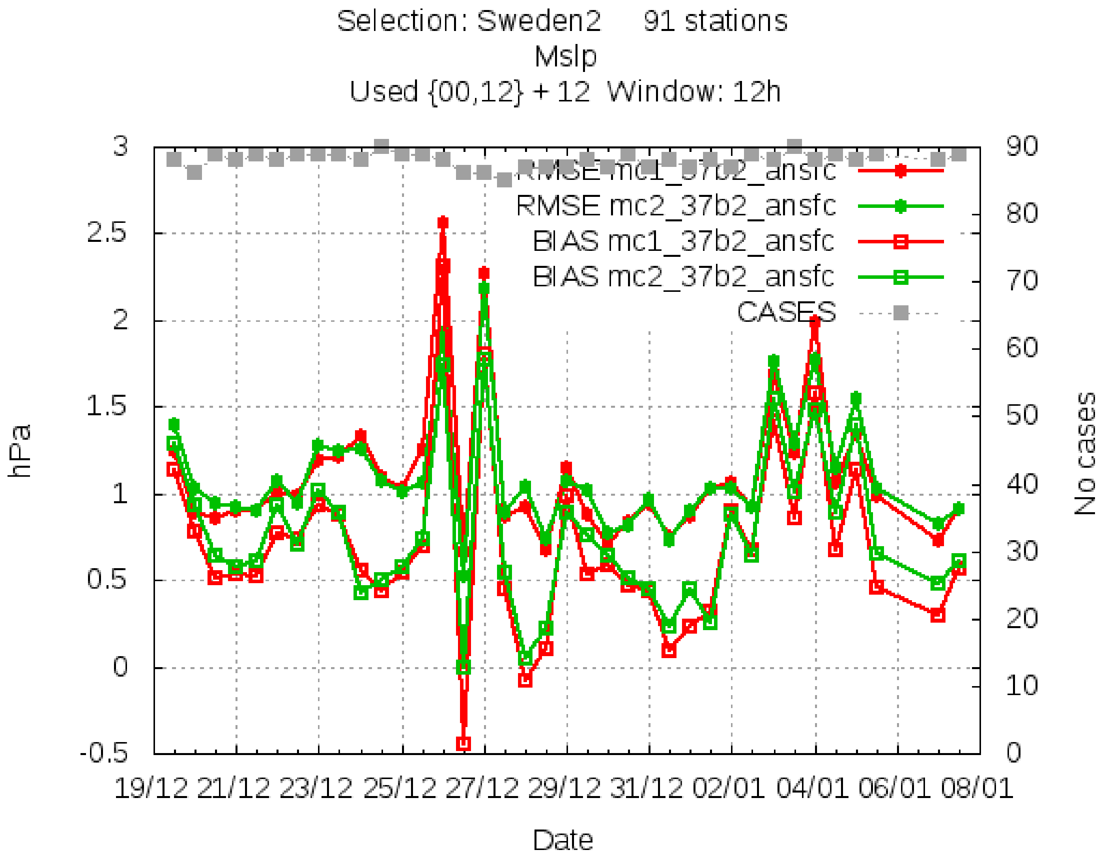


Figure 42: MSLP BIAS and RMSE (+12t lead time) as function of day for the large domain (green), and the small domain (red) for the December/January test period.

6.2.2 Verification results of 10 m wind speed

For the highest wind speeds there is a decrease in ETS score due to fewer hits (and fewer false alarms) with the large domain (Figure 43). This feature is most pronounced at the coastal stations (not shown). A short period, few stations and a limited number of occurrences means that the interpretation of the results should be done with care.

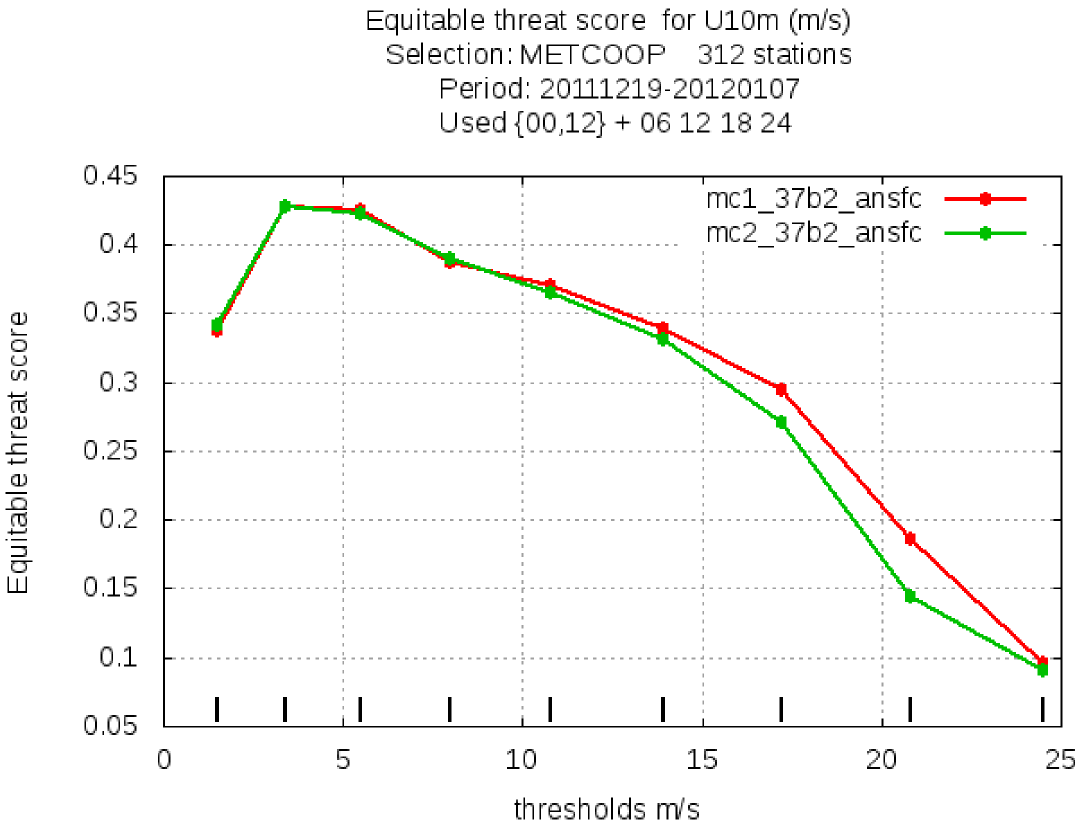


Figure 43: 10 m wind, ETS as function of exceedance thresholds for the large domain (green), and the small domain (red) for the December/January period.

6.2.3 Verification results of 2-metre temperature (T2m)

There are no differences between the domains for T2m in average for BIAS or RMSE. However, at the coastal stations the large domain gives slightly smaller RMSE and systematic errors (Figure 44). These differences are however not significant and there are at the moment no indication that domain size should influence the coastal temperature.

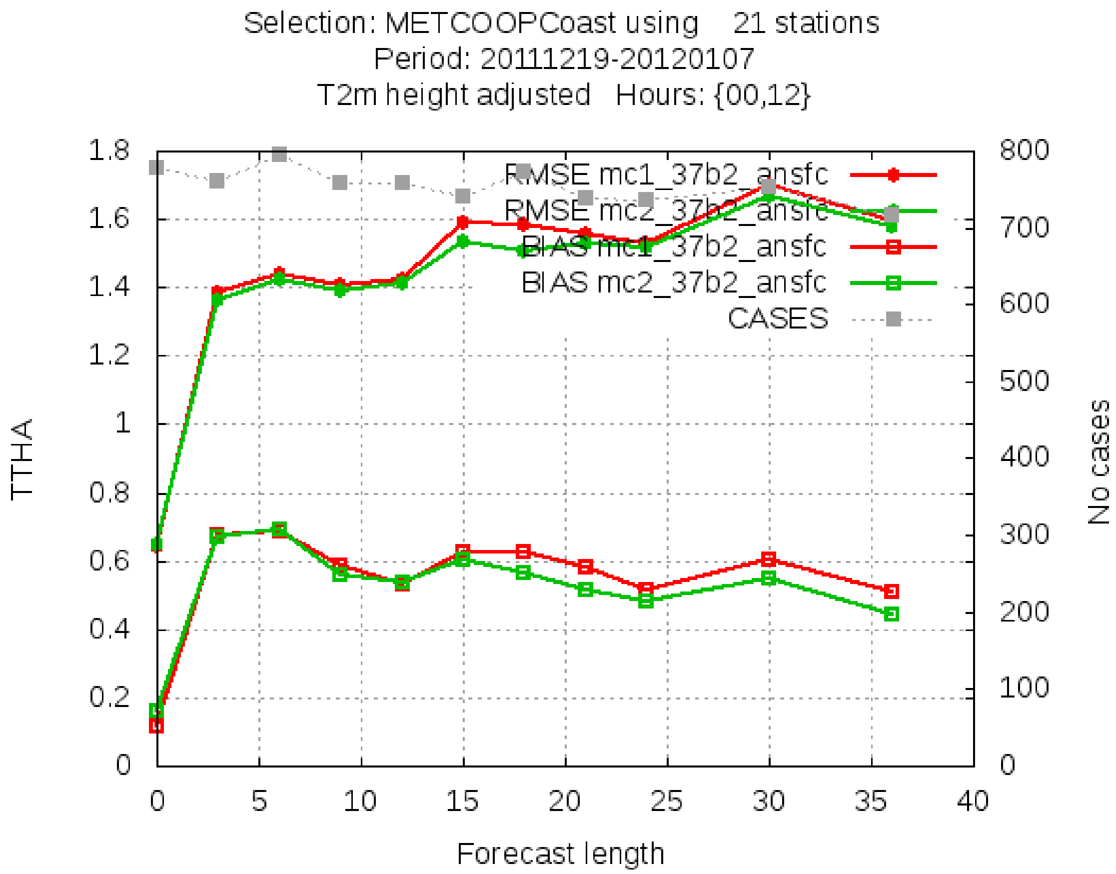


Figure 44: T2m for coast stations. BIAS and RMSE as function of lead time for the large domain (green), and the small domain (red) for the December/January period.

6.2.4 Verification of 12 hr accumulated precipitation and total cloud cover

Averaged over all Norwegian and Swedish stations there are more occurrences of precipitation in the model runs for the large domain. This is seen by an increase in the bias frequency of 1-2% for small precipitation amounts (< 1 mm/12 h) and up to 5% for 10 mm/12 hr (see Figure 45), and is reflected in a higher hit rate for all precipitation thresholds (see see Figure 46). Usually, such behaviour is connected to a similar increase in false alarm ratio. However, it is remarkable that the false alarm ratio is almost unchanged between the two domains (see Figure 47). The large domain shows higher ETS than the small domain (meaning better result), see Figure 48. Again the results are not significant and interpretation should be done with care. A hypothesis is that for a large domain the humidity related variables have more time to develop correctly from the inflow on the lateral boundaries (see Køltzow et al. [2] for an example from regional climate modeling).

Only minor differences are found between the two domains with respect to total cloud cover.

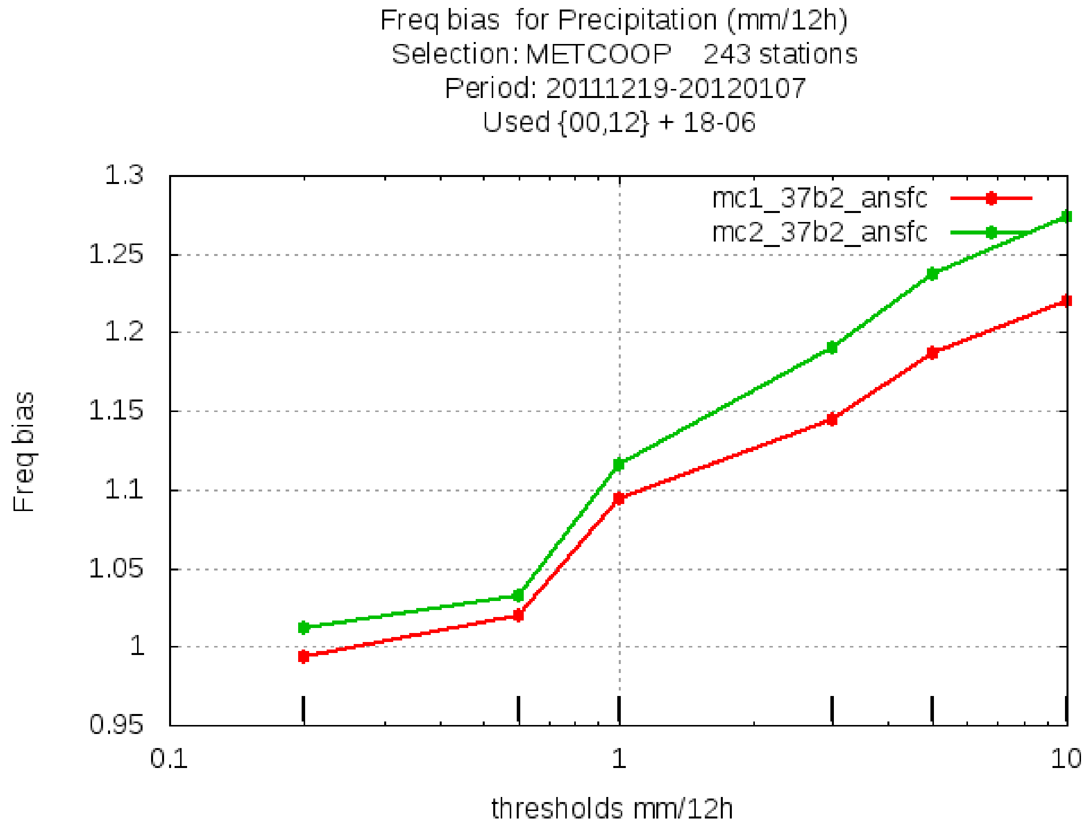


Figure 45: Frequency bias for 12 hr precipitation as function of exceedance thresholds for the large domain (green), and the small domain (red) for the December/January period.

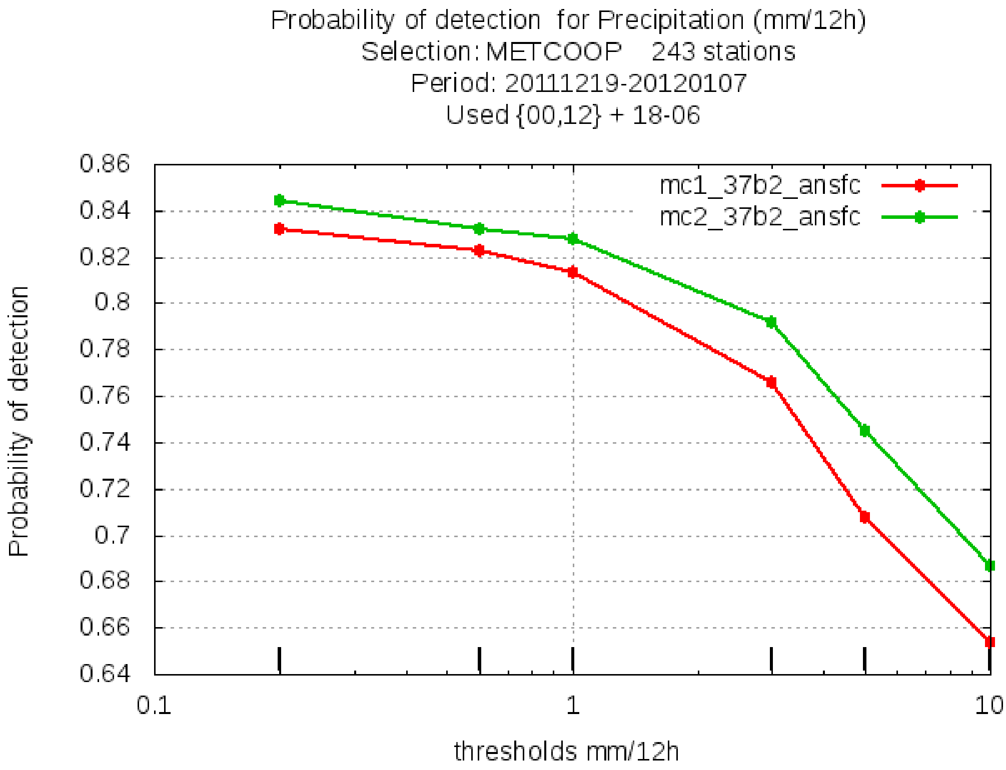


Figure 46: Hit rate for 12 hr precipitation as function of exceedance thresholds for large domain (green), and small domain (red) for the December/January period.

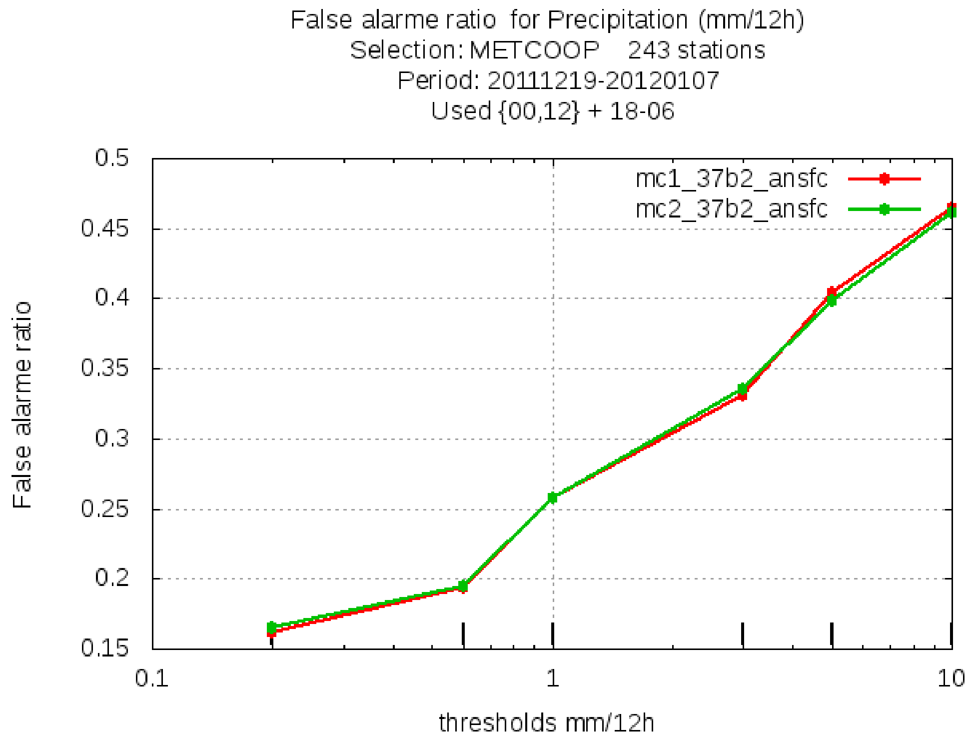


Figure 47: False alarm ratio for 12 hr precipitation as function of exceedance thresholds for large domain (green), and small domain (red) for the December/January period.

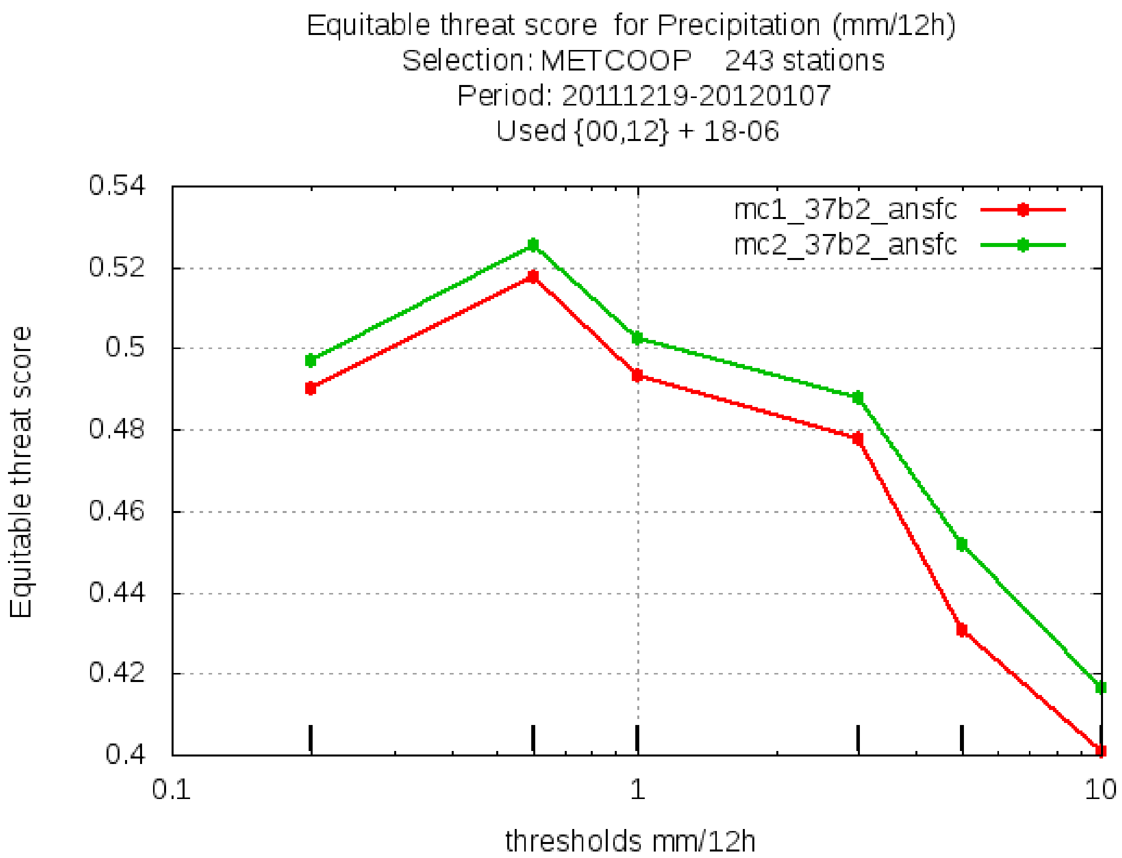


Figure 48: ETS for 12 hr precipitation as function of exceedance thresholds for large domain (green), and small domain (red) for the December/January period.

6.2.5 Summary

There are small differences in the forecast quality between the two integration domains tested for this period. A small positive impact of the larger domain is found with respect to precipitation, coastal temperatures and MSLP for high impact events. A small negative impact is found for the larger domain for MSLP in general and strong wind events.

It is expected that high resolution observations and upper-air assimilation will have a more positive influence on the larger domain. The effect of including more observations in the assimilation will be evaluated later.

7 References

[1] Kristiansen, Jørn, Dag Bjørge, John M. Edwards, Gabriel G. Rooney, (2012), *Soil Field Model Interoperability: Challenges and Impact on Screen Temperature Forecast Skill during the Nordic Winter.* J. Hydrometeor, 13, 1215–1232. doi: http://dx.doi.org/10.1175/JHM-D-11-095.1

[2] Køltzow, Morten, Iversen Trond and Haugen Jan Erik. (2008), *Extended Big-Brother experiments: the role of lateral boundary data quality and size of integration domain in regional climate modelling.* Tellus A, 60: 398–410. doi: 10.1111/j.1600-0870.2008.00309.x

[3] Køltzow Morten A, Ivarsson Karl-Ivar, Agersten Solfrid, Meuller Lars, Bjørge Dag, Vignes Ole, Dahlgren Per, Eriksen Bjart, Ridal Martin, Rudsar Rebecca. (2012) *Verification study HARMONIE AROME compared with HIRLAM, UM and ECMWF*. 01/2012 METCOOP MEMO (Technical Memorandum) http://metcoop.org/memo

[4] Roberts N.M. and Lean H.W. (2008). *Scale selective verification of rainfall accumulation from high-resolution forecasts of convective events*. Monthly Weather Review Vol 136, p. 78-97

8 Figures and tables

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