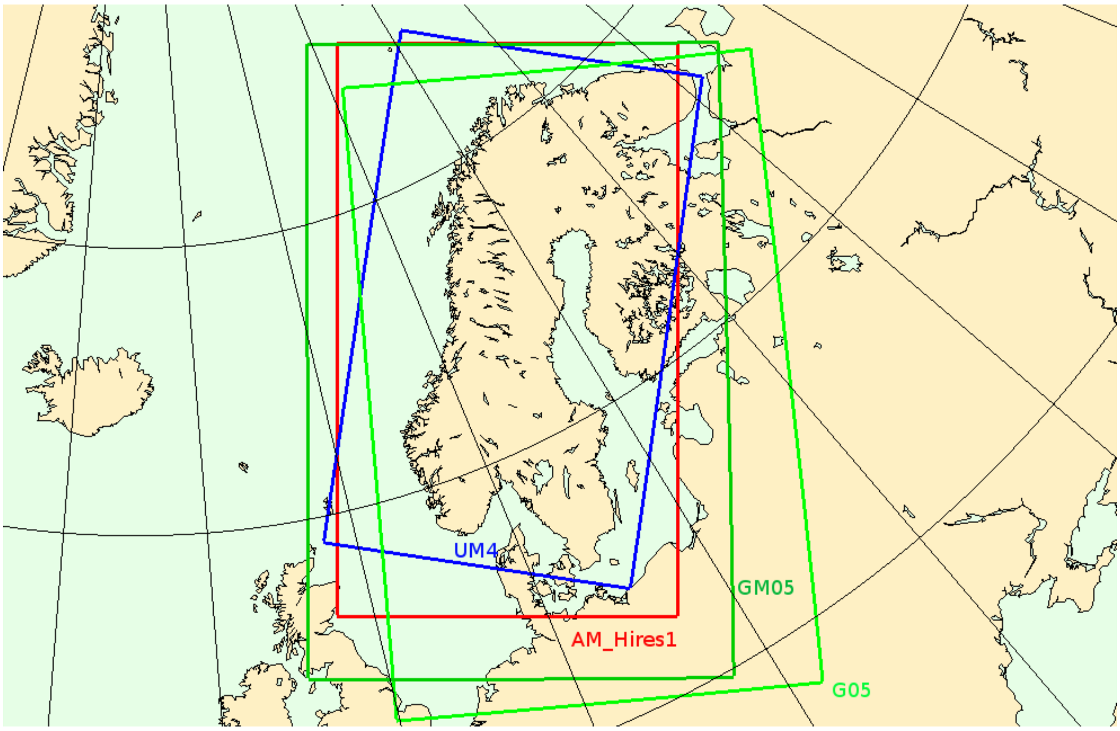
METCOOP MEMO No. 01, 2012

Verification study

HARMONIE AROME compared with HIRLAM, UM and ECMWF

Morten A Køltzow, Karl-Ivar Ivarsson, Solfrid Agersten, Lars Meuller, Dag Bjørge, Ole Vignes, Bjart Eriksen, Per Dahlgren, Martin Ridal, Rebecca Rudsar



Front:
Model domains for the four compared models in this study; HARMONIE AROME (AM_Hires1), Unified Model (UM4), and HIRLAM (GM05, used in chapter 3.4 and 3.5 and G05 used in chapter 3.6 and 3.7).

MetCoOp

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

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

ISSN: **1893-7519**
Date: 2012-10-24

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

METCOOP MEMO Nr 01, 2012
Technical Memorandum

Verification study

HARMONIE AROME compared with HIRLAM, UM and ECMWF

Morten A Køltzow, Karl-Ivar Ivarsson, Solfrid Agersten, Lars Meuller, Dag Bjørge, Ole Vignes, Bjart Eriksen, Per Dahlgren, Martin Ridal, Rebecca Rudsar

Summary

The aim of this report is to give an overview of the quality of the NWP model AROME (a specific configuration of the HARMONIE model system targeted on a 2.5 km resolution) for an integration domain covering Norway and Sweden. The quality of AROME model system (2.5 km resolution) is compared with other model systems operationally available at *met.no* and SMHI. These are ECMWF model (16 km resolution), the HIRLAM (5 km resolution) and the met.no version of the Unified Model (4 km resolution).

In these experiments the AROME model employs ECMWF lateral boundary conditions every 3rd hour. Four test periods are investigated: (1) 10 days in August 2011, (2) 18 days in December 2011/January 2012, (3) 18 days in March 2012, and (4) 30 days in May 2012. Verification of these 4 periods includes summarized score for different parameters and different regions and additionally some case studies. These periods are not long enough to span all possible weather types and do not necessarily give statistical significant results, but act together with other studies, as guidance for future investigations and direction of work in the MetCoOp-project.

AROME is in general not as good as ECMWF on MSLP and total cloud cover, but is better than ECMWF on wind, precipitation (neutral on large precipitation amounts) and temperature. The conclusion for temperature is however dependent on region and type of verification score.

AROME is in most cases better than UM, but not on medium and large precipitation amounts (neutral) and total cloud cover (worse).

AROME is better than HIRLAM on MSLP, wind and precipitation. There are still some problems in AROME for the T2m forecast, but for this parameter HIRLAM is clearly best of the four models. HIRLAM is also better than AROME on total cloud cover.

AROME gives the best result of the four models on wind and small precipitation amounts. AROME is the second best model choice for temperature and mean sea level pressure. Different models beat the AROME system on these parameters, but it is not one particular model which is better than AROME. AROME shows the worst result of the four models on total cloud cover.

This report ends with a brief discussion of 12 issues that should be investigated further for a better understanding and for possible improvements of the MetCoOp AROME 2.5 km model system set-up.

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. In 2011 it was decided that the co-operation should focus on the non-hydrostatic HARMONIE model on high resolution (2,5 km) with AROME physics (General description by Driesenaar [1]). The aim of this report is to give an overview of skills of the HARMONIE model system compared with other model systems operationally available at *met.no* and SMHI today. It is necessary for the institutes to ensure that the model system(s) that will be operational in this co-operation are able to give the users, the customers, researchers and partners weather predictions that are as good as or better than those the institutes can deliver today and those the global models can give. This verification report will serve as a recommendation for which numerical weather prediction model system to use in the co-operation between SMHI and *met.no*. It will also address some issues for the model system that have to be solved.

In this report we focus on verification of deterministic models. We are aware that if a deterministic model has good skills, a probabilistic system based on this model system will also give better results.

The first chapter will focus on the methods and procedures that are used in this verification report, including short descriptions of the model systems. Chapter 3 is the main chapter with four verification studies of different periods, where the summary is done per relevant weather element; sea level pressure, wind direction and speed, temperature, precipitation and cloud cover. In Chapter 4, the verification results from the previous chapter are summarized and put into score tables for a better overview of the verification results. Finally, in Chapter 5, twelve of the main issues that are seen from the verification studies are further discussed.

2 Method and procedures

Numerical weather prediction (NWP) model systems have been developed over the years and the high performance computer (HPC) systems have become faster and more effective, so it is now possible to run operational non-hydrostatic models with high resolution (1-3 km). Such models have the advantage that they can resolve small scale features in the orography and non-hydrostatic processes in the atmosphere. The working hypothesis in this study is that a non-hydrostatic model system with high resolution will have increased skill (i.e. to be closer to the observed weather) than the hydrostatic model systems with coarser resolution (Roebber et. Al [5]).

2.1 Model description

The models evaluated in this report come from the SMHI and *met.no* operational NWP production suites, and from experiments that have been run within the MetCoOp project. ECMWF model data has been retrieved for comparison. The model systems are described shortly in this chapter. A more detailed description of the configuration of the models can be found at MetCoOp's wiki pages (only accessible for IP-addresses from SMHI and *met.no)*:

https://wiki.met.no/nwp/results/available_models/models_versions_and_configurations

2.2 ECMWF

The European Center for Medium range Weather Forecasts (ECMWF) runs a global model (called IFS, Integrated Forecasting System) that also provides boundary fields for the HIRLAM- and HARMONIE models. More information about the ECMWF model system and ECMWF forecasts can be found at
http://www.ECMWF.int/products/forecasts/d/charts

For verification, surface fields from the operational deterministic ECMWF main model runs (starting 00 UTC and 12 UTC) have been retrieved from the MARS archive. The fields have been retrieved on the same grid (~16km) that is used for the initial state and lateral boundaries of the MetCoOp pre-operational runs (and *met.no* H8, HM55 and HM25 runs). An example of a MARS retrieval request is shown on the MetCoOp wiki pages.

2.3 Unified Model (UM)

The UM version used in this verification study is the met.no operational run. UM runs with initialisation and hourly lateral boundaries interpolated from HIRLAM8, ~4km horizontal resolution. The model is run without data assimilation. Bottom of low clouds is changed from 111m to 200m, as in HIRLAM; RH2m calculation is changed to be in respect to water (not ice). A more detailed description about UM at met.no is found in Kristiansen et.al, 2011 [3].
A version based on UM 6.1 with 38 vertical levels were employed for August 2011- The operational UM 4km version at met.no was upgraded mid December 2011 to a setup based on UM 7.7 with 70 levels. After the upgrade the precipitation fields typically showed more pronounced localized effects. Some problems with too weak inland wind speeds were reported, and fixes were introduced on 30/1 and 24/4 2012.

2.4 HIRLAM

HIRLAM is the operational model system at both institutes. HIRLAM is a hydrostatic model (Undén et. al. [8]) and the HIRLAM consortium has decided to stop further development of this model system in order to further develop the high resolution non-hydrostatic HARMONIE model system (Evaluation report by Dudhia [2]).

The HIRLAM model used in this study is SMHI's operational model suite called G05.

The G05 setup is run on a ~5.5 km horizontal resolution (0.05x0.05 degrees) on a rotated lat/long grid with 294x441 horizontal gridpoints and 60 vertical levels. The domain is shown in Figure 1.

G05 is based on the rather old HIRLAM-7.1.2 version with a number of changes, the most notable one being that the Rasch-Kristjansson scheme is updated to roughly HIRLAM-7.3.

G05 is nested into SMHI E11 (HIRLAM 11 km) model, with boundaries every hour and has its own data assimilation.

2.5 HARMONIE AROME

HARMONIE (HIRLAM ALADIN Regional Meso-scale Operational NWP In Europe) will be the successor to the HIRLAM model and is characterized by being a non-hydrostatic meso-scale model (Seity et.al. [6]). These types of models on finer scales are assumed to give a better description of the meteorological processes due to the possibility to resolve the more elaborated prediction of vertical motion and the convection.

The HARMONIE runs included in this study are from a setup called AM_Hires1, that for the MetCoOp project is run in real-time at SMHI with AROME (Application of Research to Operations at MesoscalE) physics. AM_Hires1 was based on the MetCoOp-branch of HARMONIE-36h1.4, which is a copy of the official version registered by the HIRLAM consortia (tagged in the version control system), and was run on the first version of MetCoOp's high resolution domain with 65 vertical levels and 540x900 horizontal gridpoints at 2.5 km horizontal resolution. The domain is shown in Figure 1.

2.6 Rerun of test periods

In order to avoid comparing models run under different circumstances, e.g. different input data and/or different start-up periods, it was decided to rerun the chosen test periods with the same input data and configuration of the model. The reruns have, for this report, been done for two periods, 2011-08 (chapter. 3.4) and 2011-12 (chapter. 3.5) for HARMONIE with AROME physics and HIRLAM.

The HARMONIE AROME run, AM_Hires1B, is a copy of the real-time AROME at SMHI. The 2011-08 period was rerun with the reference version, but with the namelist variable LWMOCLOUD set to "FALSE". This setting effects the post-processing of clouds somewhat, and was set to TRUE for the 2011-12 period. Changes were made in HARMONIE to post-process 'fog' and 'low clouds' as is done in HIRLAM. A post-processing of cloud base was also introduced to be able to verify against observed quantities from automatic stations.

The HIRLAM GM05 run is a copy of SMHI operational HIRLAM G05 setup. It was decided that GM05 should be run directly coupled with ECMWF boundary data, not nested in a coarser HIRLAM (which is the case for SMHI G05 HIRLAM). The domain had to be rotated slightly (see Figure 1) to be able to be run with the same ECMWF boundary data as the rerun with AROME.

All reruns had the same 'cold start', i.e., they started the first forecast from interpolated boundary fields. Forecasts lengths are +48 hours at 00 and 12 UTC and +6 hours at 06 and 18 UTC.

No Large Scale Mixing procedure was applied for the runs. For the period 2011-08 (chapter. 3.4) the AROME version was run both with upper air data assimilation (3DVAR) and with blending while the 2011-12 period was only run with 3DVAR.

The MetCoOp project group has also run ALARO 5.5 km on the MetCoOp domain (AM_Hires1). The ALARO run, AO_Hires55, was run on 5.5 km horizontal resolution, 65 levels, on a domain with 270x450x65 gridpoints covering the Hires1 domain. It was run with 3DVAR. ALARO is another set of physics schemes in the HARMONIE model system and has been tuned to run on 5,5 km resolution. Compared to AROME at 2,5 km resolution ALARO is therefore not so expensive to run on a larger domain. If the institutes want or need local area model results that are available earlier, and possibly also more frequently updated than ECMWF, the ALARO system should be considered.

The comparison between AROME and the operational models HIRLAM and UM and the global model from ECMWF will be presented in this report. Verification between AROME, ALARO and other (operational) models than mentioned above has also been done during the last year but is not included in this report. Some of these comparisons will be presented in later reports.

2.7 Experiences of AROME in general

Other meteorological institutes running AROME seems to be mainly satisfied with the model, but some problems are reported. Fog occurs too often over sea, perhaps also over land (The Netherlands, [7] van der Veen). Spurious waves in the MSLP field in very windy situations are reported (perhaps due to some instability problems; Denmark). In addition an on-off behaviour of the cloud cover is seen in AROME. This means too often clear sky or total overcast (several institutes). The cause may be that the subgrid scale variation of moisture is not large enough in the cloud cover parameterization, since the assumed probability distribution function of moisture used, gives too small spread. The convection cells created by the model seem sometimes to be too intense (several institutes). This may cause too much precipitation and too intense downdrafts in connection with the convective cells. This is an old problem, which only partly has been solved. The cause of it seems to be complex and not fully understood.

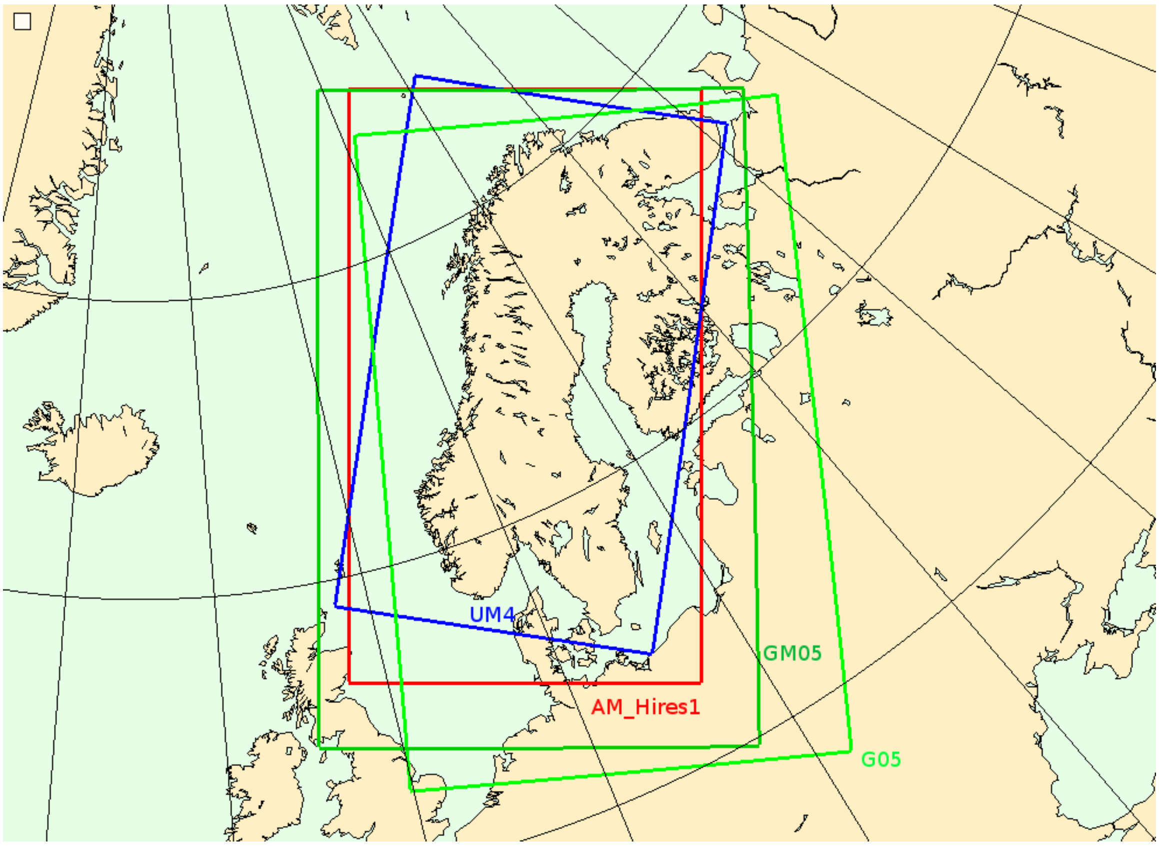


Figure 1 : Model domains for HARMONIE AROME (AM_Hires1), Unified Model (UM4), and HIRLAM (GM05, used in chapter 3.4 and 3.5 and G05 used in chapter 3.6 and 3.7)

2.8 Input data to the model runs

The data used for the reruns of the chosen peridos, AM_Hires1B and GM05, were *met.no's* operational observations and boundary data.

The boundary data are from ECMWF's BC project with new boundaries every 6 hour and a temporal resolution of 3 hours. No analysis was used but every run had 6 hour old boundaries in the same way as operational NWP runs.

Only conventional observations, SYNOP, SHIP, TEMP, PILOT, BUOY, AMDAR and AIREP were used.

SMHI's G05 model is nested in SMHI's E11 (HIRLAM 11) and uses SMHI's operational observations, which differ slightly from *met.no*'s observations. Only conventional observations are used.

The AROME model (AM_Hires1) for the March 2012 and May 2012 periods, was run with SMHI operational boundaries and observations in near real-time at SMHI for the MetCoOp project. The boundaries are from the ECMWF BC project and have a horizontal resolution of 0.25x0.25 degrees (~25 km) and a time resolution of 3 hours.

Observation basis

The verification results presented in this report compare the forecasts to the synoptic observations. The station list was updated to include all synoptic observing stations in Norway and Sweden. The numbers of observations used varies according to the period and the parameter. In Norway there has been an increase in the number of synoptic observing stations over the last few years although some stations have also closed down. It is important to bear in mind that model values are grid averages, while the observations are point observations. A comparison is meaningful when the model output is used in issued point forecasts. For model quality assessments the representativeness of the observations for comparison with model output is an issue.

Table 1: Verification variables and differences between simulated and observed variables are briefly explained.

Model Variable	**Observed variable**	**Comments**
Mean sea level pressure (MSLP)	Pressure reduced to sea level	Model variable: The reduction of model pressure to sea level depends on the model quality of temperature. Observed variable: At *met.no* the pressure reduced to sea level is calculated with an algorithm based on the pressure measured on the station, and the air temperature at the station at the same observation time. The algorithm uses also the annual mean value of humidity and the annual average air temperature at the station, and the reference height of the air pressure on the station. For more information, see: https://kvalobs.wiki.met.no/doku.php?id=kvalobs:aggregering It is assumed that the method is more or less the same at SMHI and other European countries.
2m air temperature (T2m)	2m air temperature (T2m)	T2m from model is available as raw model temperature (averaged over grid cell). Both the raw model output and a T2m height adjusted to station height (with 0.6°/100m) is verified against observed T2m.
Wind 10 m (speed and direction)	10 min averaged wind speed and direction	For significant weather, i.e. high wind speeds at the coast and in the mountains, observed wind can vary considerably in time and more than in the models. Verifying against observations every (third) hour will therefore miss out some of the observed cases with high wind speeds.
Total cloud cover [0,100]	Total cloud cover [0,8]	Model percentage cloud cover is converted into octas as the observations. In Norway the observation of total cloud cover is done manually, while in Sweden it is done automatically for most stations. Differences in score between Norway and Sweden may therefore be due to model differences or observational differences.
12h precipitation	12h precipitation	No attempt is done to correct the observations for containment failure (i.e. due to wind drift). The observed precipitation is therefore a lower limit for observed precipitation and especially in winter. An interpretation of the bias for precipitation should therefore be done with care for different regions/seasons.

There are no attempts on area averaging of the verification scores. Model performance in regions with high observational density will therefore have greater impact on the summarized verification score than regions with few observations. This is partially improved by verifying with different station lists (Table 2). The number of observations is also given in Table 2. In general there is an increase in the number of observations from the first to the last period for automatic observations (MSLP, T2m, wind, precipitation), while a small decrease for manual observations of cloud cover is seen.

Table 2. Station lists and number of observations.

Name **Station list**	**Description**	**Number ofobservations**				
		MSLP	**T2m**	**10 m Wind**	**12hr Precip**	**Cl. Cover**
METCOOP	All Norwegian and Swedish SYNOP stations	243 - 249	432 - 452	397 – 407	296 – 310	92 – 99
METCOOP **coast**	Norwegian and Swedish SYNOP stations at the “outer” coast line	43- 46	65 – 69	63 – 67	32 – 34	8 – 9
METCOOP **mountains**	Exposed mountain SYNOP stations in Norway and Sweden.	3- 5	9 – 11	9 - 11	3	0
METCOOP **north**	All SYNOP stations north of 67.5°N	55- 60	84 – 93	79 – 85	48 – 54	29 – 32
METCOOP **Mid**	All SYNOP stations between 62.5°N and 67.5°N	109- 110	180 – 181	162 – 164	116 – 123	52 – 54
METCOOP **south**	All SYNOP stations south of 62.5°N	176 - 177	287 – 308	268 – 276	197 – 205	73 – 77
METCOOP **City**	SYNOP stations close to the largest cities in Norway and Sweden.	13	15	14 – 15	10	9
0-200m.a.s.l.	All SYNOP stations below 200 m.a.s.l.	203 - 209	304 – 312	277 – 283	197 - 205	101 - 108
200-600m.a.s.l.	All SYNOP stations between 200 and 600 m.a.s.l.	72 - 73	145 – 149	130 – 133	117 – 120	38 – 40
600-1000m.a.s.l.	All SYNOP stations between 600 and 1000 m.a.s.l.	21	42 – 46	38 – 40	27 – 30	7 – 9
1000m.a.s.l.	All SYNOP stations situated higher than 1000 m.a.s.l.	6- 9	10 – 17	10 – 13	4 – 5	1 – 2

Norway	All Norwegian SYNOP stations	137- 142	233 – 254	206 – 216	151 – 164	66 – 71
Sweden	All Swedish SYNOP stations	106- 107	198 – 199	191 – 192	143 – 145	26 – 28
Norway coast	Norwegian SYNOP stations at the “outer” coast line	30- 33	41 – 45	39 – 43	12 – 14	8 – 9
Sweden coast	Swedish SYNOP stations at “outer” coast line	13	23 – 24	23 – 24	19 – 20	0

2.9 Verification methods and scores

The main goal for the verification is to use comparison methods that will give objective results with possibility to derive clear conclusions. The weather in the atmosphere is by nature chaotic and the model systems are mathematical equations mapped into a numerical model system that will give an approximate picture of the weather situation at a given time and place. The model systems in this comparison will behave differently in different weather situations and will therefore not have the same strengths and weaknesses. This makes the comparison quite difficult and it is therefore necessary to use skill scores that will complement each other in order to give a more complete picture of the skills of the four model systems.

The standard HARMONIE verification system (WebGraF) is applied for the verification presented in this report.

2.10 Explanation of common verification scores

In this chapter the different verification and evaluation scores are explained in detail. The 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.

Skill score: (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 forecast with a large positive or negative BIAS.

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

3 Verification of test periods

In this chapter four different periods are chosen for a comparison between the four models described in the previous chapter.

This chapter contains a brief explanation of the models used in the comparison, and where to find a complete set of figures. Thereafter there are verification results for these four periods.

3.1 Periods

The four different periods have been chosen for the following reasons:
* August 2011: It represents a late summer period with heavy rain causing a lot of damage over southern Norway and south western Sweden (especially August 14-16)
* Late December 2011- early January 2012: It represents a mild and very windy winter period. Considerable damage due to high wind speed, including severe wind gusts, was caused in both countries.
* March 2012: It represents a late winter/early spring period.

* May 2012: It represents a late spring/early summer period and covers all 31 days in May.

The periods are quite short, but should give a representative overview of the different skills of the models. None of the winter periods in the study are very cold, but the results will nevertheless indicate the behaviour of HARMONIE compared to the other models during cold winter conditions. The very cold 2010 winter will be investigated later.

The summer period is quite special due to heavy precipitation and is quite short, but the May period of 2012 is included in the study so the summer conditions should be quite well covered.

The study lacks a period of normal autumn, or rainy and windy autumn, but the winter period has high wind speeds and should cover this parameter quite well.

3.2 Models

HARMONIE-METCOOP (AROME, cy36h1.4 with 3DVAR and horizontal resolution 2,5 km) is compared with operational forecasts available from the ECMWF model, the HIRLAM Model (G(M)05, SMHI) and the Unified Model (UM4, *met.no*), hereafter named AROME, ECMWF, G(M)05 and UM4. The models are described in more details in previous chapter 2.1.

The HIRLAM runs, named GM05 and G05, differ in that the periods August 2011 and December – January 2012 run on the same domain as AROME, while the two other periods are run on the operational SMHI domain.

3.3 Additional verification figures

More figures from the HARMONIE verification system WebGraF are available for the verification studies for:

- the period13-23 August 2011 (chapter. 3.4):

http://metcoop.met.no/verif/201108_SummerReport2012_export/

- the period of December 2011- January 2012 (chapter 3.5) :

http://metcoop.met.no/verif/201112_SummerReport2012_export/

- the period of March 2012 (chapter 3.6) :

http://metcoop.met.no/verif/201203_SummerReport2012_export/

- the period of May 2012 (chapter 3.7):

http://metcoop.met.no/verif/201205_SummerReport2012_export/

This site is only available for IP addresses from *met.no* and SMHI.

3.4 Verification study of 13 – 23 August 2011

In this study the rerun of the MetCoOp model HARMONIE AROME (2,5 km) is compared with the rerun of HIRLAM (GM05) and UM (4 km) from *met.no* and the ECMWF forecasts (domain shown in Figure 1).

The verification results further in this chapter are divided in the main weather elements that are most relevant for users and give a good overview of the weather situation. The figures show different skill scores for the four models and give a picture of the behaviour of the different models, how well they fit the observations and how the models compare with each other.

3.4.1 Meteorological description of the period

The unsettled weather was mainly due to low pressure activity in Skagerrak and easterly winds as low-troughs occupied southern Scandinavia. During August 12 to 17, a number of lows passed from the British Isles towards the middle of Scandinavia. A high pressure system started to develop over the Barents Sea. This caused south-easterly winds, and warm and moist unstable conditions over a large part of Scandinavia during August 14 and 15. During August 18 to 20 the high pressure system in the north weakened, but low pressure systems continued to pass over Scandinavia. At the end of the period a weak ridge caused slightly more stable weather conditions.

This period was dominated by cases with large (convective) precipitation amounts and was extremely wet in parts of Norway, especially south-eastern parts of the country with 195 % of mean precipitation; while Norway as a whole received 140 % of normal values.

Even though temperatures seldom climbed above 25 degrees Celsius during daytime, the mean temperature for the period June - August was 1.2 degrees above normal. This was mainly due to unsettled weather with clouds during night time.

In Norway the large precipitation amounts resulted in a flood on August 16. Large precipitation amounts were also experienced August 14. (15-18UTC) in the Gothenburg region in southern Sweden.

3.4.2 Verification results of mean sea level pressure (MSLP)

In terms of RMSE, ECMWF is in general better than the other models. AROME and GM05 show very similar RMSE scores, while UM4 shows larger errors (Figure 2). There are only minor BIASes for all models, but AROME has a tendency to underestimate the MSLP in the last part of the forecast. For stations situated high above sea level (more than 600 m.a.s.l.) the quality of AROME, GM05 and ECMWF are quite equal (not shown). It is also clear that the errors increase with height above sea level for all models. However the daily average RMSE is well below 1 hPa, the exception being the mountainous regions.

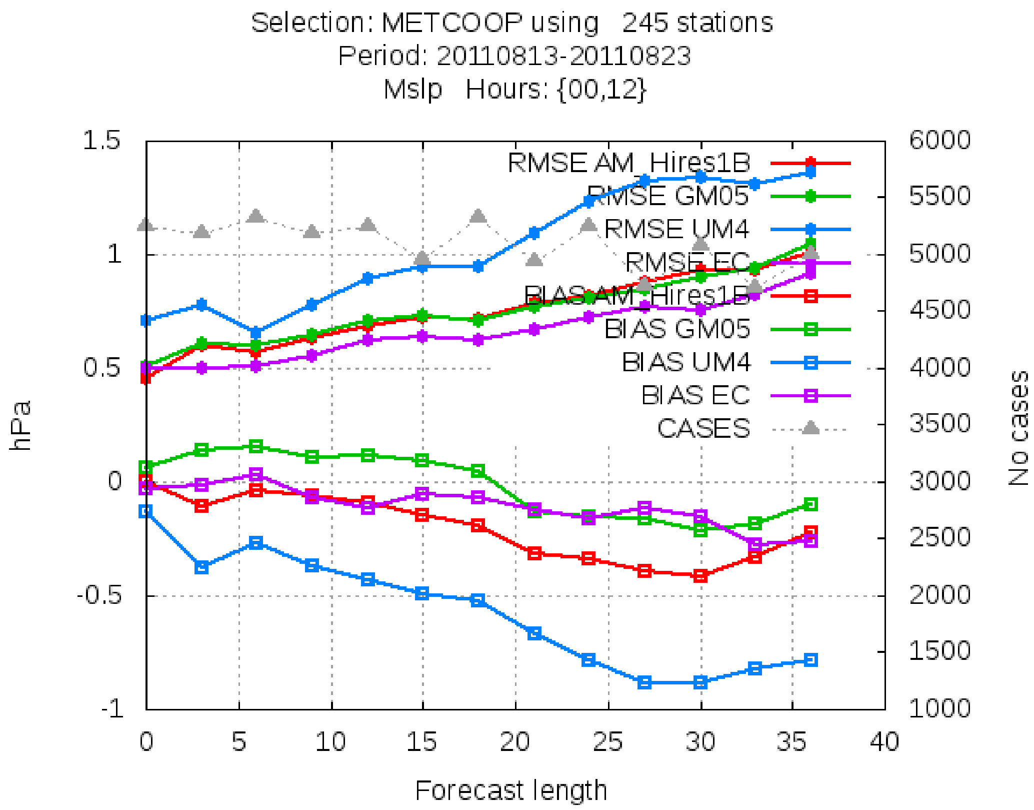


Figure 2: August2011, RMSE and BIAS for MSLP (averaged over all stations).

3.4.3 Verification results of 10 m wind direction

For wind direction AROME shows almost no BIAS and less systematic errors than the other models (Figure 3). It has also a slightly smaller or equal RMSE compared with the other models. All models experience smaller RMSEs at the coast compared to inland stations.

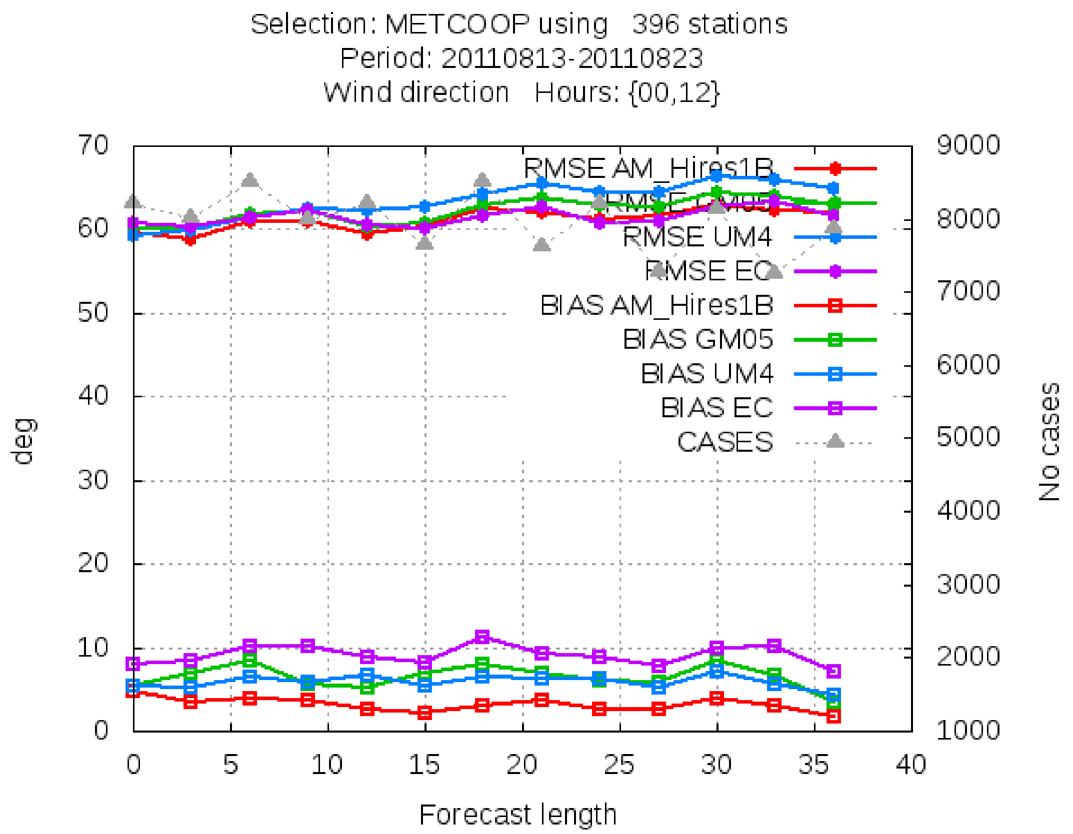


Figure 3: August2011, RMSE and BIAS for wind direction (all stations).

3.4.4 Verification results of 10 m wind speed

AROME shows a negligible BIAS averaged over all Norwegian and Swedish station, while the other models show a small positive BIAS (up to 0.5 m/s). The RMSE is below 2 m/s for all models and slightly better in AROME, GMO5 and UM4 compared to ECMWF. Averaged over all Norwegian stations, there are only minor differences between the different models (somewhat similar RMSE and only minor BIASes for AROME, GM05 and ECMWF, while UM4 shows a small negative BIAS (0.25 m/s). However for Sweden there are clear differences in score. A positive BIAS is found in ECMWF (0.75 m/s), GM05 and UM4 (0.5 m/s), while AROME shows a small negative BIAS (-0.25 m/s). AROME has slightly smaller RMSE than GM05 and UM4, which again is better than the ECMWF model. The errors are in general larger in Norway than in Sweden and are most probably associated with higher wind speeds in Norway.

There are further differences in the AROME behaviour between different station lists (regions) indicating that the systematic errors averaged over all stations hide model weaknesses. AROME underestimates the wind speed on average in the mountains (0.5 – 1.0 m/s), in the northern part (1.0 m/s) and the mid-part (0 – 0.5 m/s) of the integration domain and for stations situated between 200 and 1000 m.a.s.l. (0.25 – 1 m/s). Furthermore, AROME slightly overestimates the wind speed on average at the coast (0.25 m/s) and shows larger RMSE average over the 15 largest cities compared to the other models.

In terms of frequency (BIAS), all models show a deficiency in not forecasting wind above 10 m/s. AROME shows a better agreement with observed frequency (averaged over all stations, stations above 600 m.a.s.l., Norwegian coast and stations in the mid-part of the integration domain) than the other models, but we emphasize that there are differences between different regions. For stations situated between 200 to 600 m.a.s.l. AROME shows too few forecasts of moderate wind speeds (3 – 8 m/s).

AROME gives higher (better) ETS for all evaluated thresholds using all stations. AROME is followed by UM4 and GM05, while ECMWF shows less ability to forecast the exceedance of different thresholds (Figure 6). For all models the quality decreases with increasing wind speed. These findings are valid for both Norway and Sweden. Based on the ETS score AROME is better than the other models for high wind speeds at stations above 200 m.a.s.l., stations in the northern and mid-part of the integration domain. At the Norwegian coast UM4 and AROME clearly perform better than GM05 and ECMWF, while at the Swedish coast the difference is less. The higher ETS for AROME is mainly a result of a higher hit rate (HR) associated with an increase of the false alarm ratio (FAR) for windy regions, while the FAR is equal or slightly less than the other models in other regions. However, compared with the other models AROME shows lower ETS over the largest cities with a lower hit rate than ECMWF and a higher FAR than the other models

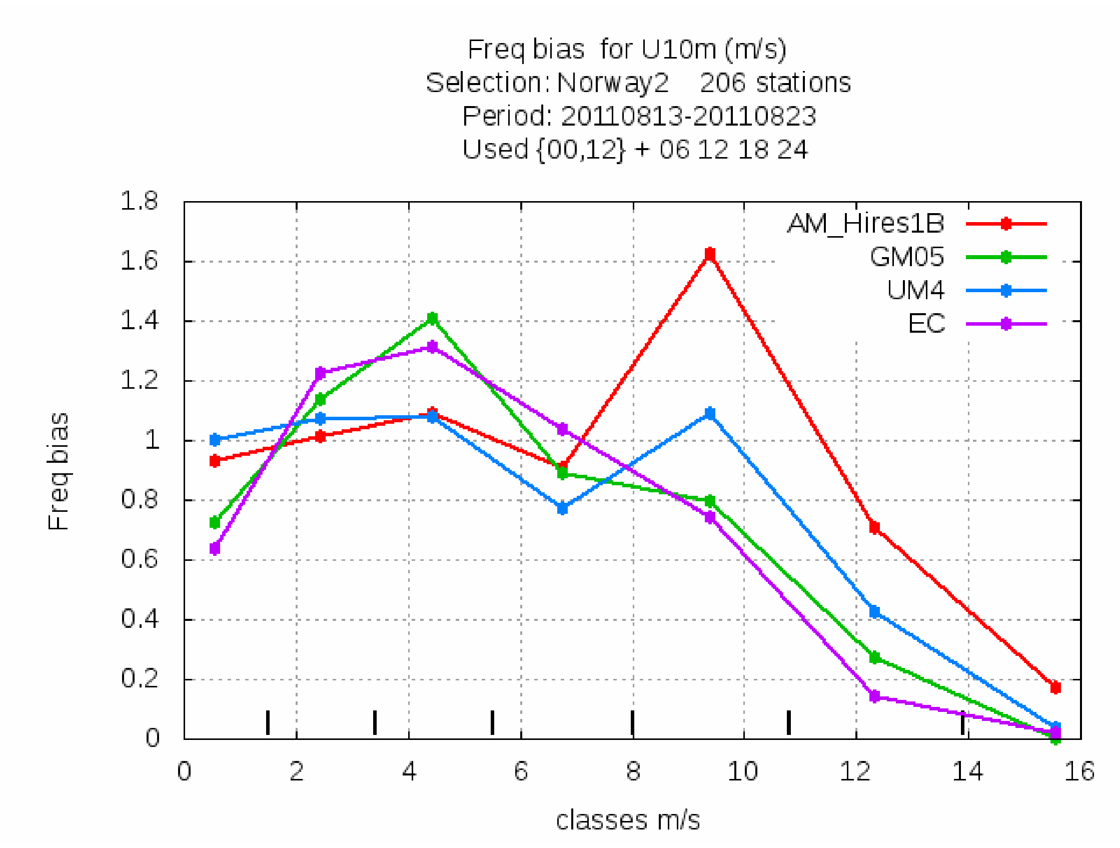


Figure 4: August 2011, Frequency BIAS (FB) wind speed averaged over all Norwegian stations

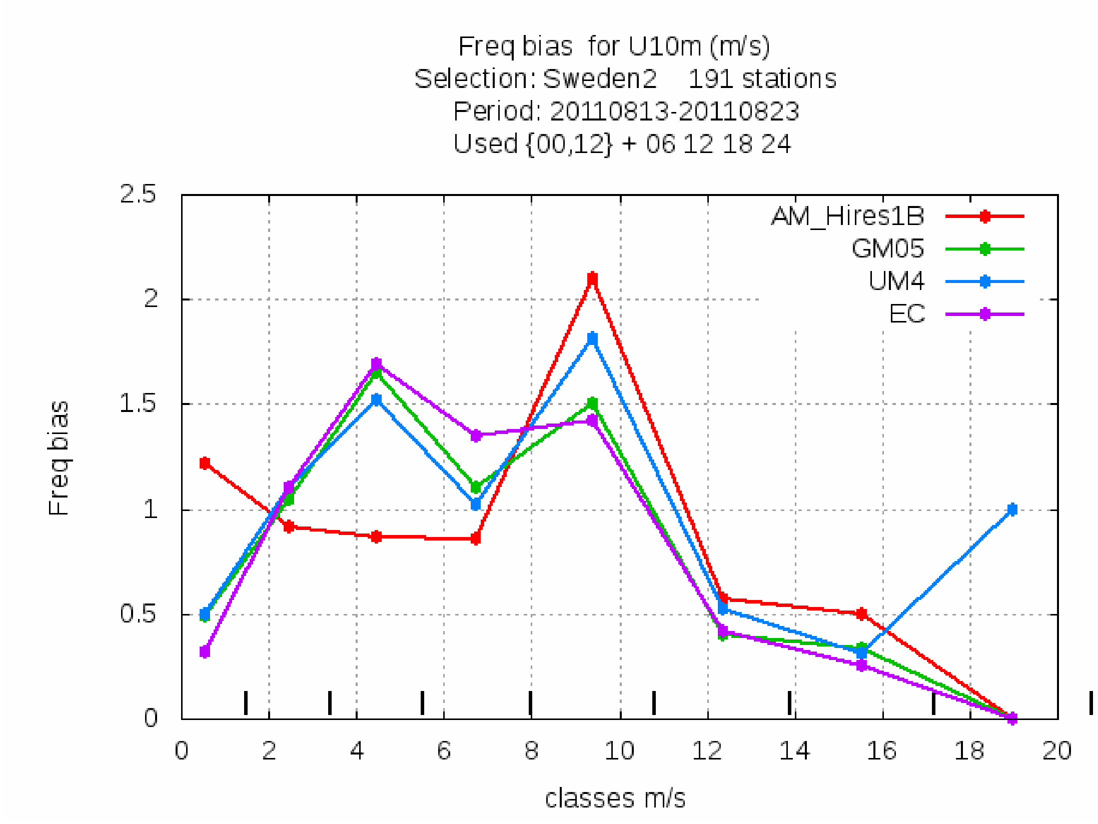


Figure 5 : August 2011, Frequency BIAS (FB) wind speed averaged over all Swedish stations

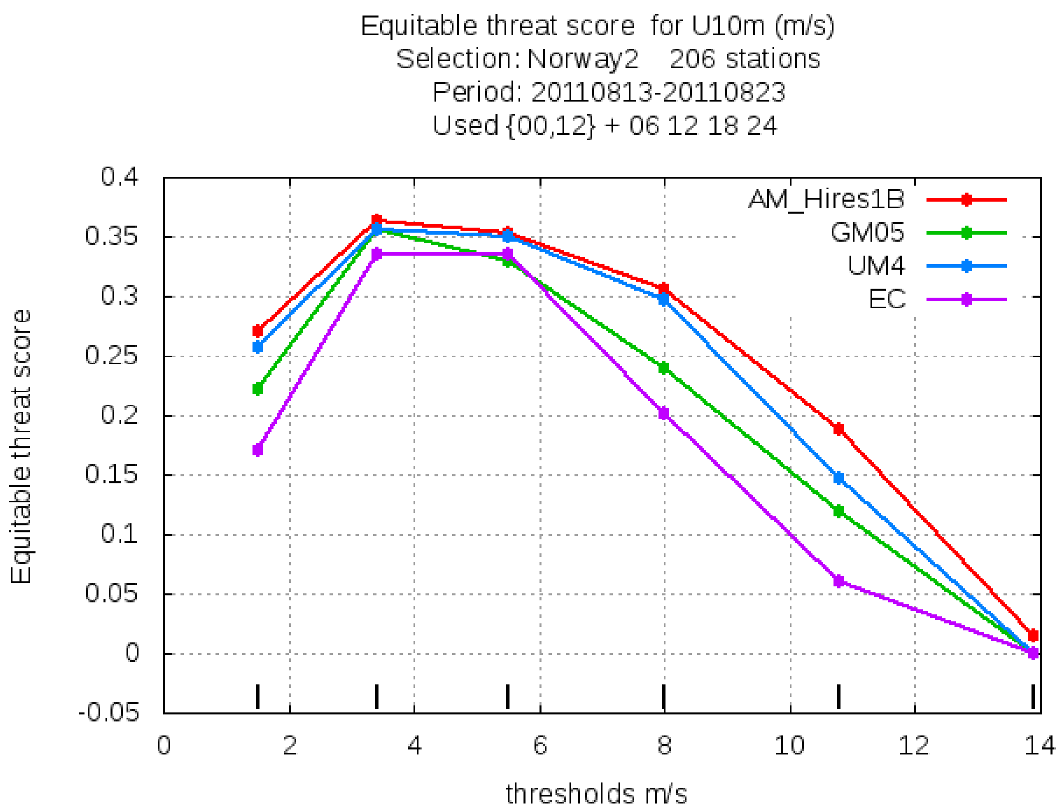


Figure 6: August 2011, ETS wind speed averaged over all Norwegian stations

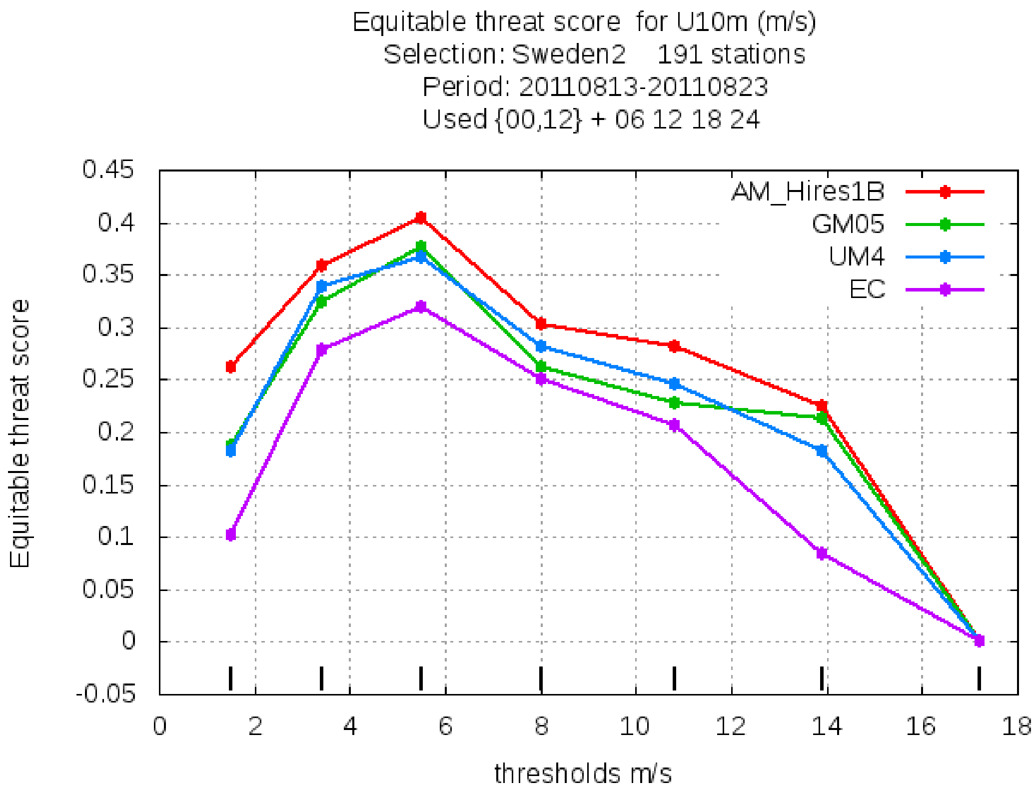


Figure 7 : August 2011, ETS wind speed averaged over all Swedish stations

3.4.5 Verification results of near surface temperature (T2m)

AROME shows in general good verification results with respect to T2m with smaller RMSE and BIAS than the other models (Figure 8). However, this is partially equalled with height adjustment (from model to station height), meaning that at least some of the smaller errors are due to better resolution. After height adjustment, AROME has a RMSE which is similar to ECMWF and GM05, but has smaller RMSE than UM4. AROME and GM05 show smaller systematic errors (BIAS) than ECMWF which has negative BIAS and UM4 which has positive BIAS.

With respect to the relative quality of AROME it is better for Norwegian stations compared to Swedish stations. In Norway, AROME shows less RMSE than the other models and negligible BIAS, while in Sweden both GM05 and ECMWF display smaller RMSE and BIAS. Scatter plots and BIAS maps show that the largest part of the errors from Swedish stations is due to too warm night temperatures. The negative BIAS for AROME is most pronounced at the coast, in the mountains and in the northern part of the integration domain.

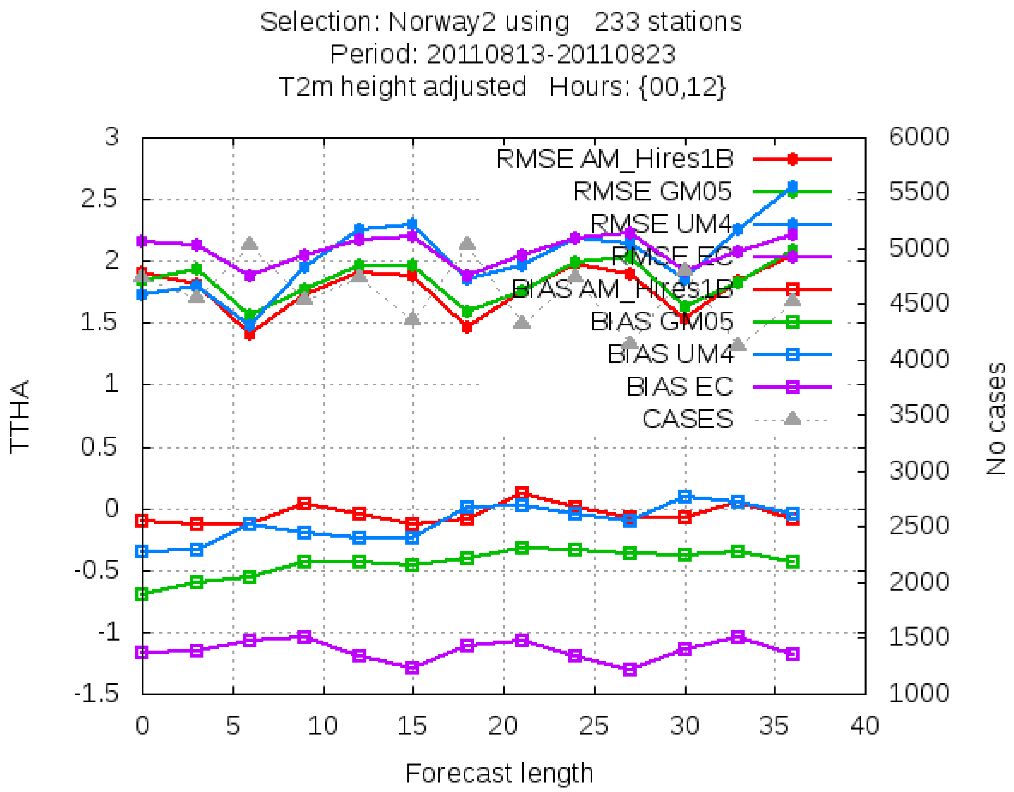


Figure 8: August 2011, RMSE/BIAS T2m averaged over all Norwegian stations.

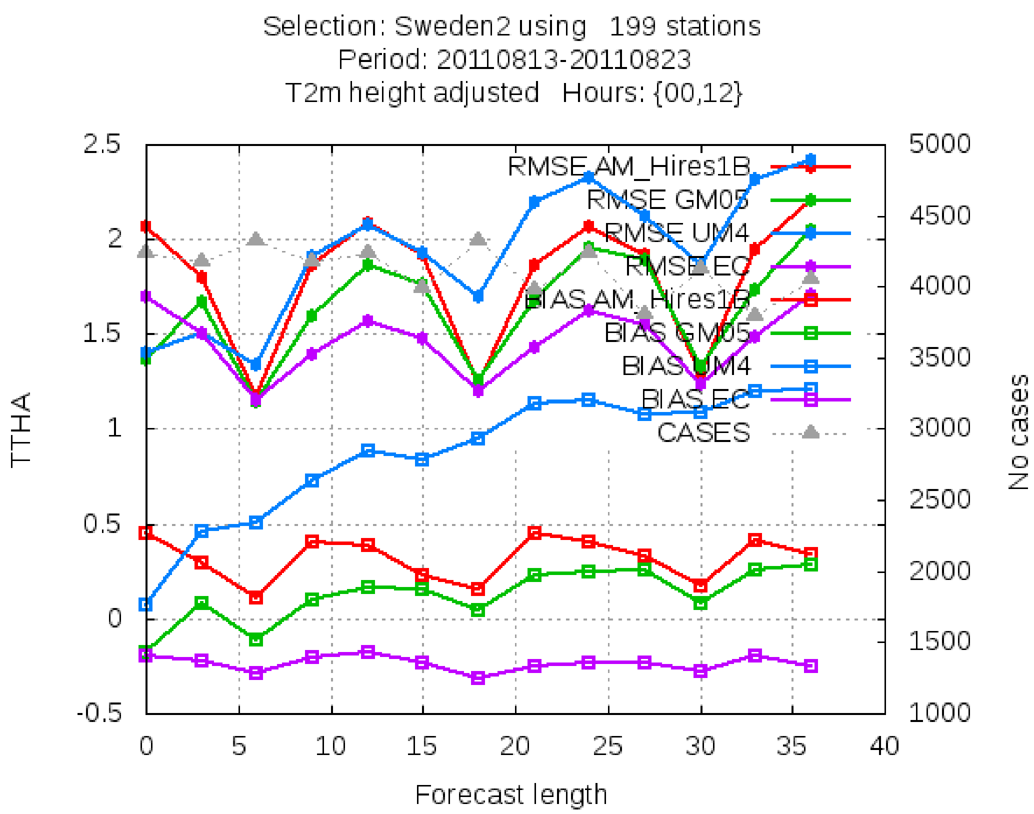


Figure 9: August 2011, RMSE/BIAS T2m averaged over all Swedish stations.

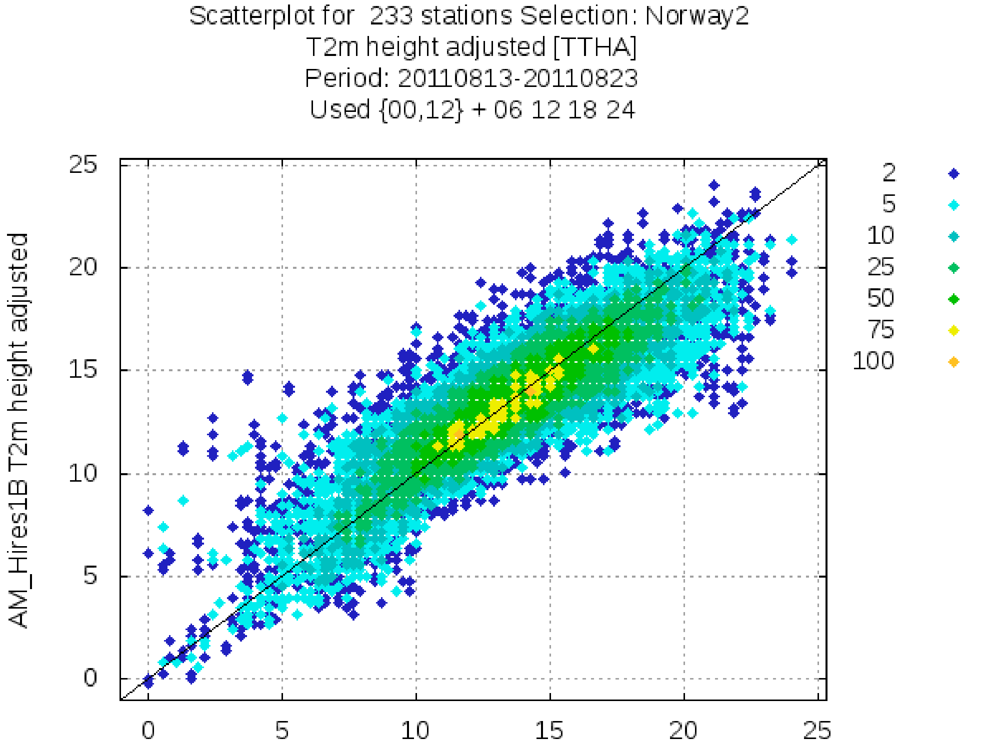


Figure 10: August 2012, scatter plot of AROME T2m averaged over all Norwegian stations.

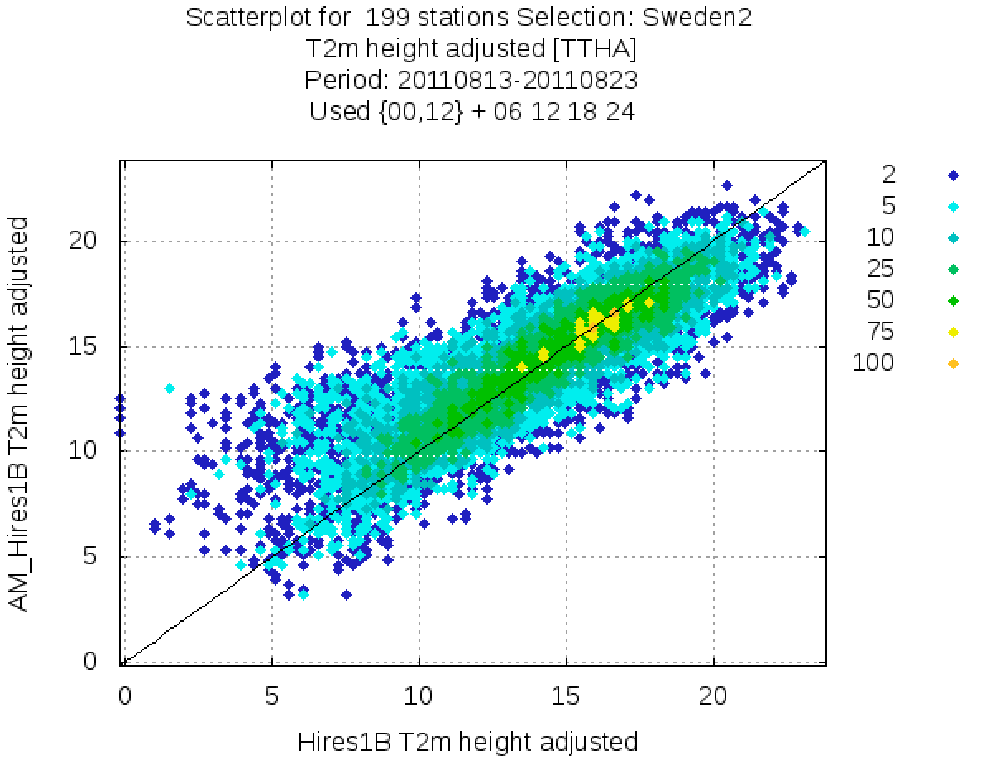


Figure 11: August 2011, scatter plot of AROME T2m averaged over all Swedish stations.

3.4.6 Verification results of precipitation (12hr)

All models slightly overestimate the precipitation amount in the period (0 – 0.5 mm/12h) and show an increase in positive BIAS with height. Even though the observational errors are significantly smaller in summer than in winter it is reasonable to believe that the overestimation is a desired feature since observed precipitation amounts can be regarded as a minimum of actual precipitation.

The total amounts of precipitation for all models are reasonable, but there are larger deficiencies regarding the frequency of occurrences. AROME is in better agreement with the observations than the other models. AROME overestimates the occurrences of precipitation by 10% followed by GM05 (20%), UM4 (30%) and ECMWF (almost 50%).

In terms of ETS there is no clear "best model" for the period (averaged over all thresholds). AROME scores better for small precipitation amounts (less than 0.5–1.0 mm/12h), UM4 scores better for intermediate amounts (1–10 mm/12h). While no clear "best model" is found for precipitation amounts larger than 10 mm/12h. AROME shows lower ETS than the other models (Figure 13). There are, however, large differences in terms of quality for different regions. In general, AROME shows lower hit rate for all thresholds, but also lower false alarm ratio for small precipitation amounts.

From the 15th to 16th of August large precipitation amounts led to flooding and severe damage in south east Norway. In the area south and east of Trondheim, north of Lillehammer and close to the Swedish boarder between 50 and 75mm/24h were observed at 14 stations and a maximum of 111mm/24h was measured at one of the *met.no* climate stations Håsjøen. Such precipitation amounts are especially critical since parts of the region are usually quite dry. The observational series at Håsjøen is too short to calculate a normal, but the normal August precipitation at Langen, close to Håsjøen is 75 mm for August.

In the next figures 24h precipitation from AROME, UM4, GM05 and ECMWF is presented together with observations. ECMWF shows a broad south-east band of precipitation with a maximum 24h precipitation exceeding 50mm/24h. The three other models display much more fine scale patterns with several south-east bands of precipitation and with higher maximum precipitations (UM4 and GM05 exceeding 80mm/24h and AROME exceeding 70 mm/24h). Even though the location of the precipitation in ECMWF is good it underestimates the largest observed amounts. The high resolution models are in better agreement regarding the maximum precipitation amounts, but show difficulties in the placement.

The large spatial variation in the high resolution models highlights the importance of also evaluating the forecast with other skill scores than traditional verification scores based on station data. This will allow taking fully advantage of the information available from the precipitation forecasts from the high resolution models and will be investigated later.

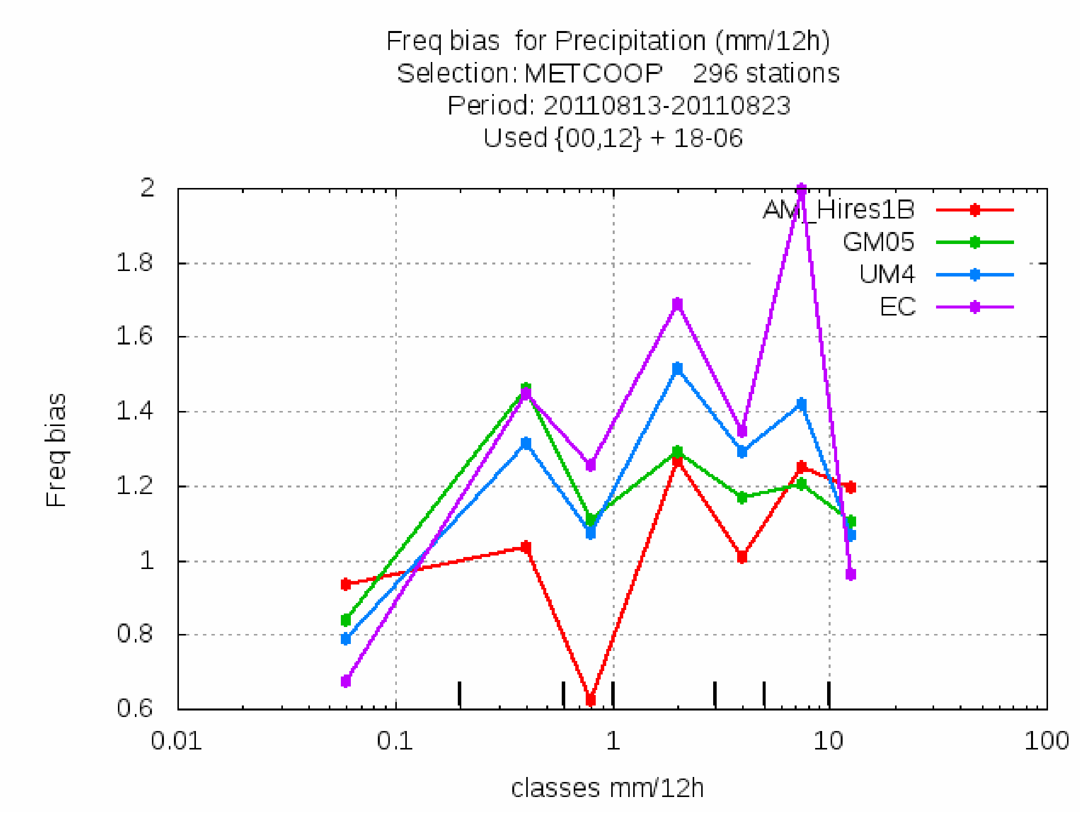


Figure 12: August 2011, Frequency BIAS (FB) 12hr precipitation averaged over all stations.

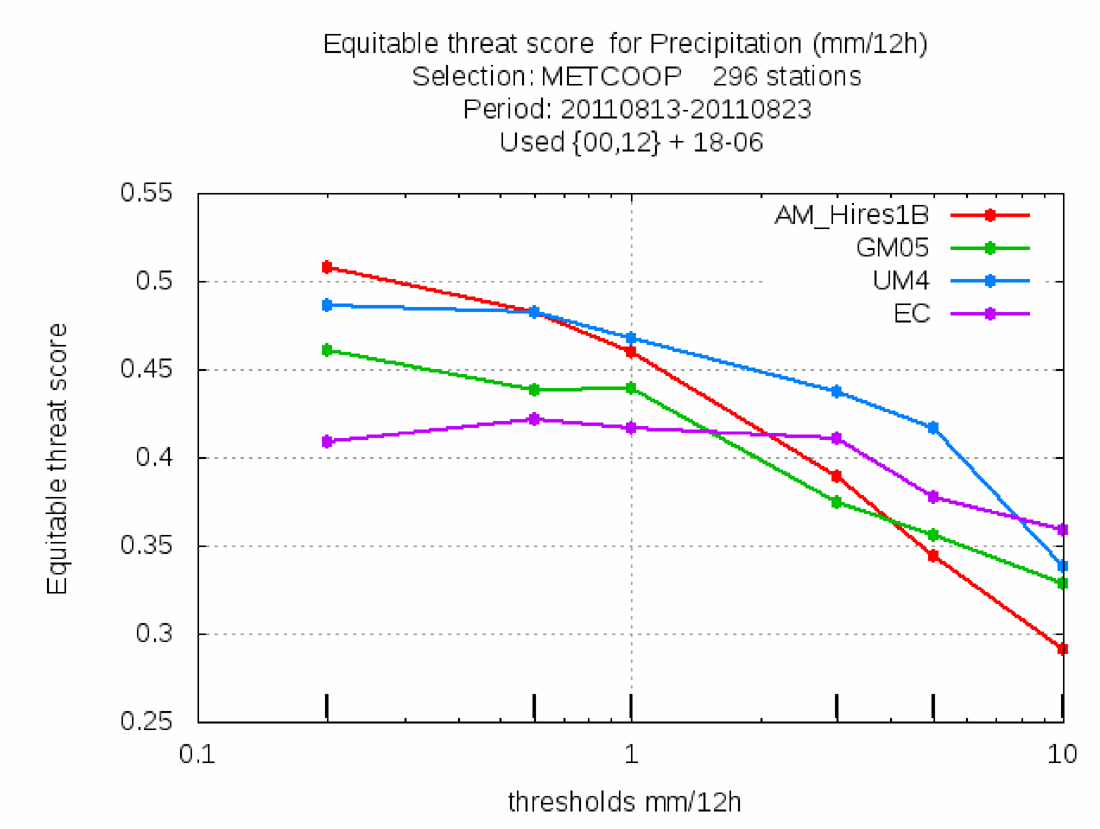


Figure 13: August 2011, ETS 12hr precipitation averaged over all stations

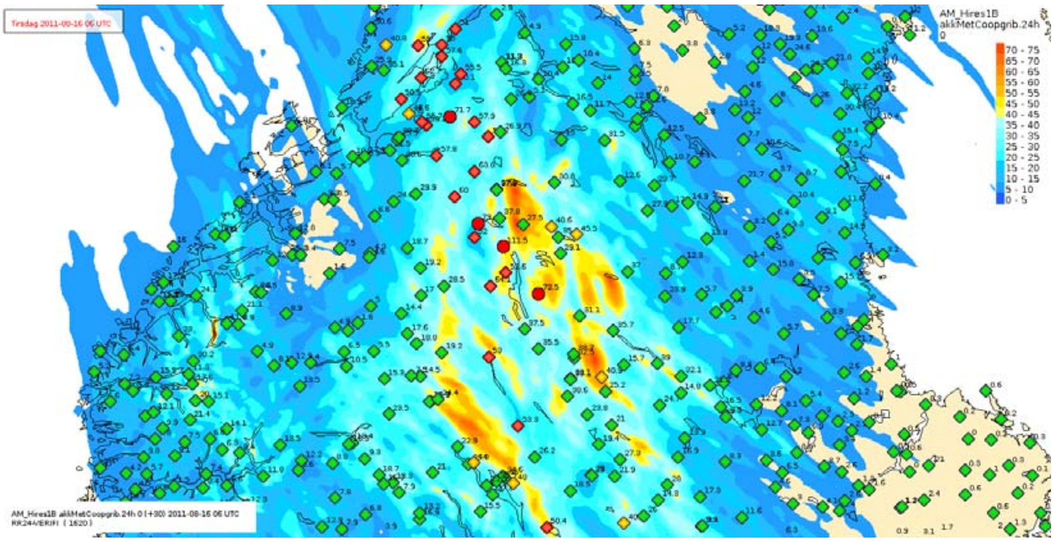


Figure 14: August 16 06UTC 24hr precipitation AROME and observations (red observations > 50mm/24hr, yellow observations >40mm/24h)

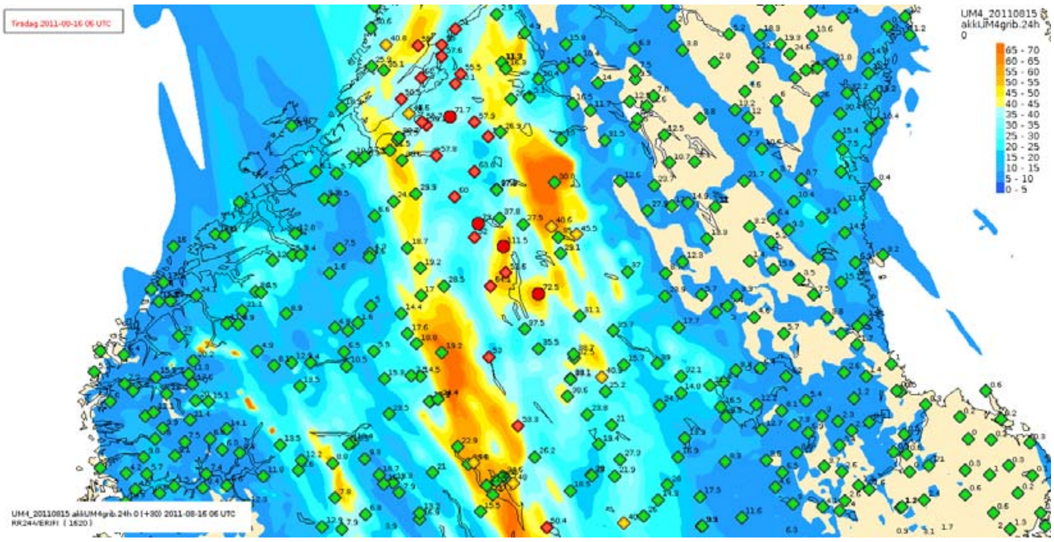


Figure 15: August 16 06UTC 24hr precipitation UM4 and observations. (red observations > 50mm/24hr, yellow observations >40mm/24h)

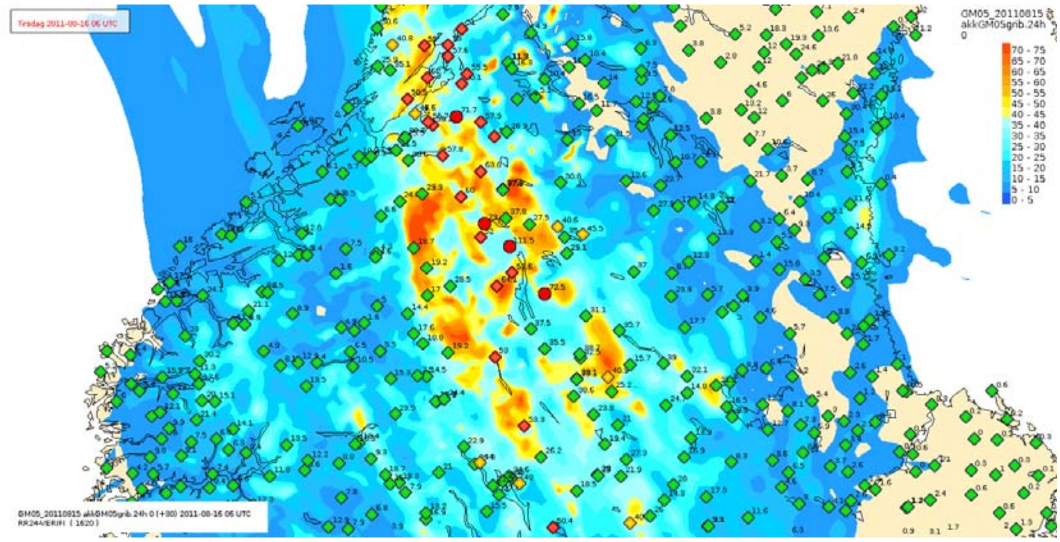


Figure 16: August 16 06UTC 24hr precipitation GM05 and observations. (red observations > 50mm/24hr, yellow observations >40mm/24h)

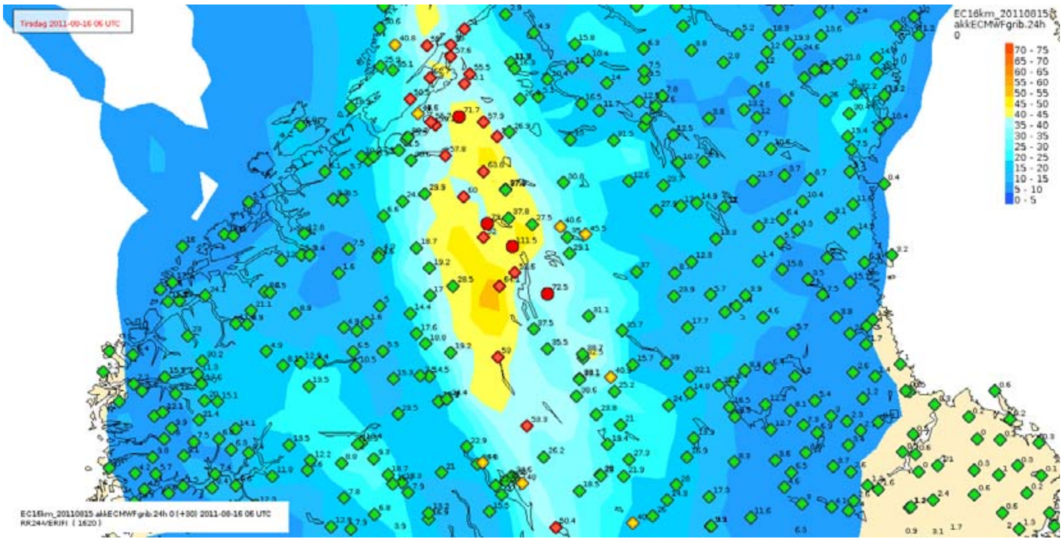


Figure 17: August 16 06UTC 24hr precipitation ECMWF and observations. (red observations > 50mm/24hr, yellow observations >40mm/24h)

3.4.7 Verification results of total cloud cover (TCC)

Averaged over all stations, UM4 give the smallest BIAS and RMSE, followed by GM05 and ECMWF and then AROME with the largest RMSE. In Sweden all models overestimate the observed total cloud cover (from 0.5 – 1.0 octa) with AROME and ECMWF showing the largest overestimation. AROME also shows largest RMSE for lead times up to 15h. In Norway the situation is different, the models only have small BIASes and AROME has a negligible BIAS, but clearly the largest RMSE for all lead times.

Averaged over all Norwegian and Swedish stations UM4 shows an impressive Frequency BIAS (FB), while AROME, GM05 and ECMWF have the tendency to forecast the extremes (no clouds/ only clouds). However these results are dominated by the Norwegian stations (3 times as many as the Swedish stations) and there are large differences in model scores between Norway and Sweden. In Sweden all models show too few occurrences with no clouds and too many occurrences with total cloud cover.

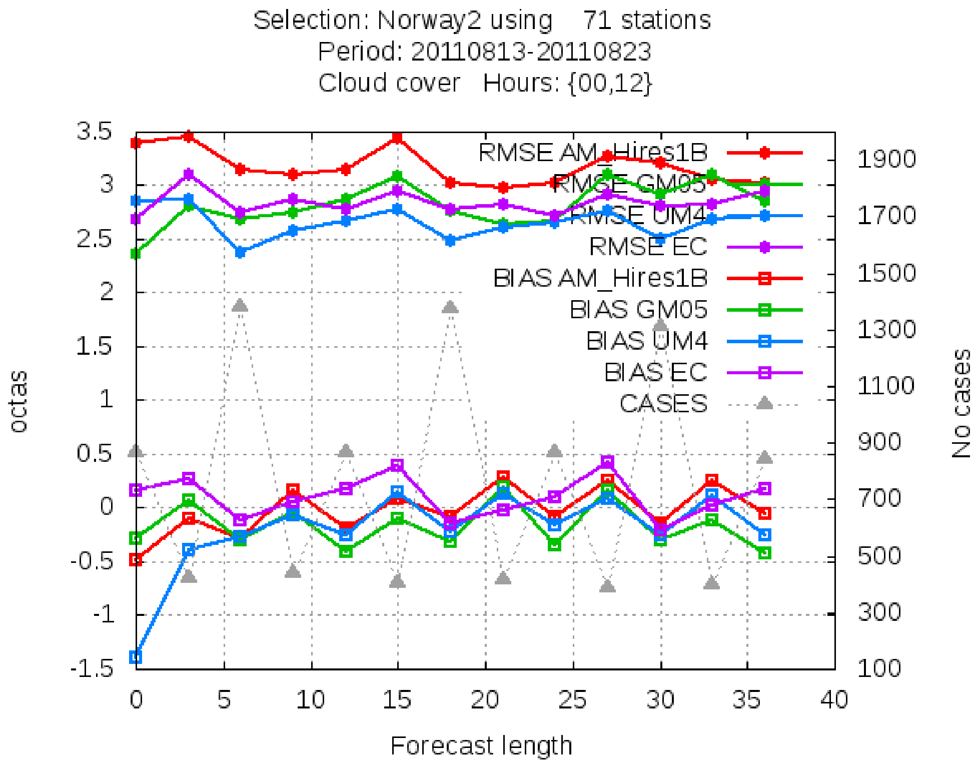


Figure 18: August 2011, RMSE/BIAS Total Cloud Cover averaged over all Norwegian stations.

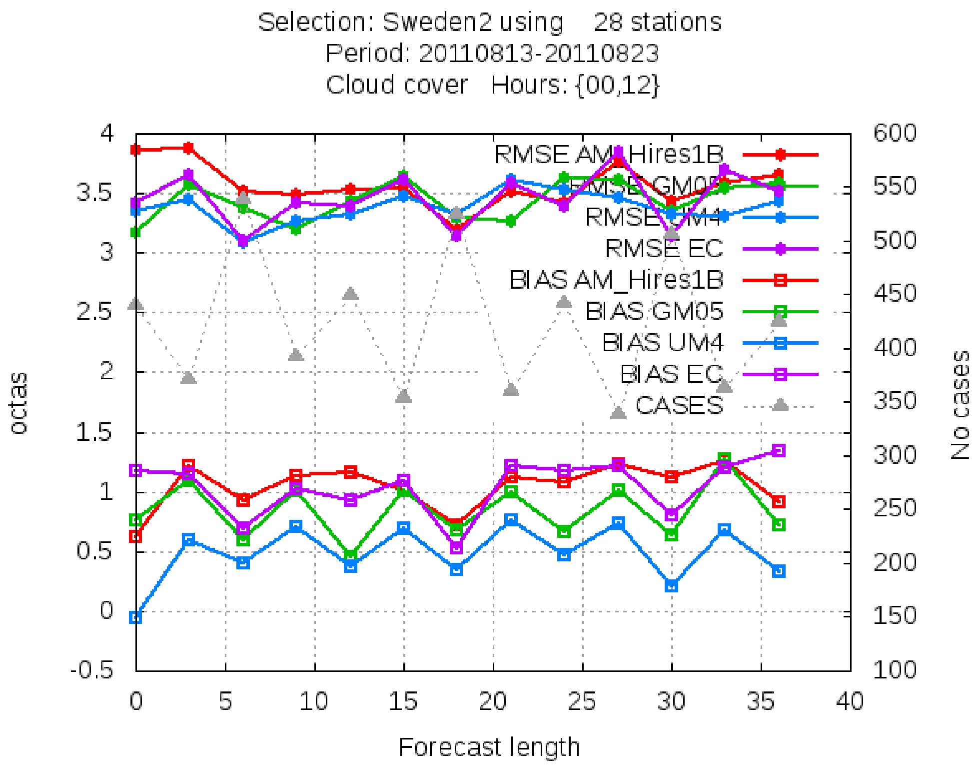


Figure 19: August 2011, RMSE/BIAS for Total Cloud Cover averaged over all Swedish stations.

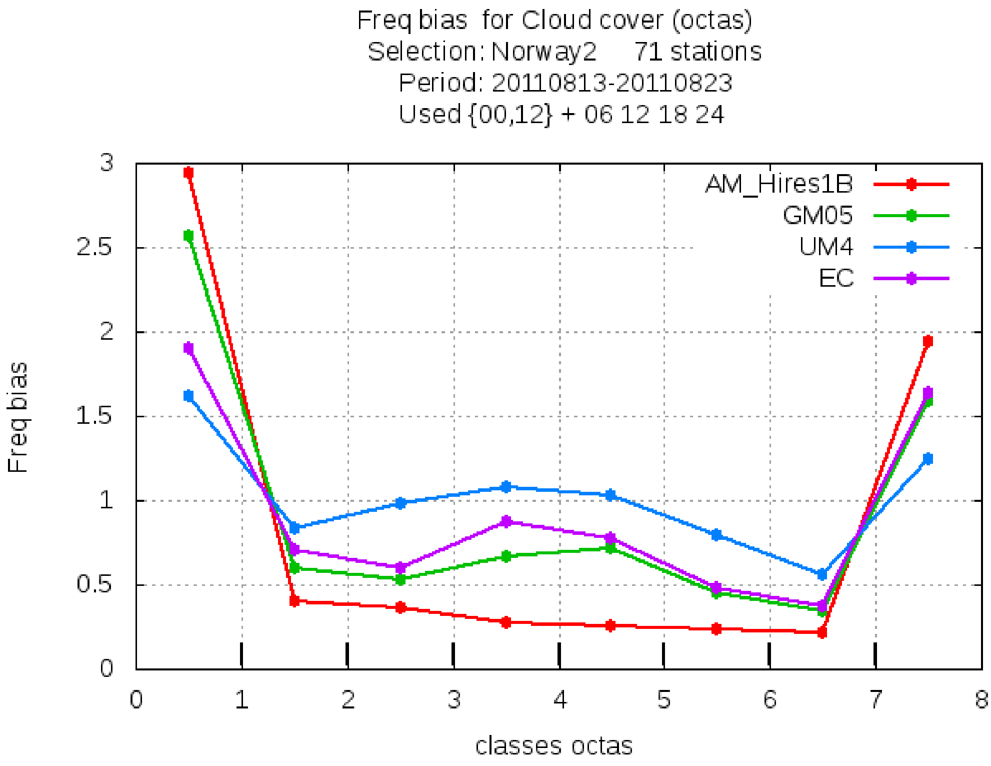


Figure 20: August 2011, Frequency BIAS (FB) for Total Cloud Cover averaged over all Norwegian stations.

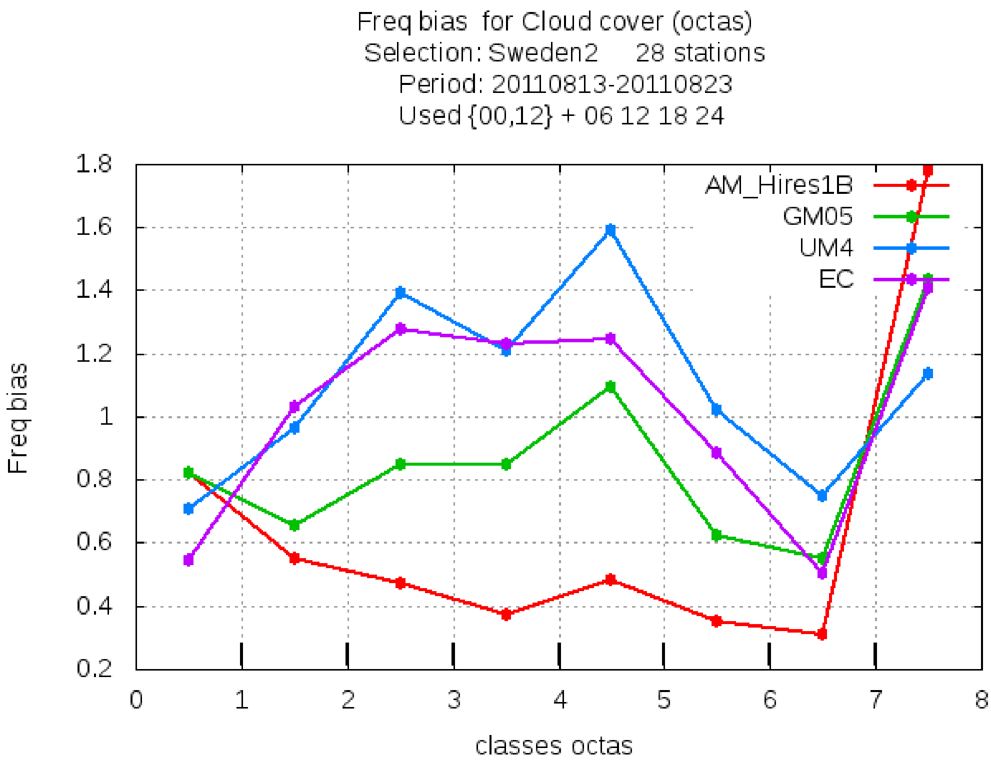


Figure 21: August 2011, Frequency BIAS (FB), Total Cloud Cover averaged over all Swedish stations.

3.5 Verification study of December – January 2012

In this study the rerun of the MetCoOp model HARMONIE AROME (2,5 km) is compared with the rerun of HIRLAM (GM05) and UM (4 km) from *met.no* and the ECMWF forecasts (domain shown in Figure 1).

The verification results following in this chapter are divided into the main weather elements that are most relevant for users and give a quite good overview of the weather situation. The figures show different skill scores for the four models and are shown in order to give a picture of the behaviour of the different models, how well they fit the observations and how the models compare with each other.

3.5.1 Meteorological description of the period

This period, 19.12.2011 – 07.01.2012 was dominated by storms. Three extreme weather warnings were issued by *met.no* (Cato, Dagmar and Emil): Cato (25/12) due to storm surge and strong wind in northern Norway, Dagmar (25/12) due to storm surge and strong wind in southern Norway and in the middle of Sweden and Emil (3–4/1) due to high wind speeds in southern Norway.

In the beginning of the period 19 – 21 December, the main track of the low pressure systems was from the north Atlantic towards central Scandinavia. Between 22 – 27 December, the low pressure system had a more northward track and became very intense. This caused strong and very mild south-westerly winds over most of Scandinavia. During December 28 to 31 the weather calmed down and there was an occasional weak ridge over Scandinavia. This caused the temperature to drop slightly. During January 1–5 there was a new period with very intensive low pressure systems moving towards northeast, but the tracks went a little more southward than during the Christmas period. Again it was very mild and windy in the southern parts of Sweden and Norway, but in the north there were colder easterly winds and snowfall.

3.5.2 Verification results of mean sea level pressure (MSLP)

In terms of RMSE ECMWF is better than the other models, with increasing errors in AROME, GM05 and UM4. The errors in all models increase with forecast length (Figure 22). An increase in BIAS is found for all models as a function of lead time (ECMWF has the smallest BIAS). In general the superiority of ECMWF is more pronounced in Sweden than in Norway.

AROME, and GM05 show (unlike ECMWF and UM4) a clear peak in RMSE 26th and 27th of December when the extreme weather Dagmar and Cato hit Norway/Sweden. The situations consist of strong gradients in the pressure fields leading to larger errors. A somewhat similar case, but not as pronounced is 4th of January when the extreme weather Emil hit southern Norway, Sweden and Denmark. The limited area models have problems meeting the quality of MSLP in the ECMWF model in situations with severe synoptic events. This should be investigated further, a first test being to update the limited area models with lateral boundary conditions every hour and not every third hour as employed in this study.

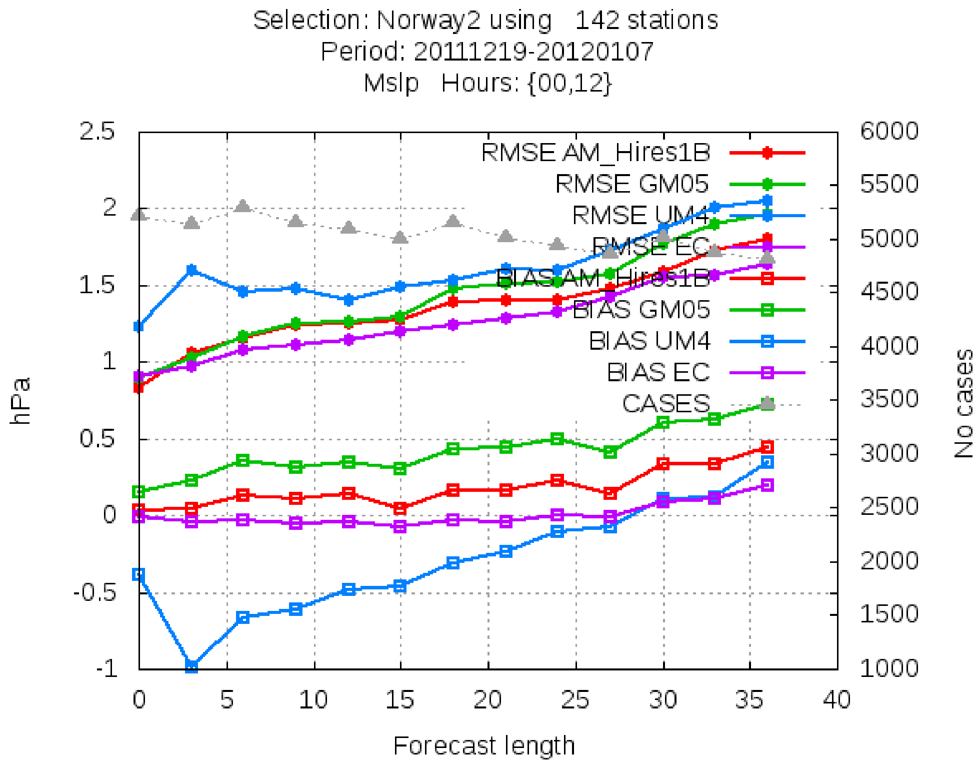


Figure 22: December/January, MSLP, averaged over all Norwegian stations.

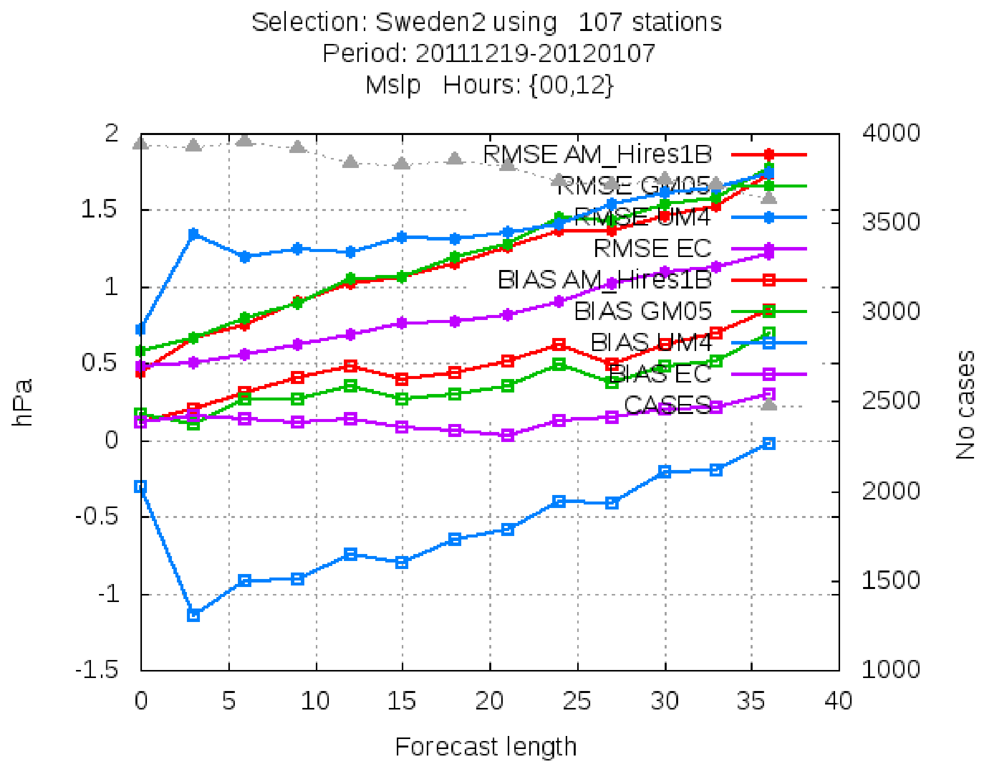


Figure 23: December/January, MSLP, averaged over all Swedish stations.

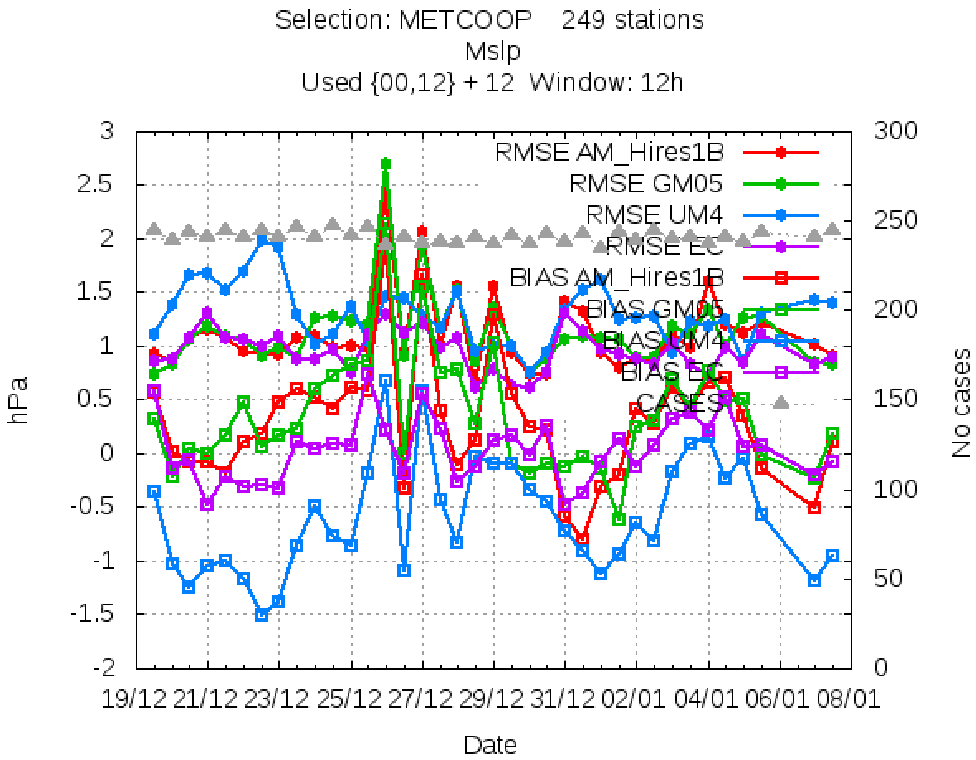


Figure 24: December/January, time series of RMSE of MSLP, averaged over all stations.

3.5.3 Verification results of 10 m wind direction

Averaged over all stations there are small differences between the RMSE and BIAS of the different models wind direction. However, AROME shows slightly smaller RMSE than the other models (more pronounced in Norway than Sweden, and more pronounced inland). The RMSE is much smaller for all models at the coast. ECMWF shows slightly better results than the fine scale models at the coast in this period.

3.5.4 Verification results of 10 m wind speed

Averaged over all stations GM05, AROME and ECMWF show a small overestimation, while UM4 slightly underestimates the wind speed. In terms of RMSE, GM05 is slightly better than the other models and AROME has slightly larger RMSE than the other models. However, there are clear differences between Sweden and Norway. In Norway, AROME shows the largest (positive) BIAS and largest RMSE (Figure 25), while in Sweden AROME shows a small (positive) BIAS and has the smallest RMSE (Figure 26). In general, AROME is better in the mountains than the other models. It has smaller RMSE and no systematic errors while the other models heavily underestimate the wind speed. At coastal stations all models are very similar in terms of systematic errors and RMSE, except for slightly larger RMSE for UM4. For the stations close to the largest cities AROME shows not so good verification results as the other models. In general AROME has a positive systematic error for all regions, with the exception of stations between 200 and 600m where an underestimation of approximately -0.5 m/s is found.

Looking at the Frequency BIAS (FB) (Figure 27) AROME shows better results than the other models for wind speeds above ~12 m/s. While GM05, UM4 and ECMWF underestimate the frequency of occurrences above 10 m/s AROME captures the observed frequency quite well up to almost 20 m/s.

Investigating the quality of wind speed with ETS (Figure 28) reveals that AROME also for this skill score shows better results than the other models for wind speeds above 10 m/s. For AROME higher hit rate compensates for a higher false alarm ratio. UM4 is slightly better than GM05 while ECMWF shows the poorest result of the four models when using the ETS skill score. AROME is superior to the other models for high wind speeds and for high elevated stations (Figure 29) and also in the northern part of the domain (not shown). Furthermore, compared to the other models AROME shows lower ETS over the largest cities (not shown) than the other models (due to higher false alarm ratio).

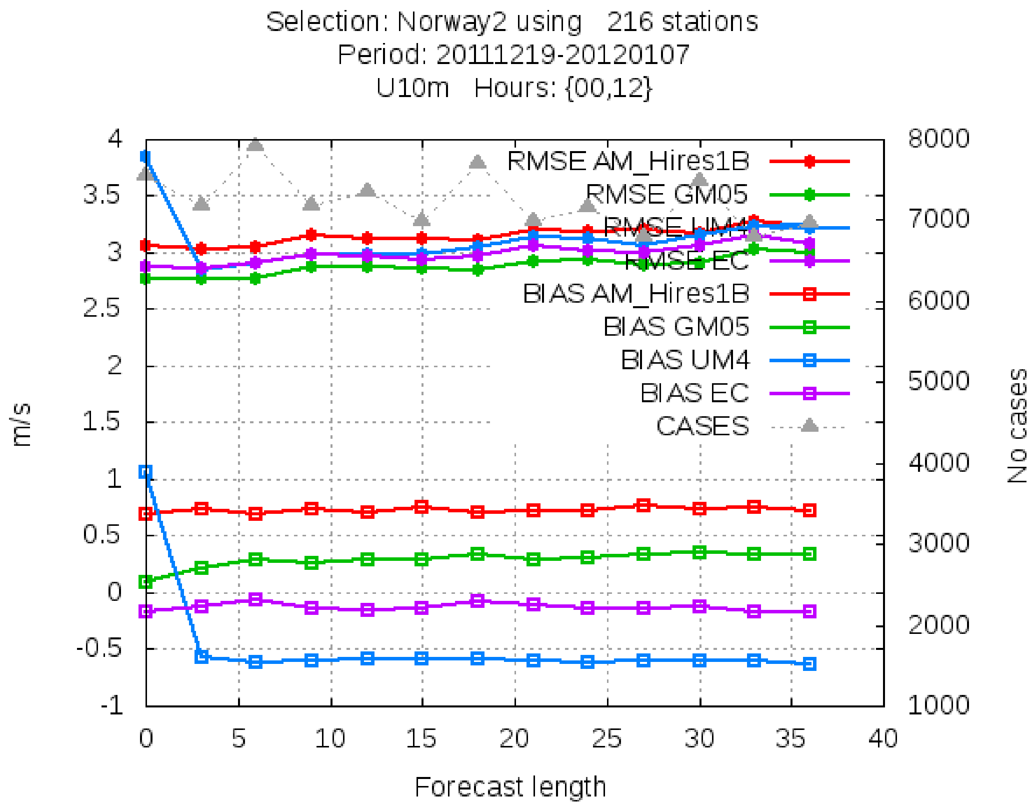


Figure 25: December/January, 10 m wind, averaged over all Norwegian stations.

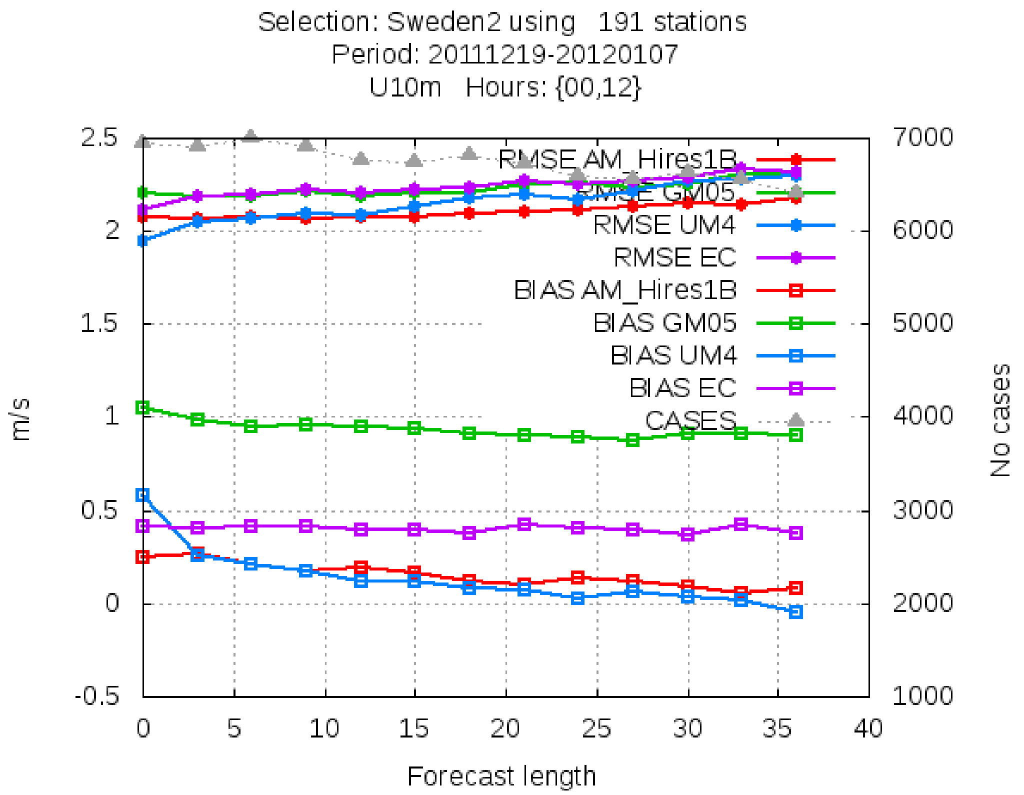


Figure 26: December/January, 10 m wind, averaged over all Swedish stations.

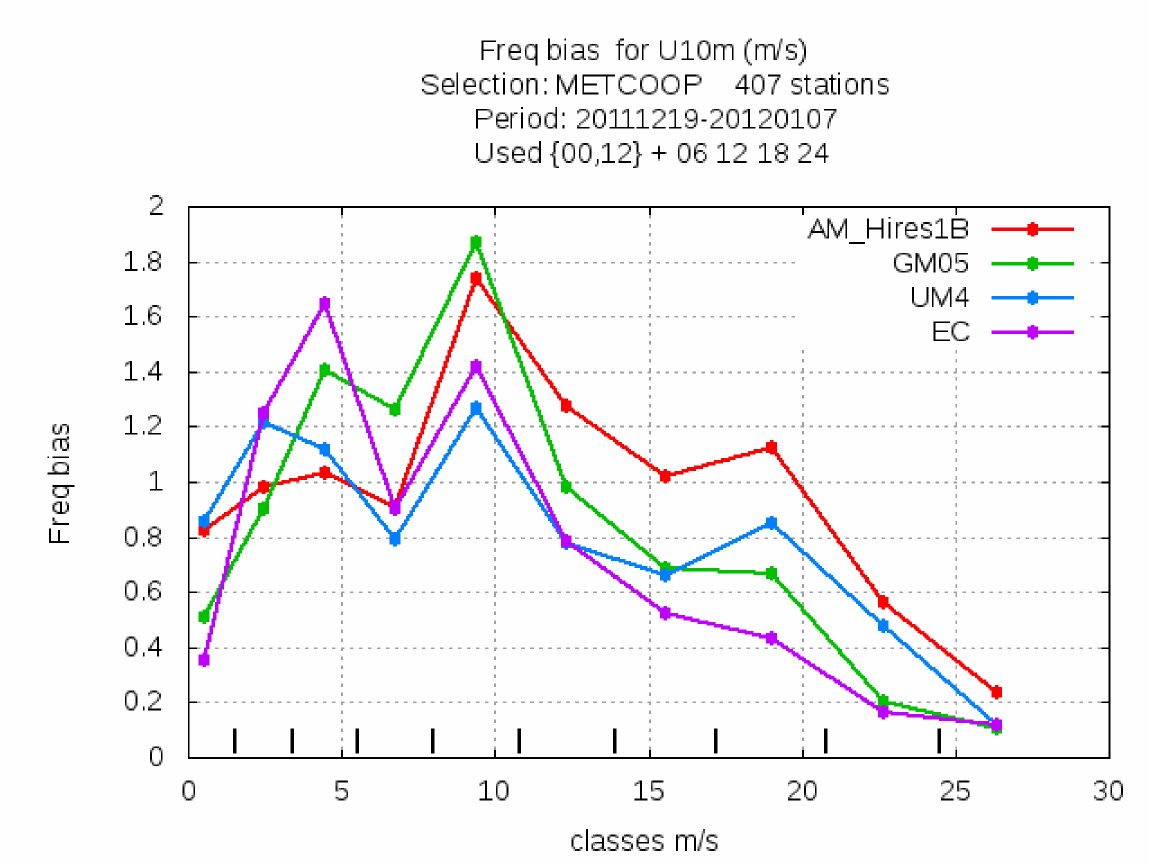


Figure 27: December/January, Frequency BIAS (FB) for10 m wind for all stations.

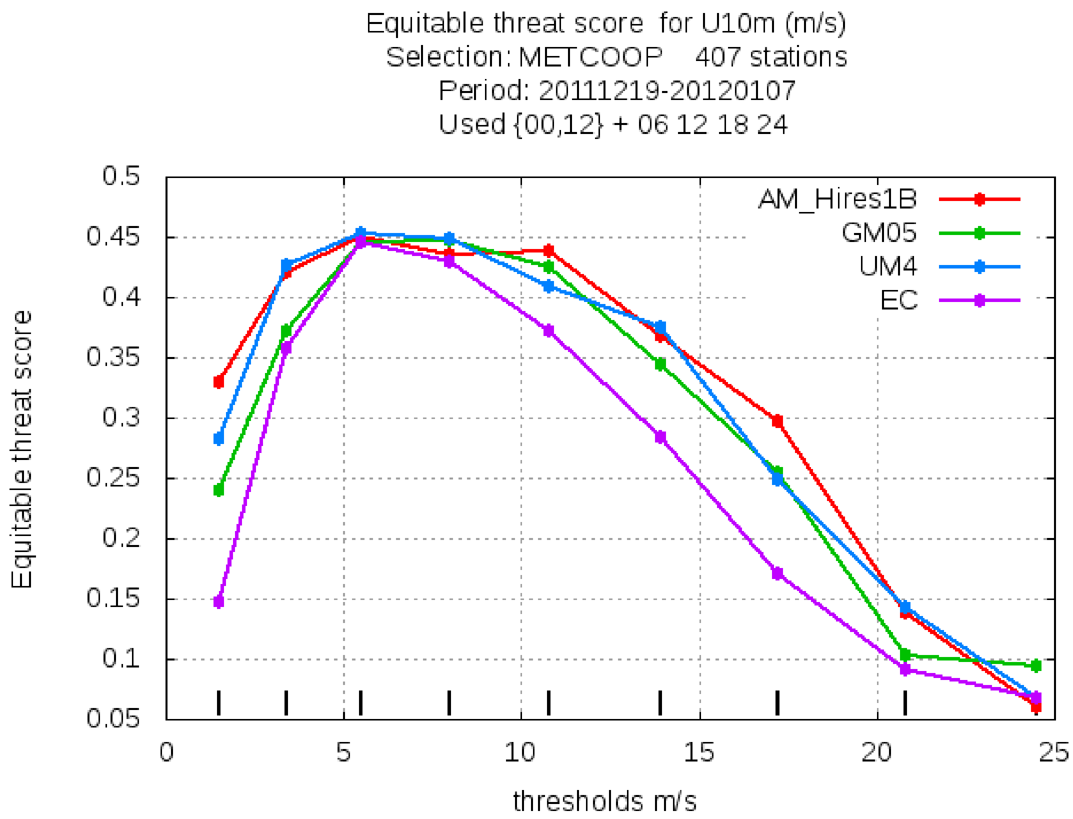


Figure 28: December/January, ETS, averaged over all stations.

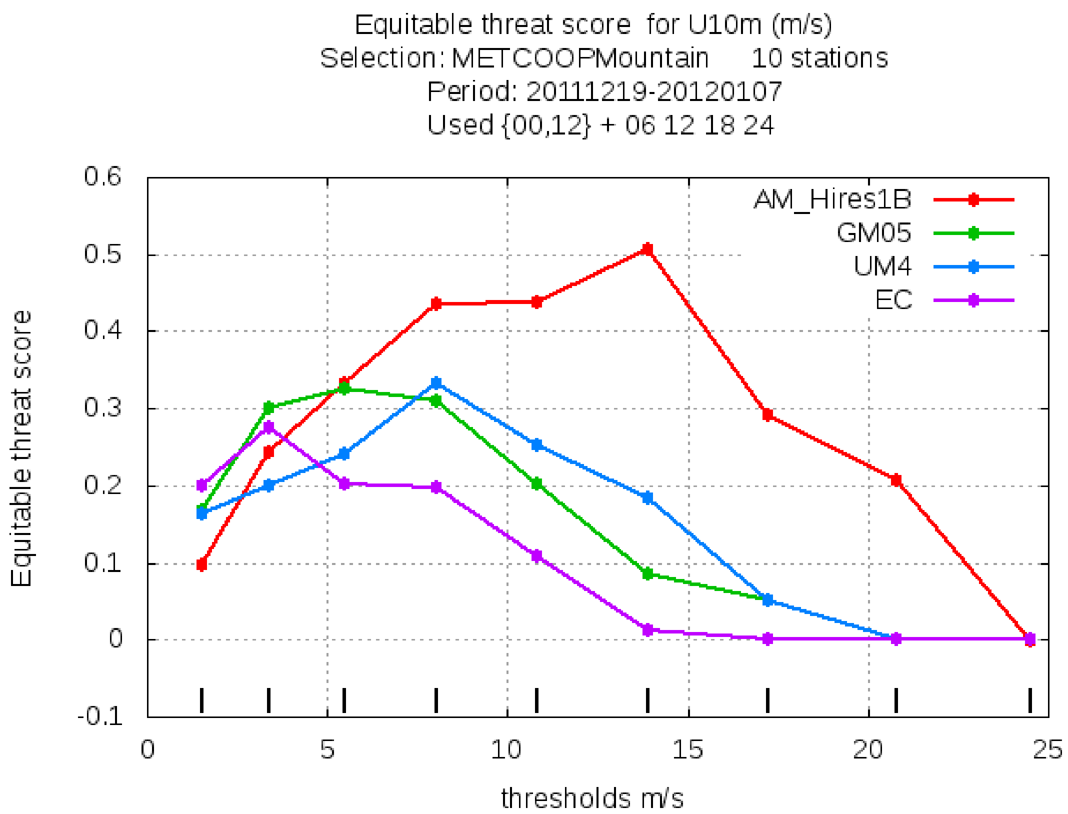


Figure 29: December/January, ETS, averaged over mountain stations.

For the storm (25.12.11) called *Dagmar* some snapshots of the wind speed forecasts from the different models are shown (Figure 30– Figure 33). The yellow areas are areas where the models forecast more than 40 knots. As seen, all models show high wind speeds over open sea, while they differ in the details at the coast. Except for AROME all models heavily underestimate the wind strength in the mountains, especially ECMWF which only in small regions exceeds 30 knots. AROME shows large areas with wind speeds above 40 knots and also areas of more than 50 knots, which are in good agreement with the available observations.

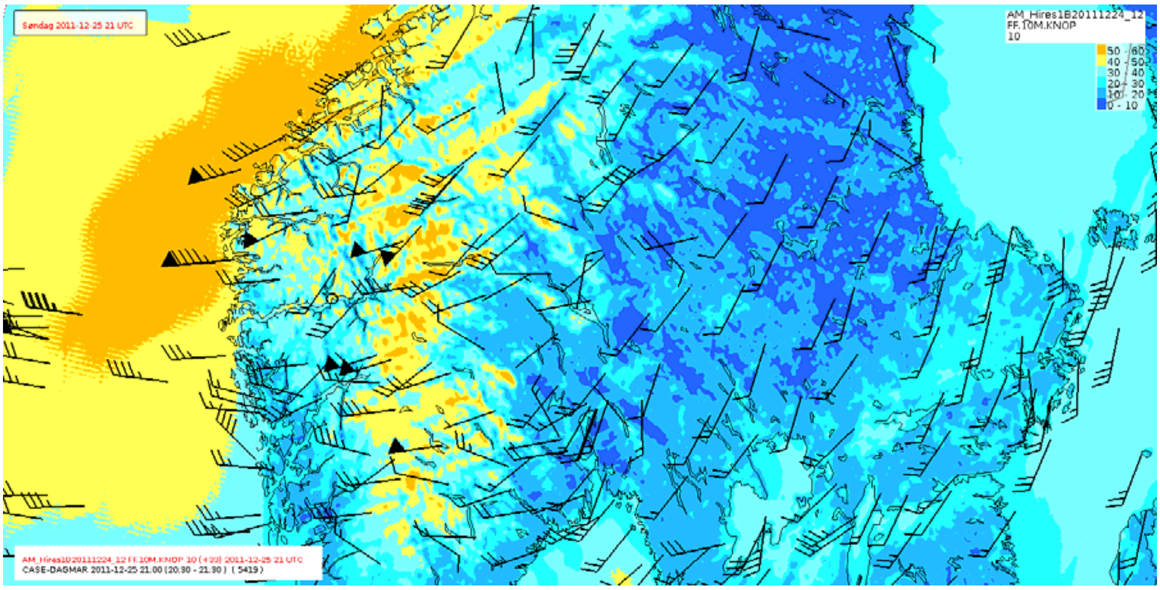


Figure 30: Snapshot of +33hr AROME wind strength (knots) forecast and observations valid 25/12 2011 -21UTC

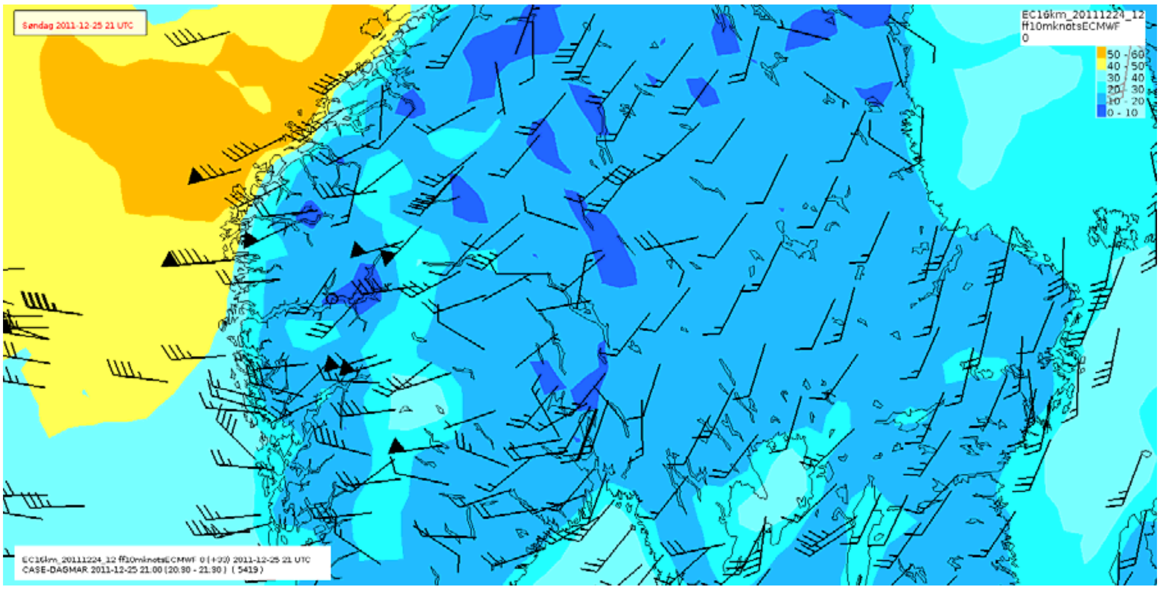


Figure 31: Snapshot of +33hr ECMWF wind strength (knots) forecast and observations valid 25/12 2011 -21UTC

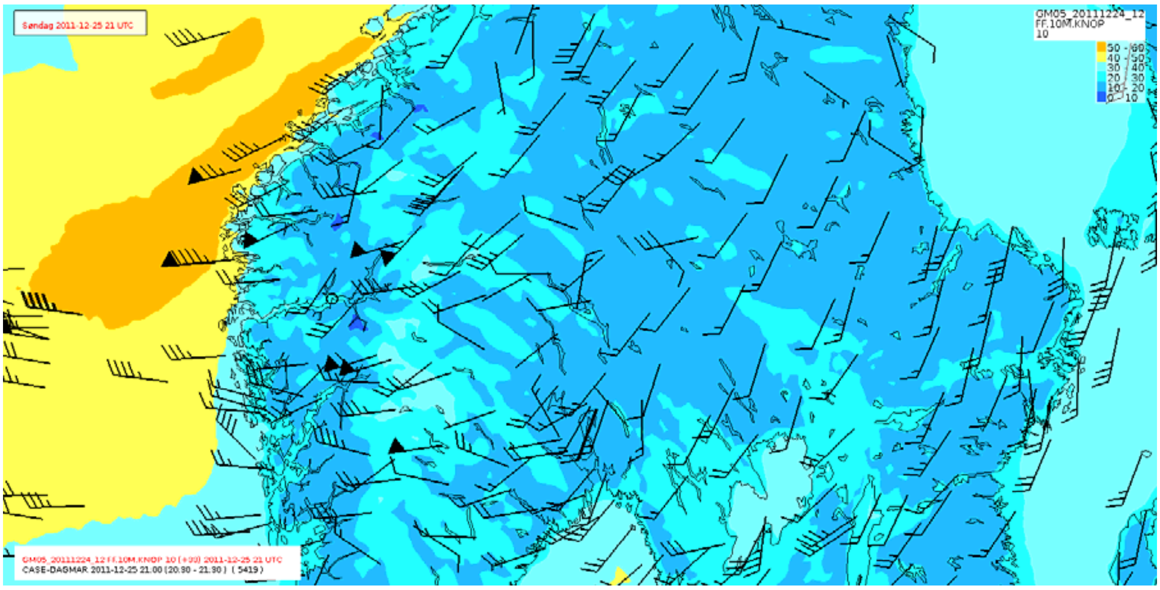


Figure 32: Snapshot of +33hr GM05 wind strength (knots) forecast and observations valid 25/12 2011 -21UTC

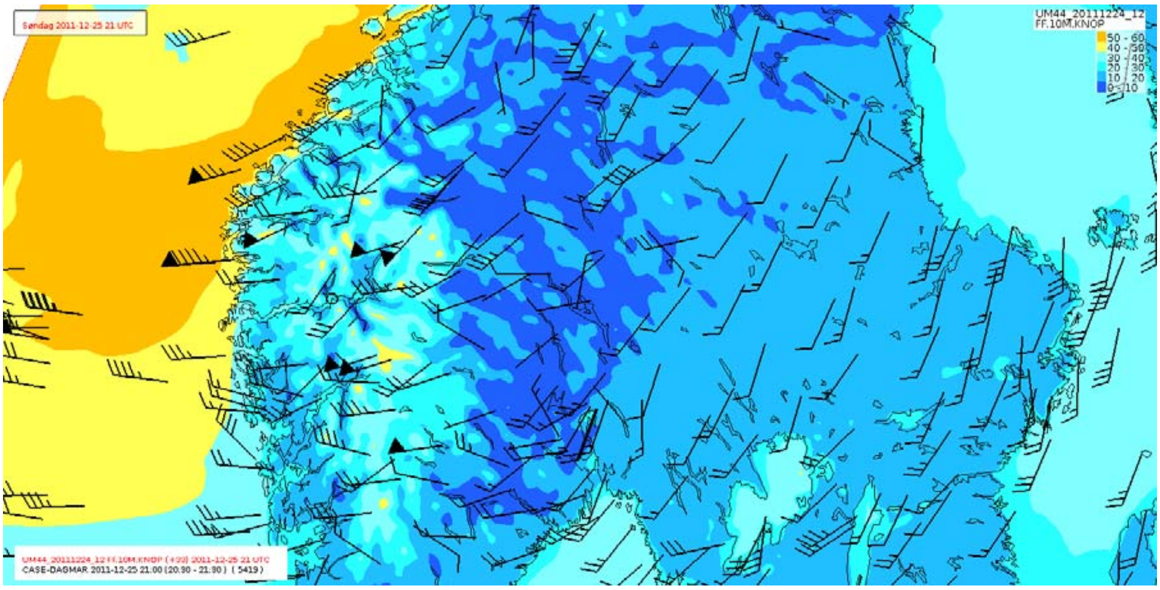


Figure 33: Snapshot of +33hr UM4 wind strength (knots) forecast and observations valid 25/12 2011 -21UTC

Figure 34 – Figure 37 show the four models forecasts of 10 m wind from a situation containing mountain waves at the northern coast of Norway. These Figures illustrate the fine scale models (and in particular AROME's) capability of modelling high wind speeds and a more detailed forecast in complex terrain. However the models forecasts in situations with mountain waves should be further investigated.

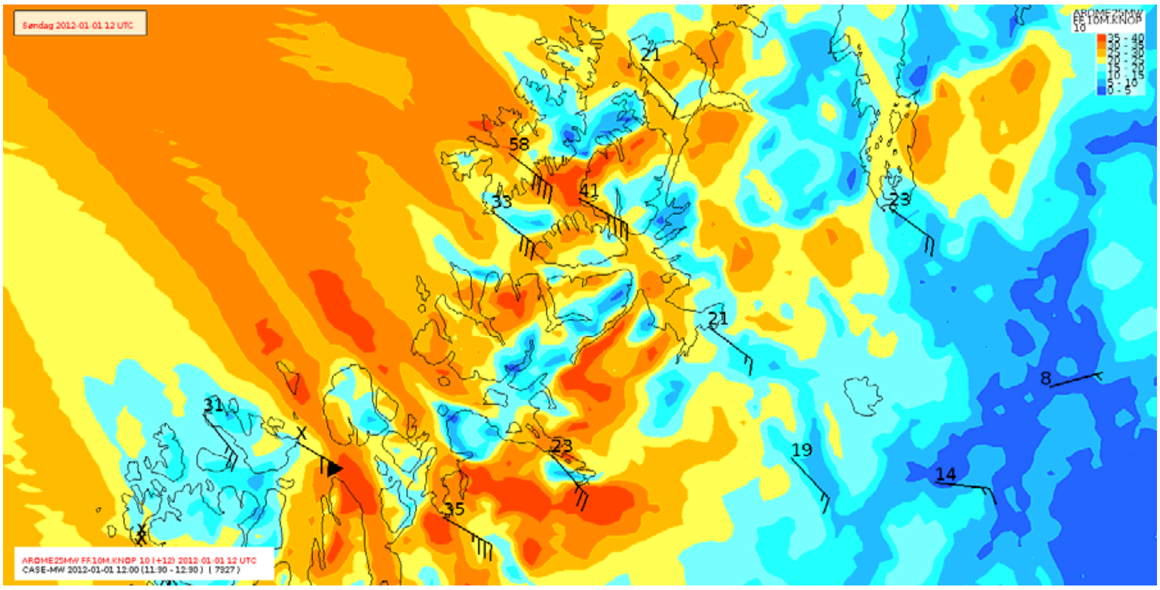


Figure 34: Snapshot of +12hr AROME 10 m wind strength (knots) forecast and observations (mean wind and max wind last hour) valid 01/01/12 12UTC for a case with mountain waves at the Northern Norwegian Coast

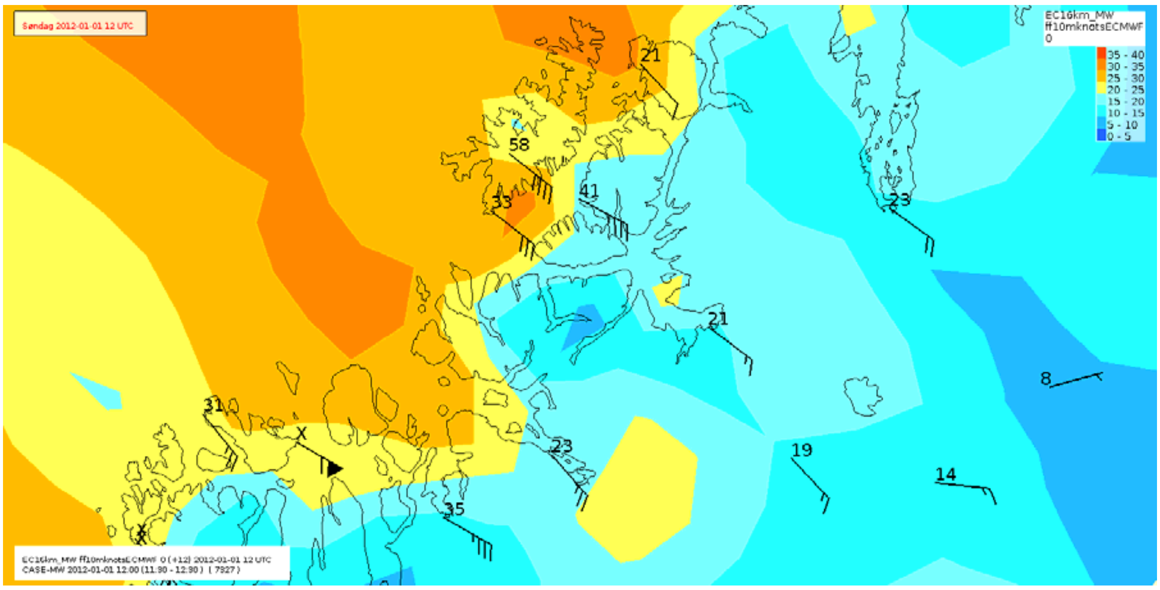


Figure 35: Snapshot of +12hr ECMWF 10 m wind strength (knots) forecast and observations (mean wind and max wind last hour) valid 01/01/12 12UTC for a case with mountain waves at the Northern Norwegian Coast.

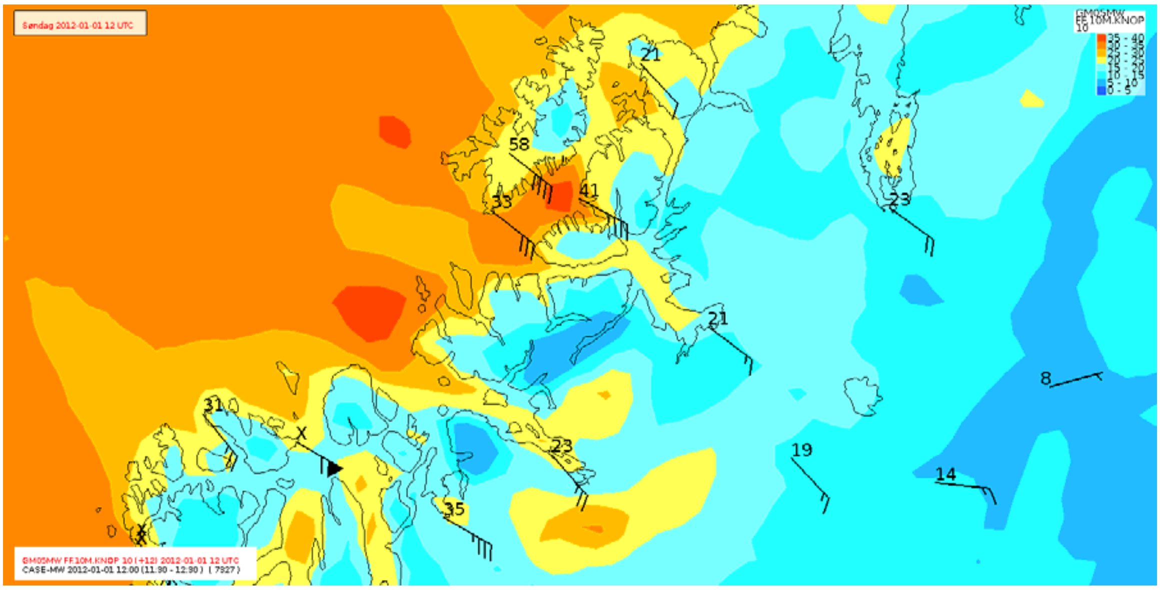


Figure 36: Snapshot of +12hr GM05 10 m wind strength (knots) forecast and observations (mean wind and max wind last hour) valid 01/01/12 12UTC for a case with mountain waves at the Northern Norwegian Coast

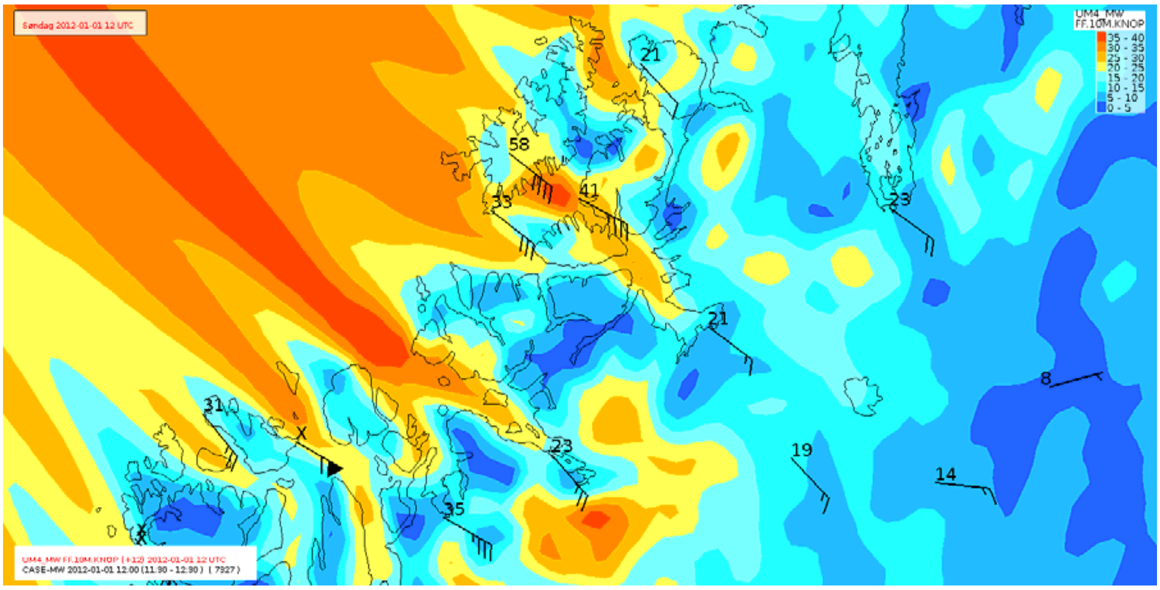


Figure 37: Snapshot of +12hr UM4 10 m wind strength (knots) forecast and observations (mean wind and max wind last hour) valid 01/01/12 12UTC for a case with mountain waves at the Northern Norwegian Coast

3.5.5 Verification results of near surface temperature (T2m)

Averaged over all stations, GM05 (followed by AROME) is best for all lead times in terms of systematic errors and RMSE. AROME shows a small positive BIAS (0.25°C), while ECMWF shows a slowly increasing negative BIAS (>1°C at +36h) and UM4 shows a fast increasing negative BIAS (>2°C after +36h).

In Norway (Figure 38), AROME shows a clear positive BIAS (+1°C), while AROME in Sweden (Figure 39) shows a small negative BIAS (<0.5°C). In the mountains there is a clear negative BIAS for all models, most probably associated with inversion episodes in the winter. However, since this period did not contain any long lasting periods of inversion this should be investigated further. At the coast a small, but consistent positive BIAS of 0.4°C is found for AROME.

A very clear feature seen in all scatter plots for all models (but most pronounced for AROME) is the overestimation of the coldest observed temperatures (Figure 40). The models are not able to forecast the coldest observed temperatures. The station maps show many inland stations in Norway with a warm BIAS, while a slightly cold BIAS is seen in the most southern part of Sweden.

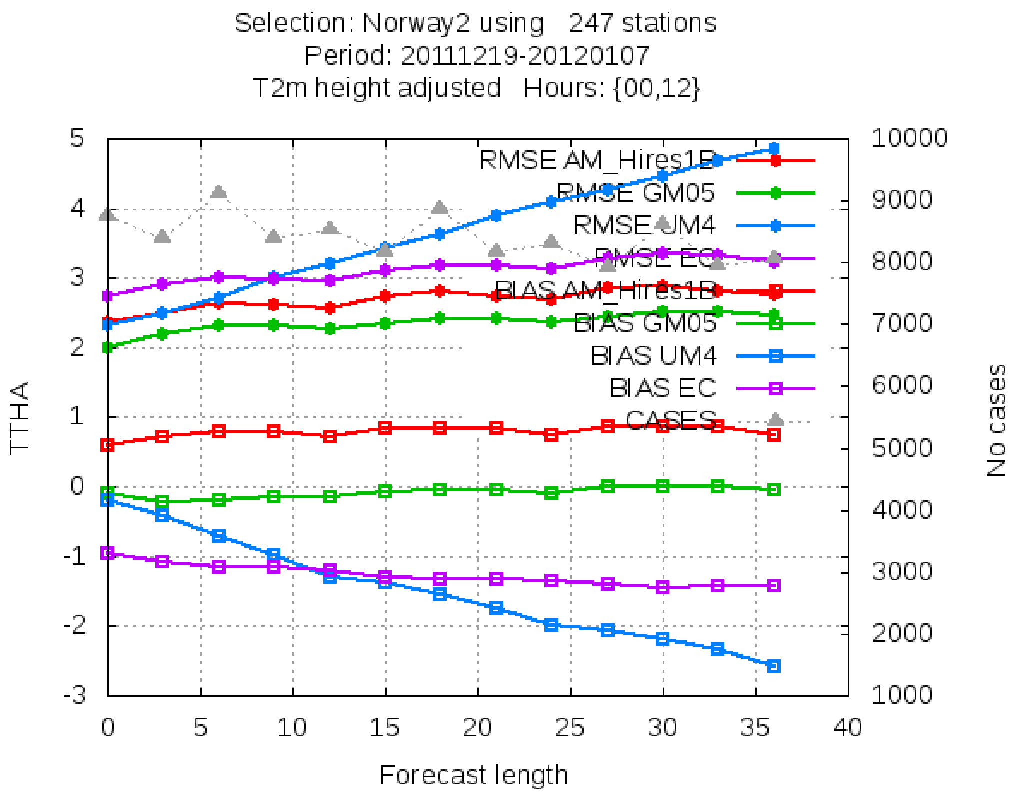


Figure 38: December/January, T2m, averaged over all Norwegian stations.

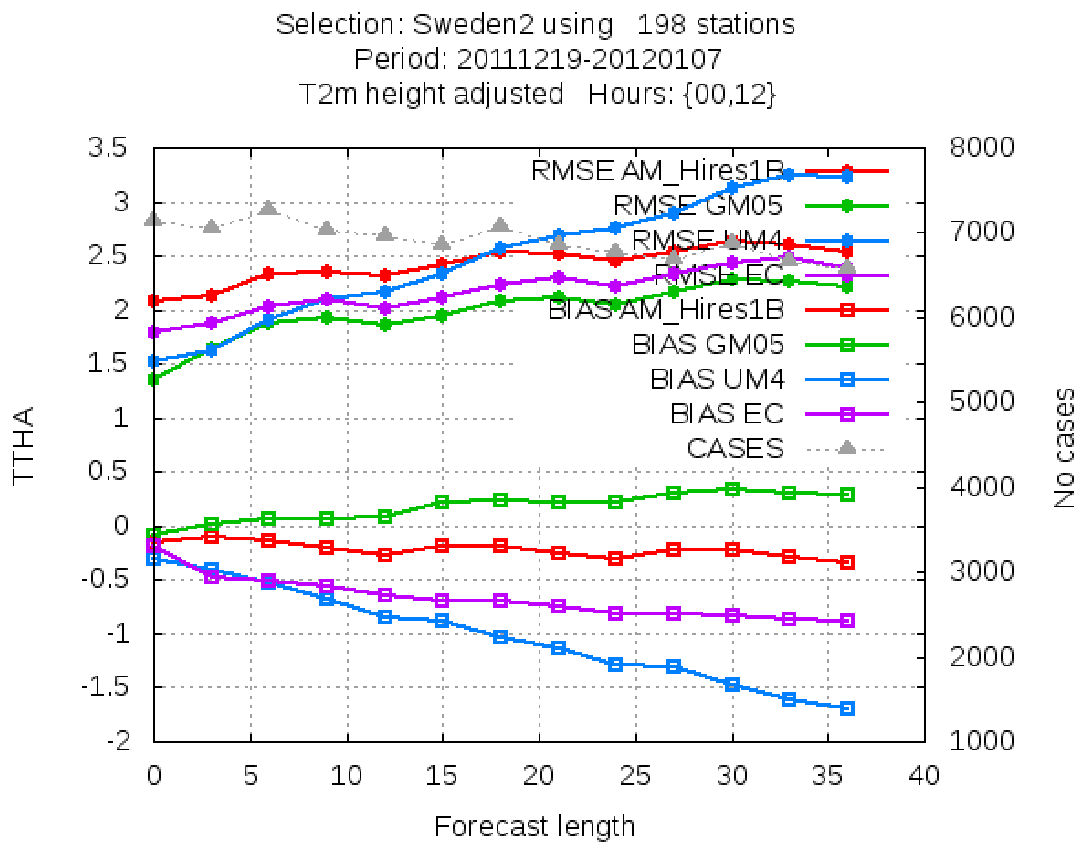
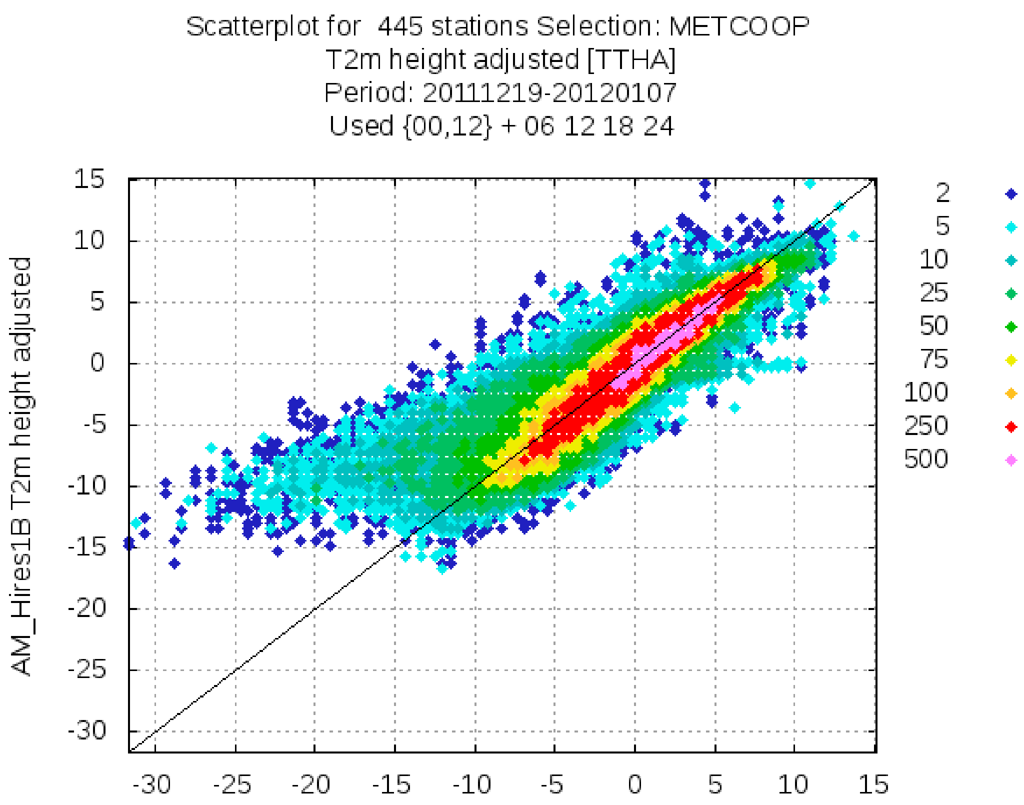
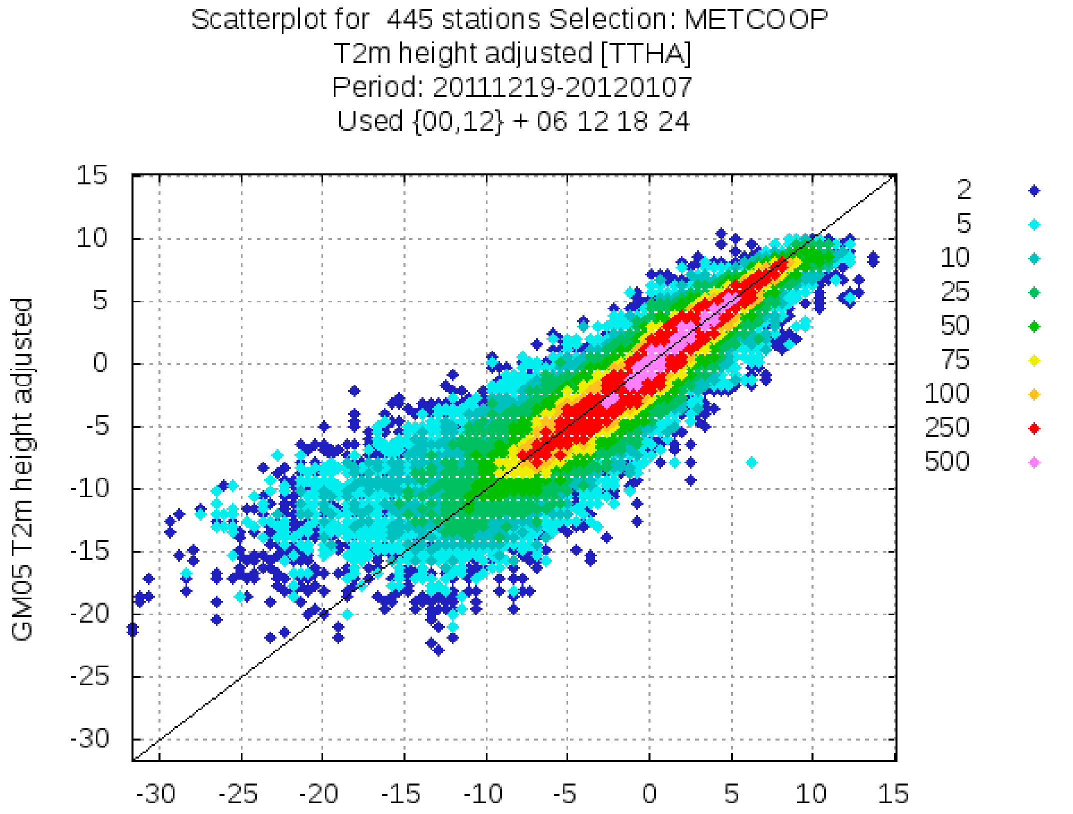
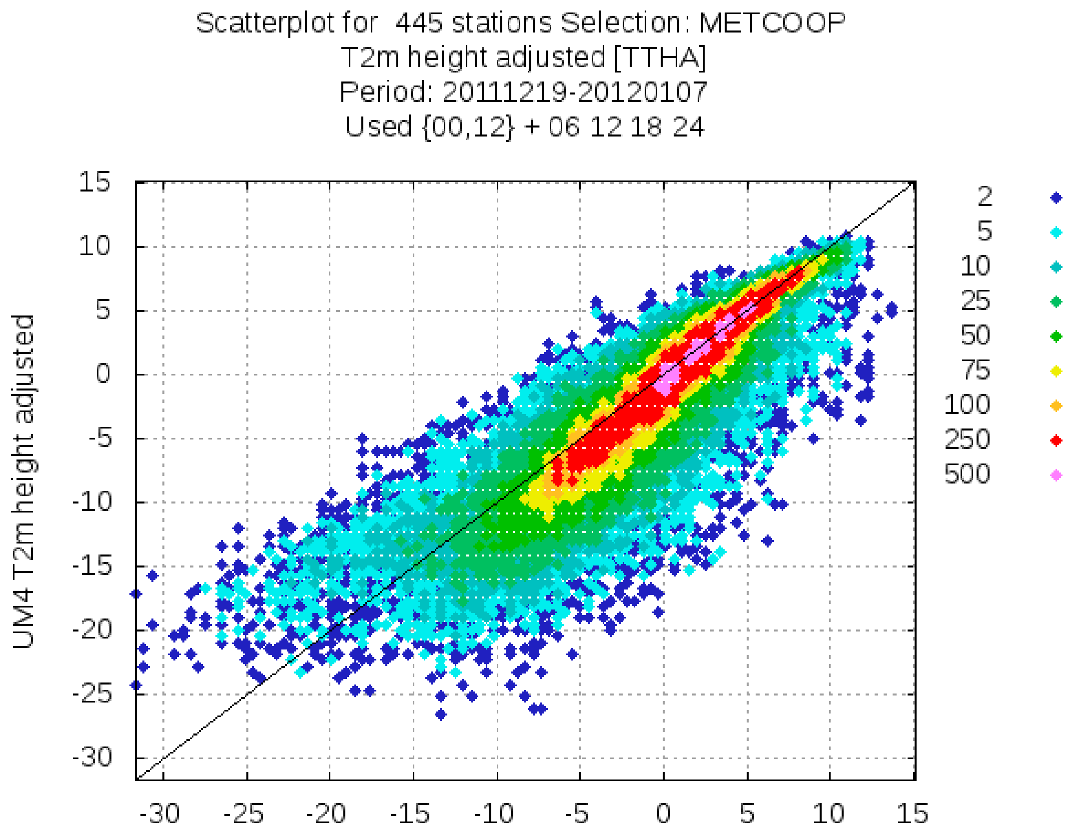


Figure 39: December/January, T2m, averaged over all Swedish stations.




Scatterplot for 445 stations Selection: METCOOP
T2m height adjusted [TTHA]
Period: 20111219-20120107
Used {00,12} + 06 12 18 24
GM05 T2m height adjusted
2
5
10
25
50
75
100
250
500


Scatterplot for 445 stations Selection: METCOOP
T2m height adjusted [TTHA]
Period: 20111219-20120107
Used {00,12} + 06 12 18 24
UM4 T2m height adjusted
2
5
10
25
50
75
100
250
500

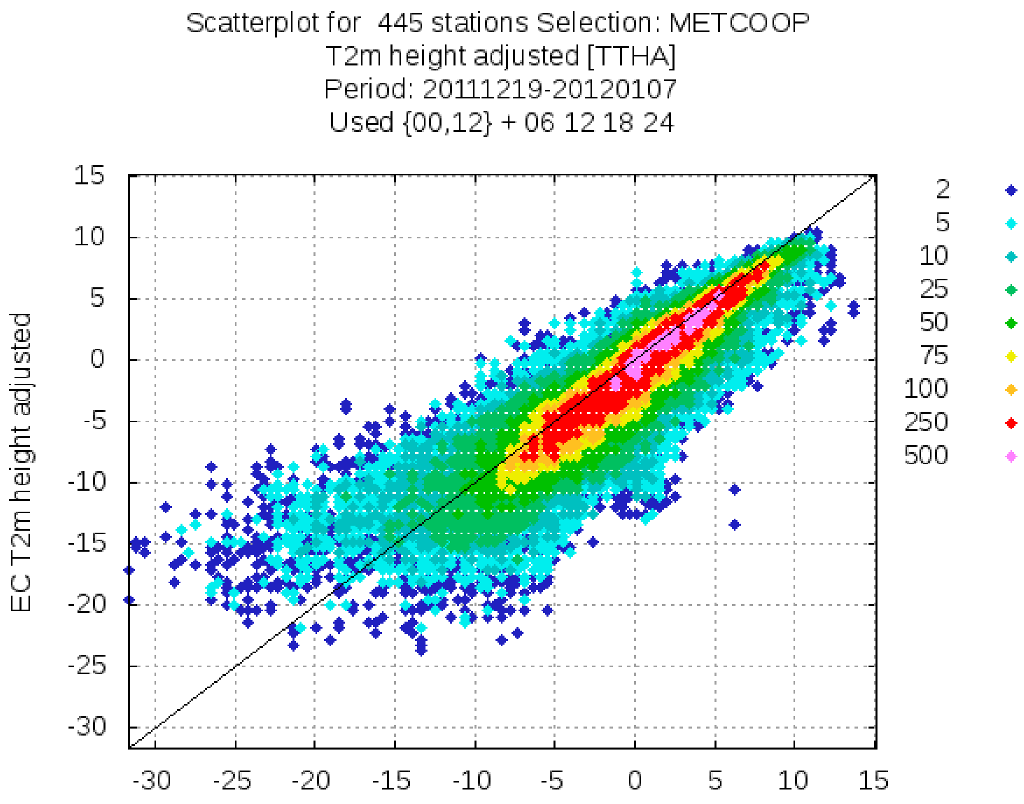


Figure 40: December/January, T2m scatter plots, averaged over all stations with AROME, GM05, UM4 and ECMWF.

3.5.6 Verification results of precipitation (12h)

All models overestimate slightly the precipitation amounts averaged over all stations (AROME 0.1 mm/12h, ECMWF 0.3 mm/12h, GM05 0.4 mm/12h and UM4 0.6 mm/12h). This is a desired feature since measured precipitation can be looked at as the minimum of actual precipitation due to observation errors under snow and windy conditions. It can be discussed whether AROME should have an even larger overestimation. A very heavy overestimation of the precipitation amounts is seen at high elevations (between 1 mm/12h in GM05 and almost 3 mm/12h AROME). The interpretation of these latter results is difficult due to few stations and the large uncertainty associated with the observations.

In terms of RMSE the models are on average ranked as follows: AROME, ECMWF, UM4 and GM05. This ranking is also valid for the Norwegian stations (Figure 41), while in Sweden (Figure 42), ECMWF is superior to the three other models with respect to RMSE. RMSE may be smaller only due to a lower variation of the forecast. This is probably the case for ECMWF over Sweden. Over mountains which are more common in Norway, AROME places details of the precipitation field correctly since AROME has a more detailed topography than ECMWF. But over flat terrain (Sweden) details of the precipitation may be more random due to convection. This may cause a larger RMSE, relative to ECWMF over Sweden compared to Norway.

The frequency of different precipitation amounts are reasonably well captured by AROME with Frequency BIAS (FB) between 1 (small amounts) and 1.2 (large amounts). The other models show higher FBes with ECMWF having the highest FB (1.25 (small amounts) and 1.35 (intermediate amounts)).

AROME shows better ETS (Figure 43) for small precipitation amounts (< 1mm/24h), and almost as good as UM4 for light and heavy precipitation. AROME and UM4 are better than ECMWF and are superior compared to GM05 with regard to the ETS.

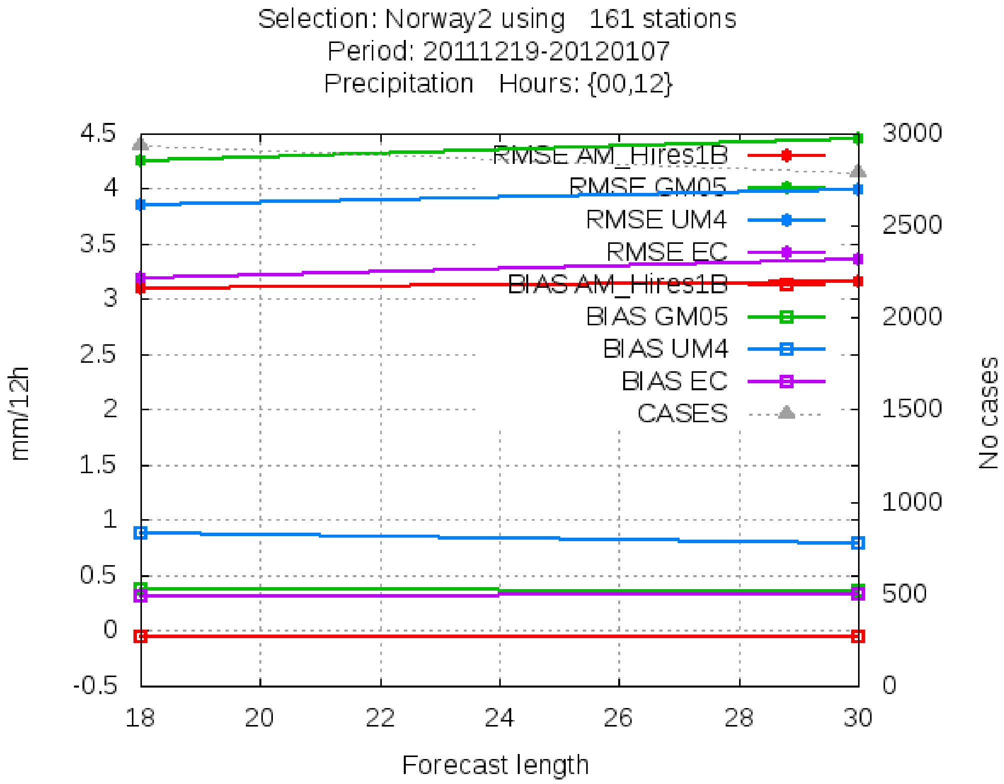


Figure 41: December/January, RMSE/BIAS, 12hr precipitation averaged over all Norwegian stations.

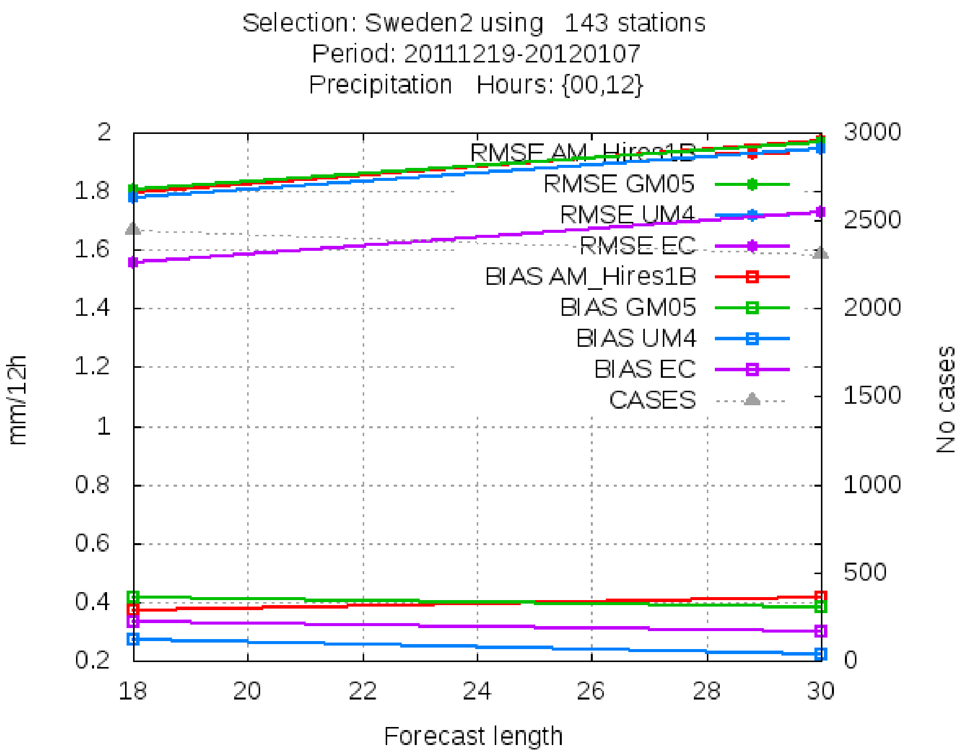
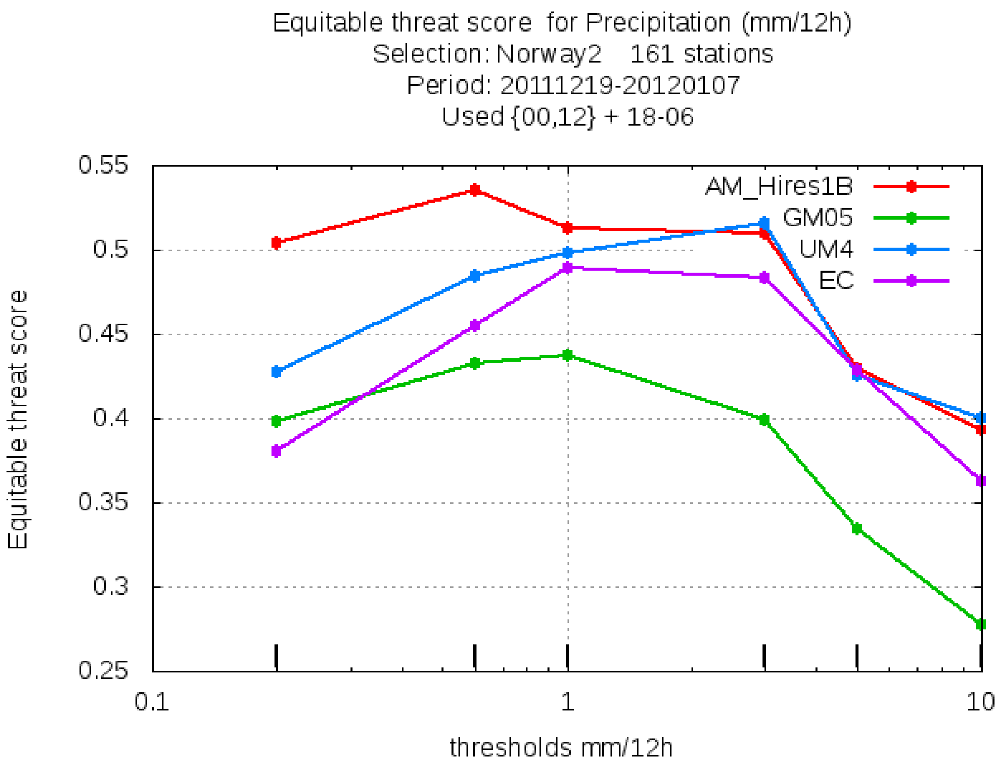


Figure 42: December/January, RMSE/BIAS, 12hr precipitation averaged over all Swedish stations.



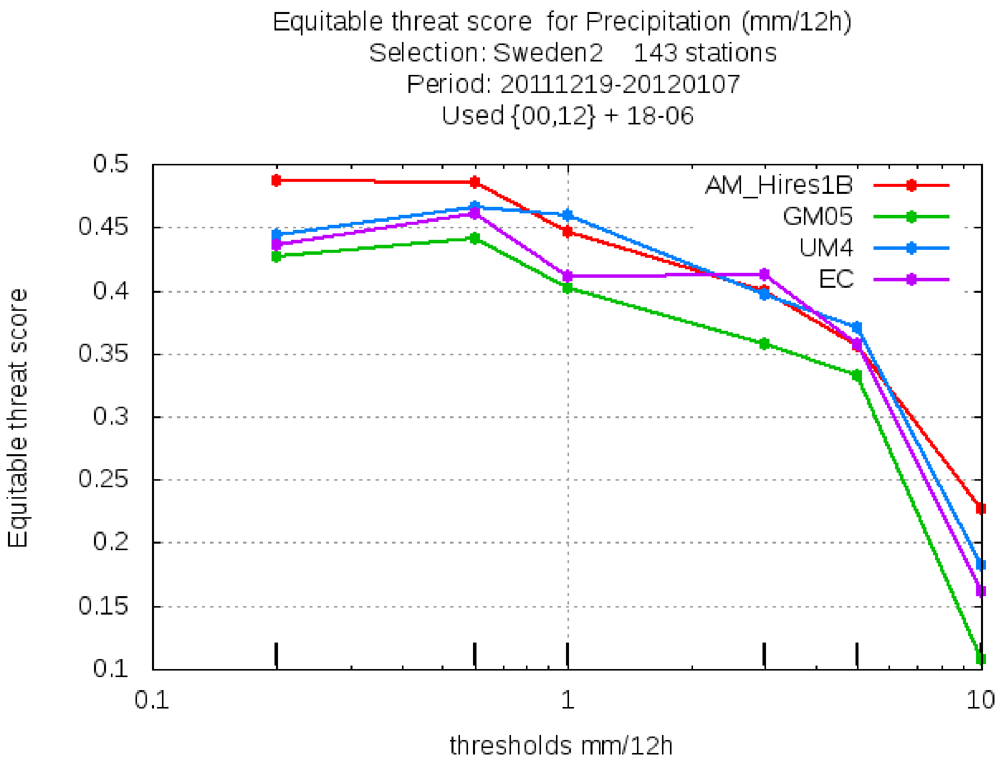


Figure 43: December/January, ETS, 12hr precipitation averaged over all Norwegian (top) and Swedish (bottom) stations.

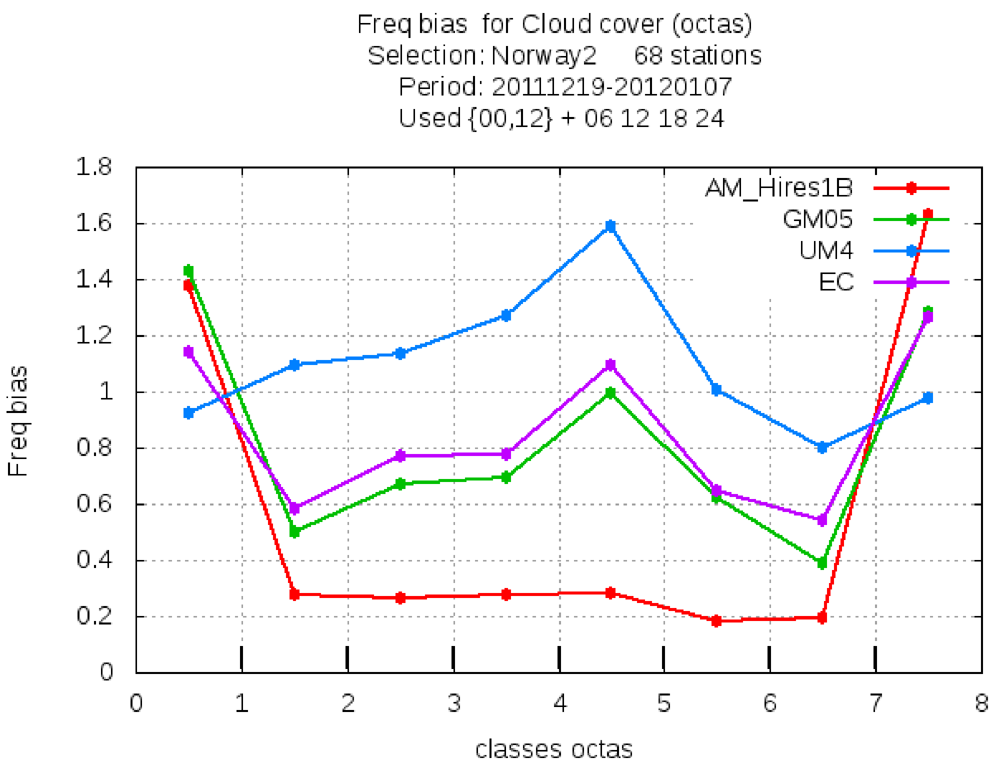
3.5.7 Verification results of total cloud cover (TCC)

Averaged over all stations UM4 is best in terms of a negligible BIAS and the smallest RMSE followed by GM05 and ECMWF. AROME shows larger RMSE than the other models and an overestimation of 0.5 to 1.0 octas. These findings are valid both for Norwegian and Swedish stations.

From the Frequency BIAS (FB) plot we see that GM05 overestimates the occurrences of no clouds with up to 20%, while ECMWF and UM4 underestimate with approximately the same. AROME shows only a minor overestimation. For total cloud clover UM4 shows a minor underestimation while it is heavily overestimated in GM05 and ECMWF with 20% and with more than 50% in AROME. On the intermediate cloud covers (1/8 –7/8) UM4 overestimates the occurrences, GM05 and ECMWF slightly underestimate, while AROME heavily underestimates the frequency of occurrences.

The picture is somewhat different when comparing Norway and Sweden (Figure 44). In Norway AROME heavily overestimates both clear-sky and cloudy conditions and UM4 seems to be in best agreement with the observations. In Sweden, all models underestimate the frequency of clear-sky conditions. AROME is overestimating the cloudy conditions with 35% while the three other models capture this feature quite well. On average, GM05 is probably in best agreement with the observations.

In terms of ETS GM05 is best for small cloud amounts while UM4 is better in more cloudy conditions. For cloudy conditions AROME scores very poorly. In contrast to the Frequency BIAS (FB) these results are valid in both Sweden and Norway. However, the superiority of UM4 and GM05 are more pronounced in Sweden than in Norway.



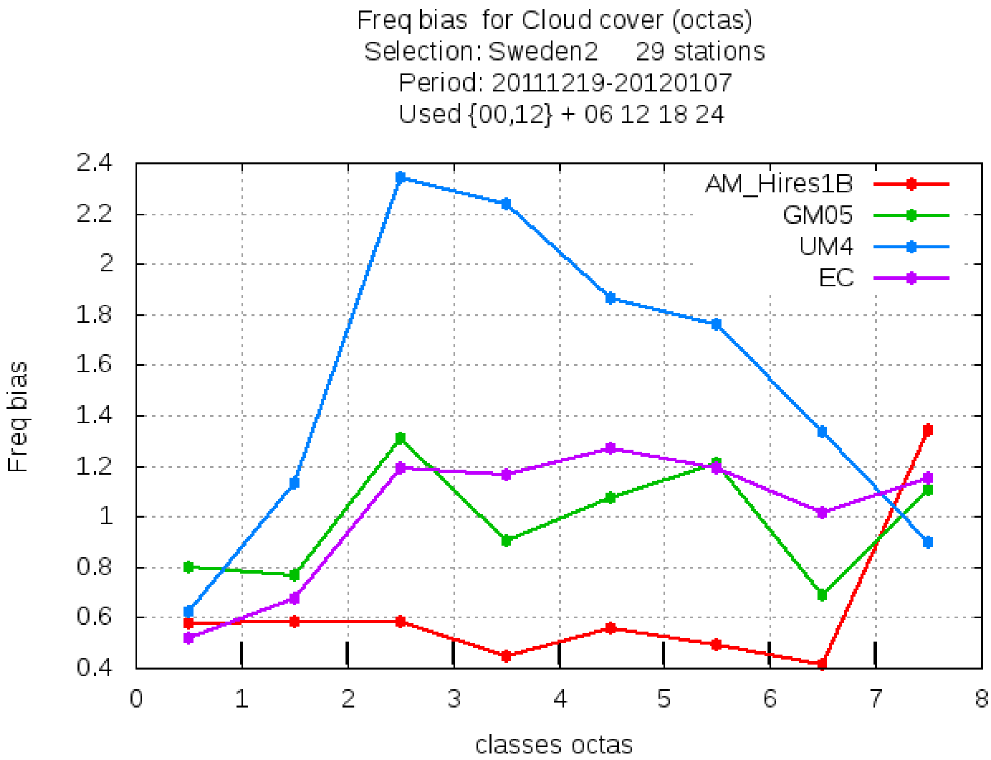


Figure 44: December/January, Frequency BIAS (FB) for total cloud cover averaged over all Norwegian (top) and Swedish (bottom) stations.

3.6 Verification study of March 2012

In this study the daily runs from MetCoOp of HARMONIE AROME (2,5 km) are compared with daily runs of HIRLAM (G05) from SMHI and UM (4 km) from *met.no* and the ECMWF forecasts.

The verification results following in this chapter are divided into the main weather elements that are most relevant for users and give a quite good overview of the weather situation. The figures show different skill scores for the four models and are shown in order to give a picture of the behaviour of the different models, how well they fit the observations and how the models compare with each other.

3.6.1 Meteorological description of the period

March 2012 started out with high pressure over Scandinavia and calm weather. During the second and third week several low pressure systems travelled from west to east in the northern part of the domain, causing strong wind and large amounts of precipitation especially along the western and northern parts of Norway. At the end of the third week some small scale systems developed in the northerly flow; a warning about a potential polar low was issued March 17th. On the 19th a cold front passed southern Sweden with observed peak wind of 60 knots. The monthly mean temperature for Norway was 4.3 degrees higher than normal. Monthly mean precipitation was 50% of the normal in south-east of Norway, but 250–400% in Trøndelag and northern Norway. In the middle part of Sweden several record high temperatures were measured.

The late winter/early spring period from 5 – 23. March 2012 is also the Met Office FAAM campaign period. A selection of output from the Met Office UM 4km model runs has been made available for MetCoOp, and results will be added to the verification study when time permits in a later report.

3.6.2 Verification results of mean sea level pressure (MSLP)

In terms of RMSE ECMWF is better than the other models, and AROME and G05 score better than UM4. The errors in all models increase with altitude. Increased BIAS is found for all models as a function of lead time (ECMWF has the smallest BIAS). In general the superiority of ECMWF is more pronounced in Sweden than in Norway. G05 shows slightly better result than AROME in Sweden.

The daily RMSE of AROME are very similar those for the ECMWF model for parts of the examined period, but deviate on some dates. This is especially seen 9–11. March and on the 17. March. This should be investigated further together with the similar but more pronounced low pressure cases from the December/January period.

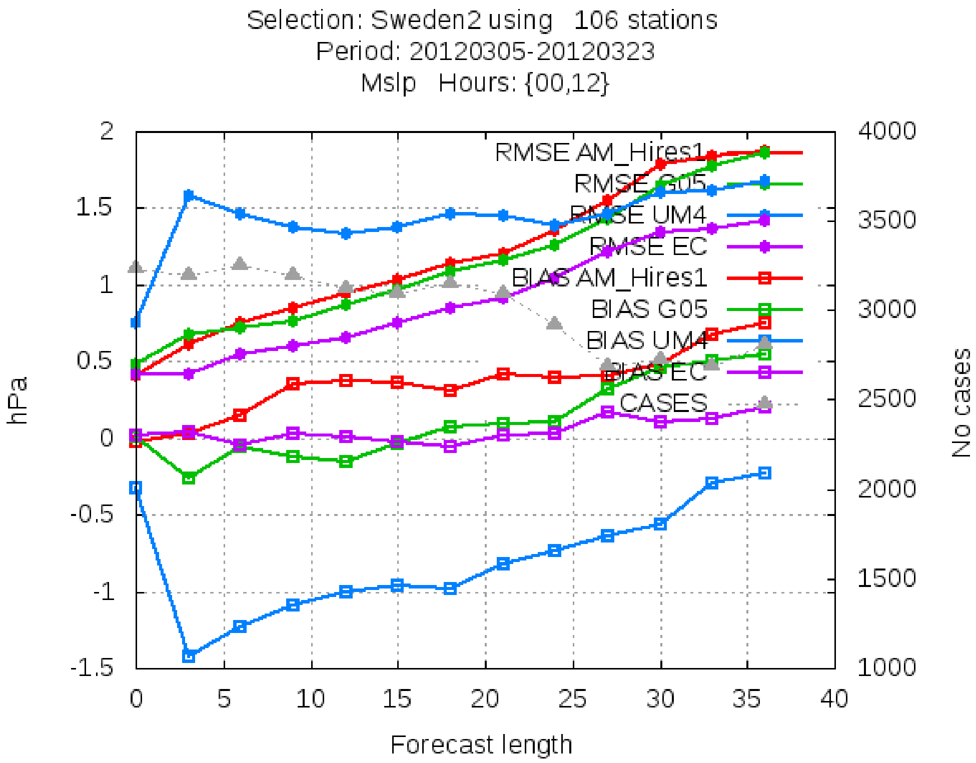


Figure 45: March 2012, RMSE/BIAS, MSLP all Norwegian (top) and all Swedish (bottom) stations.

3.6.3 Verification results of 10 m wind direction

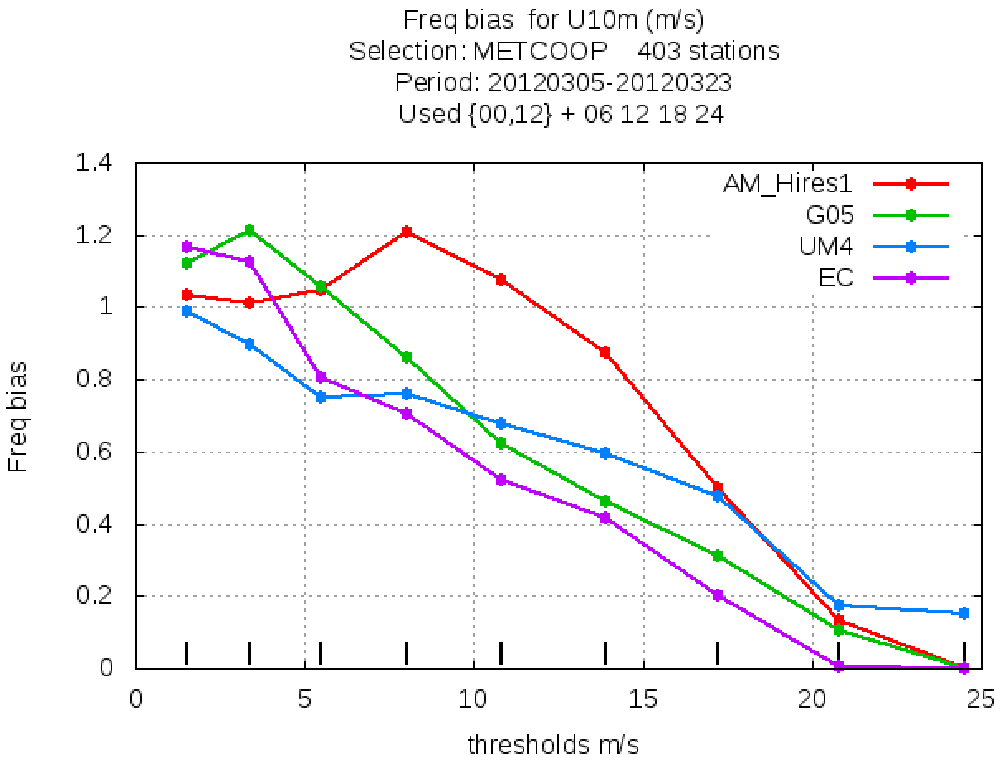
Averaged over all stations, AROME shows the smallest RMSE, closely followed by ECMWF and G05, while UM4 has the largest RMSE. However, AROME shows slightly smaller RMSE than the other models (more pronounced in Norway than in Sweden, and more pronounced inland). The RMSE is much smaller for all models at the coast, but ECMWF is better than AROME for lead times exceeding +20h.

3.6.4 Verification results of 10 m wind speed

G05 shows the smallest RMSE averaged over all stations, followed by AROME, ECMWF and UM4. While AROME and G05 show a small positive BIAS, ECMWF on average slightly underestimates the wind speed. The underestimation of UM4 is more pronounced (-0.5 m/s). The relative ranking of the models in terms of RMSE varies with regions, but AROME is superior to the other models in the mountains and performs quite well at the coast.

With respect to the frequency of different wind speeds AROME is in better agreement with observations for wind speed between 7.5 m/s and 17.5 m/s. This is most pronounced inland and in the mountains. At the coast UM4 seems to produce too many incidents with wind speeds higher than 15 m/s.

AROME produces on average the highest ETS (Figure 47), while ECMWF produces the lowest score. AROME is especially better than the other models for high wind speeds (>7.5 m/s), in the mountains and at Swedish stations. In Norway the quality of AROME, UM4 and G05 is rather similar. In the northern part of the area, UM4, G05 and ECMWF all show better result than AROME for the highest wind speeds (above 10 m/s) associated with an increased False Alarm Ratio (FAR) in AROME not compensated for by higher Hit Rate (HR). In general the higher ETS for AROME is due to higher HR which more than compensates for an increased FAR. Additionally some large errors can be seen in some main cities in Norway and Sweden. These large errors can be traced back to problems at certain dates, as seen by example in the time series from Trondheim (Figure 48). A snap shot of the wind forecasts from AROME (Figure 49) and HIRLAM (Figure 50) together with observations are shown. The figures show that AROME has much more wind on the lee side of the mountains compared with observations and with HIRLAM. However, both HIRLAM and the observations display areas with high wind, but these areas are much smaller than what is forecast by AROME. The reason for this behaviour is unknown and further investigations are needed.



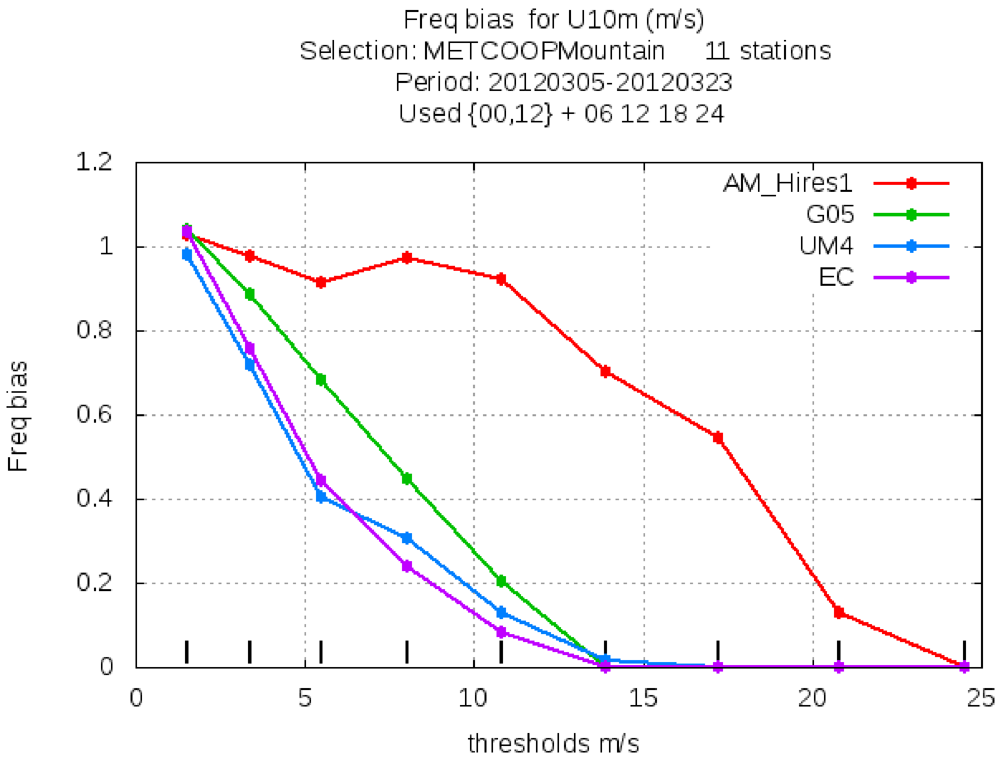


Figure 46: March 2012, Frequency BIAS (FB), wind speed for all stations (top) and mountain stations (bottom)

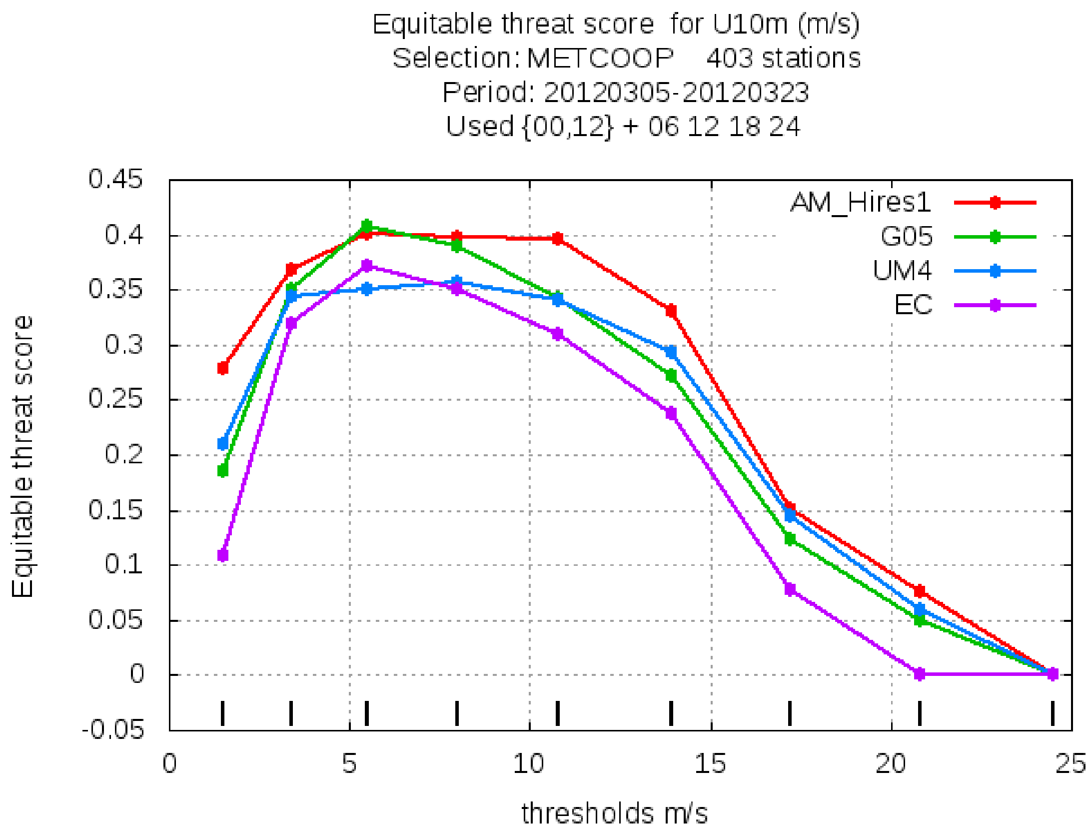


Figure 47: March 2012, ETS, averaged over all stations.

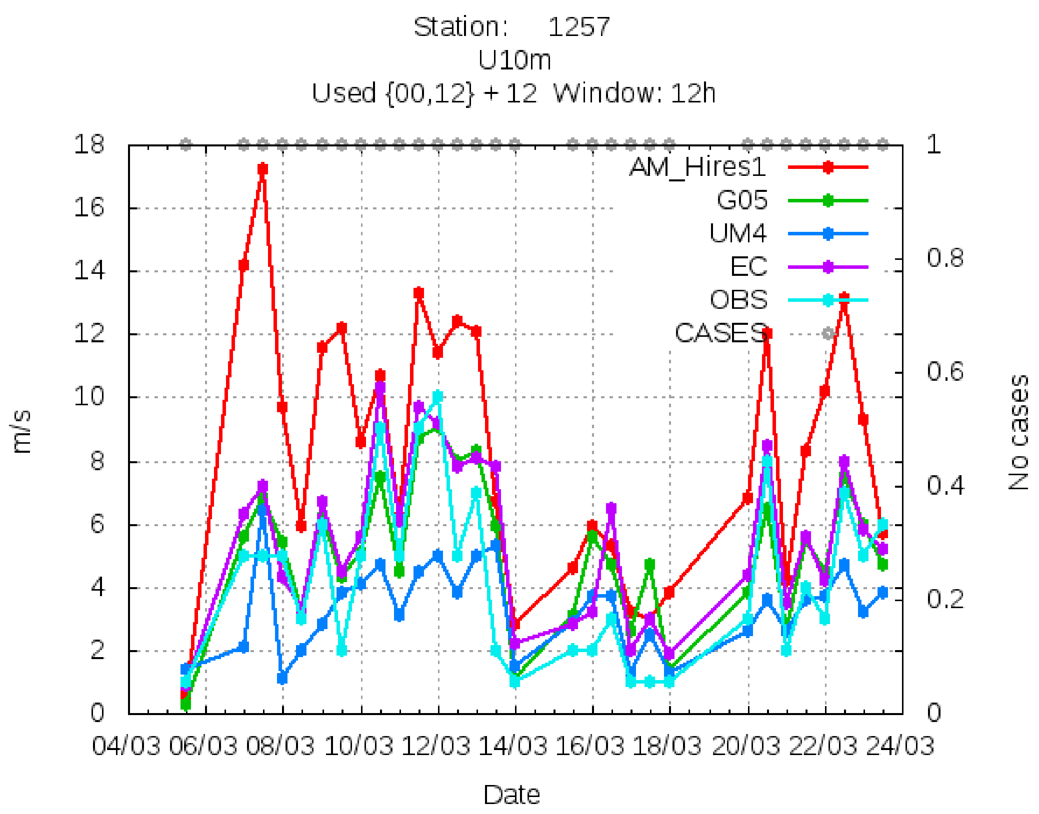


Figure 48: March 2012, Time series of 10 m wind speed from Trondheim station, Voll.

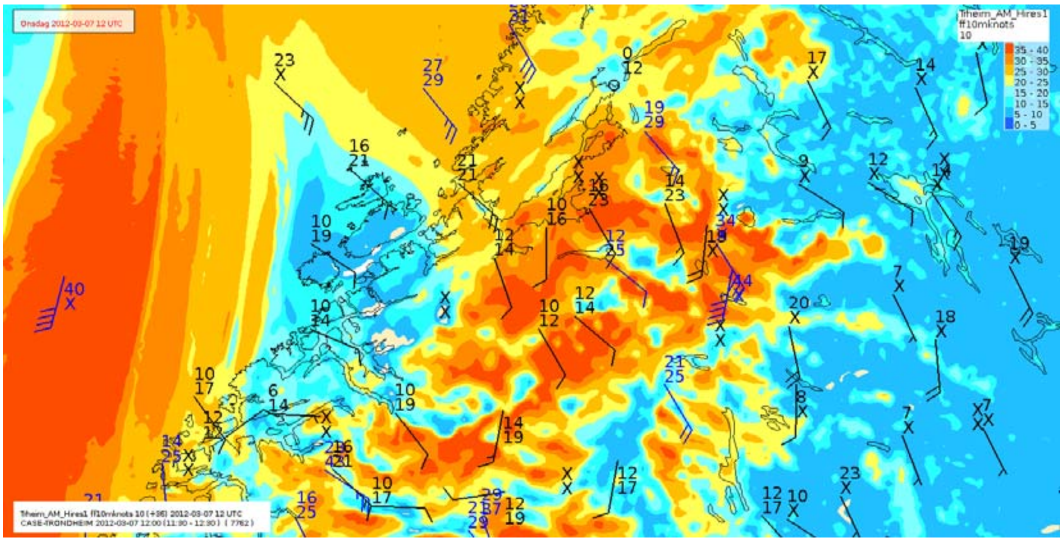


Figure 49: March 2012 AROME wind speed (knots) valid for 7th march 12UTC with observed wind and max observed last hour (blue when observed wind speeds above 25 knots).

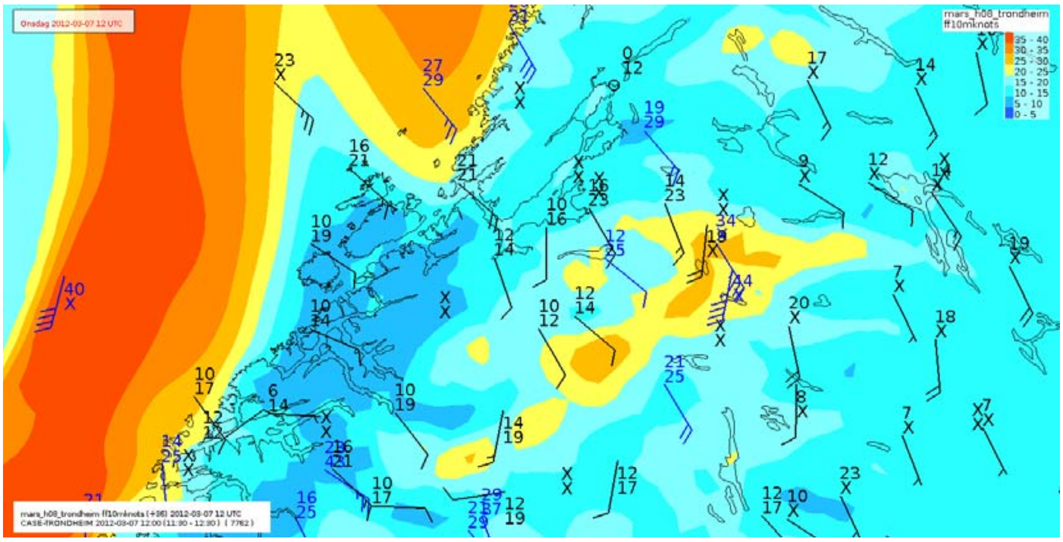
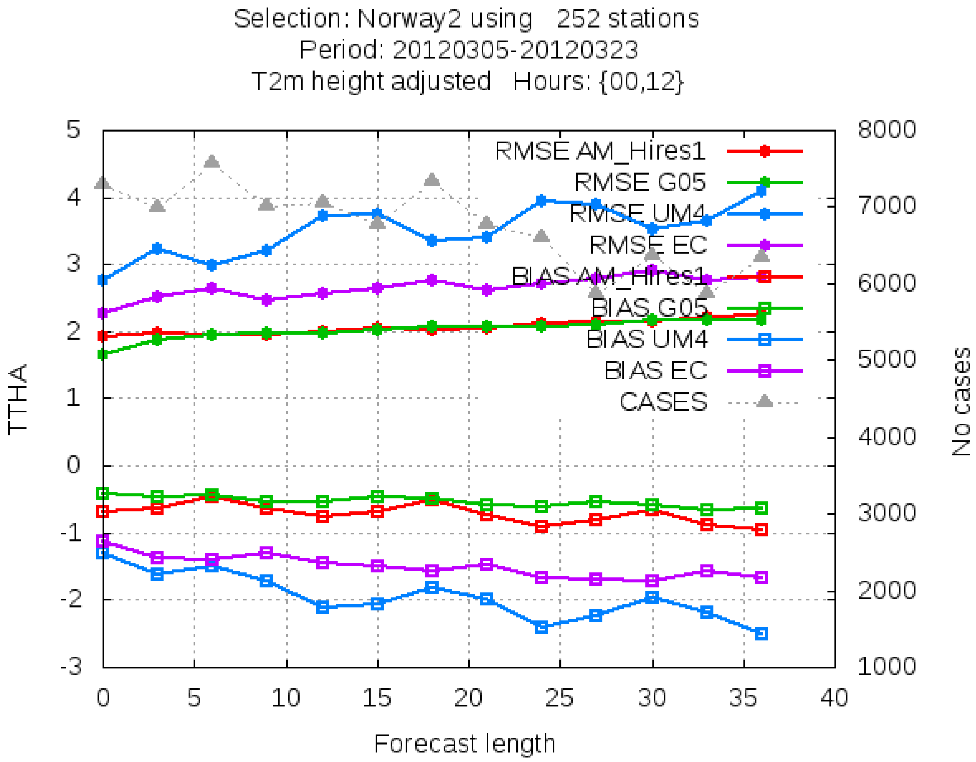


Figure 50: March 2012 HIRLAM (operational met.no version) wind speed (knots) valid for 7th march 12UTC with observed wind and max observed last hour (blue when observed wind speeds above 25 knots).

3.6.5 Verification results of near surface temperature (T2m)

All models show a systematic underestimation of T2m averaged over all stations (G05 0.5°C, AROME 0.5–1.0°C, ECMWF 1–1.5°C and UM4 1–2°C). In terms of RMSE G05 is slightly better than AROME, which again is much better than ECMWF. UM4 has the largest RMSE with respect to temperature. In Norway, AROME and G05 are of similar quality and better than ECMWF and UM4. However, in Sweden G05 is clearly better than the other models and AROME and ECMWF are of similar quality. A special feature is seen at the coast where ECMWF shows a clearly smaller RMSE than G05 which again is slightly better than AROME. AROME underestimates the temperatures with 0.5–1.0°C. Based on the scatter plots we see that the warm BIAS for cold observed temperature, as seen for other periods, is present in March as well. Additionally, for all models we find a tendency of incorrect forecasts of maximum temperatures. The latter is most pronounced at the coast.



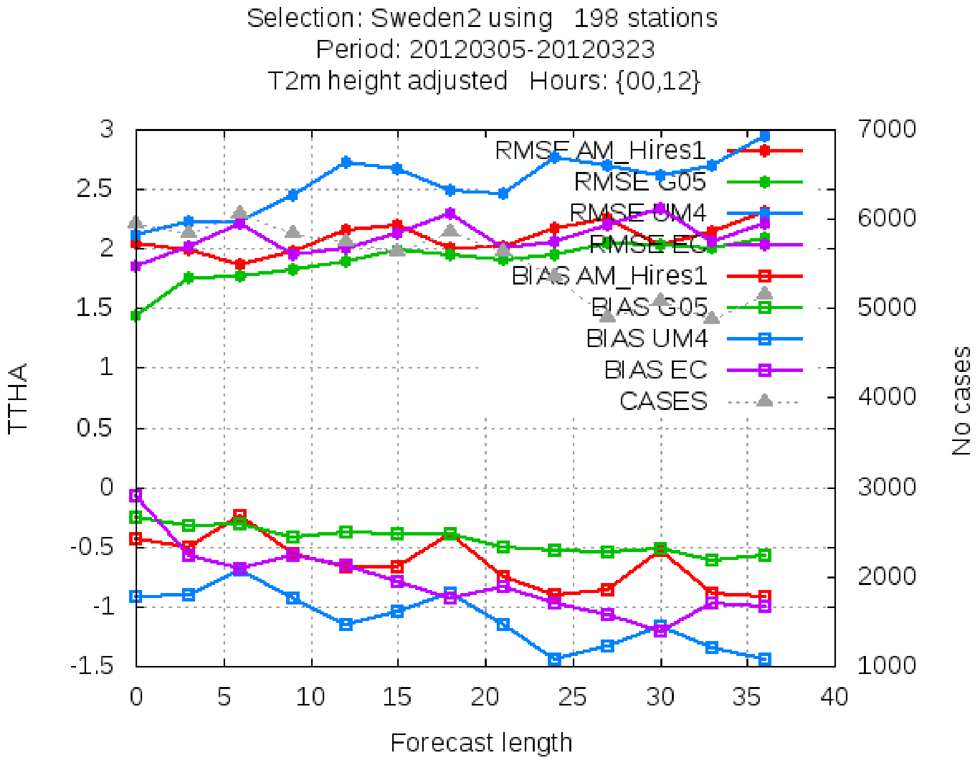


Figure 51: March 2012, RMSE/BIAS for Norwegian (top) and Swedish (bottom) stations.

3.6.6 Verification results of precipitation (12h)

Averaged over all stations AROME shows no systematic error for the examined period, while the three other models overestimate precipitation amounts with approximately 0.6mm/12h. AROME also shows the smallest RMSE, slightly better than ECMWF and better than G05 and UM4. An exception is in the mountains where AROME (and ECMWF) overestimate the precipitation and AROME shows the largest RMSE. However, with few mountain stations, and high observational uncertainty in the mountains during winter, it is difficult to draw conclusions from this region.

The precipitation frequency is in general very well described with AROME, with a slight overestimation (less than 10%) for all thresholds. For small precipitation amounts the other models overestimate the frequency with 15% (UM4), 35% (G05) and 45% (ECMWF), while similar numbers for 10 mm/12h are 30% (ECMWF), 50% (G05) and 85% (UM4). There is however a difference in the behaviour in Sweden and Norway for AROME (Figure 53): In Norway, the model is in very good agreement with the observed frequency, while in Sweden the occurrences are overestimated (30% for small precipitation amounts and more than 300% for 10 mm/12h).

Averaged over all stations AROME has the best ETS on small precipitation amounts (Figure 52). For intermediate and larger precipitation amounts UM4, ECMWF and AROME have very similar scores. The high ETS of AROME is due to a lower false alarm rate than the other models, but has a lower hit rate.

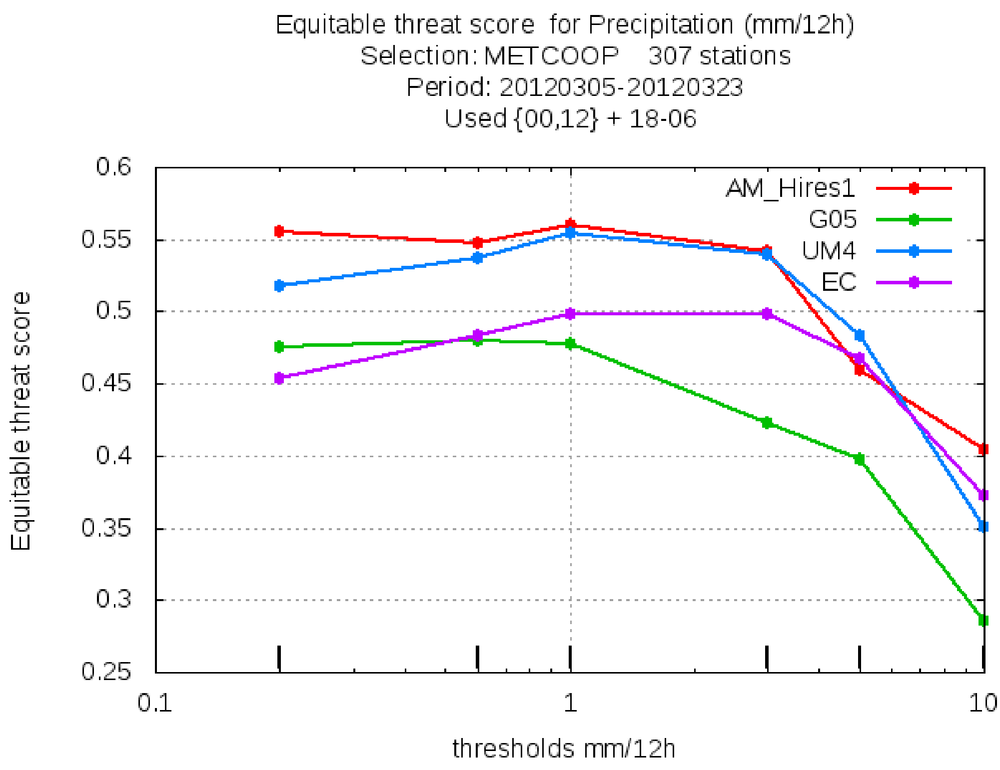
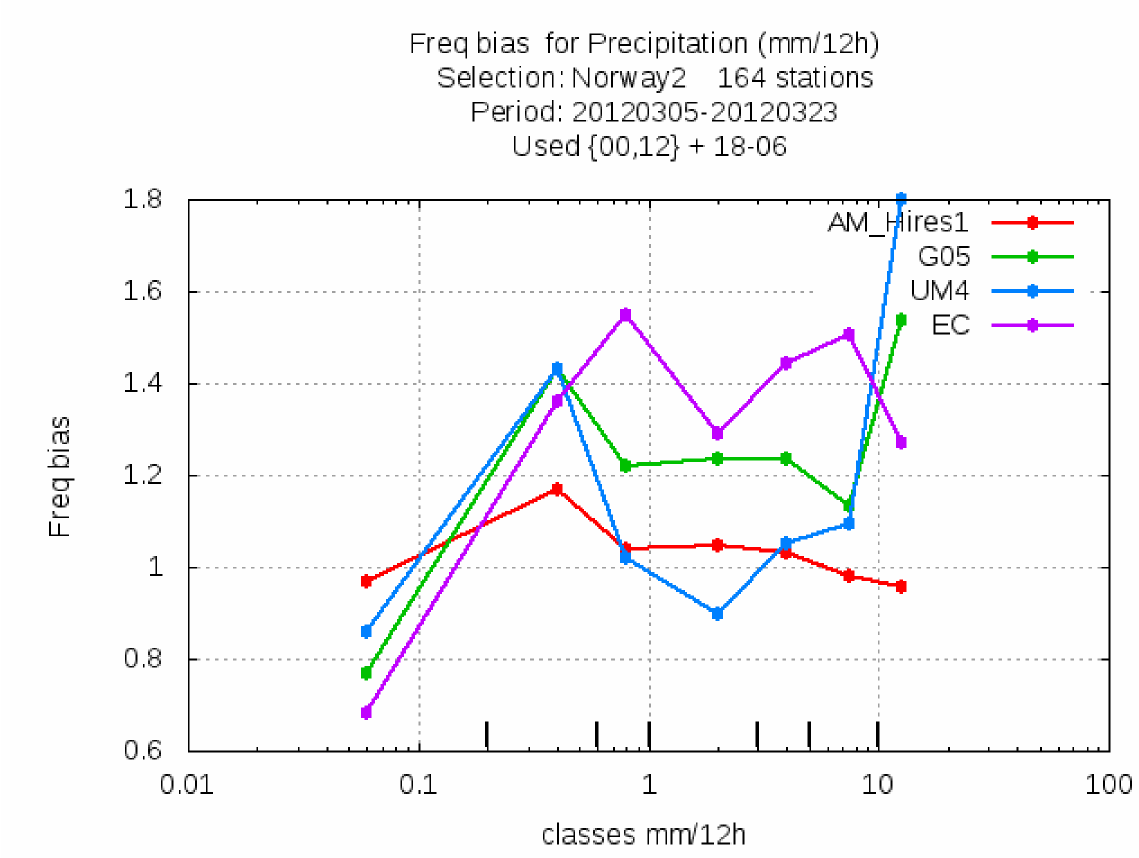


Figure 52: March 2012, ETS, 12hr precipitation all stations.



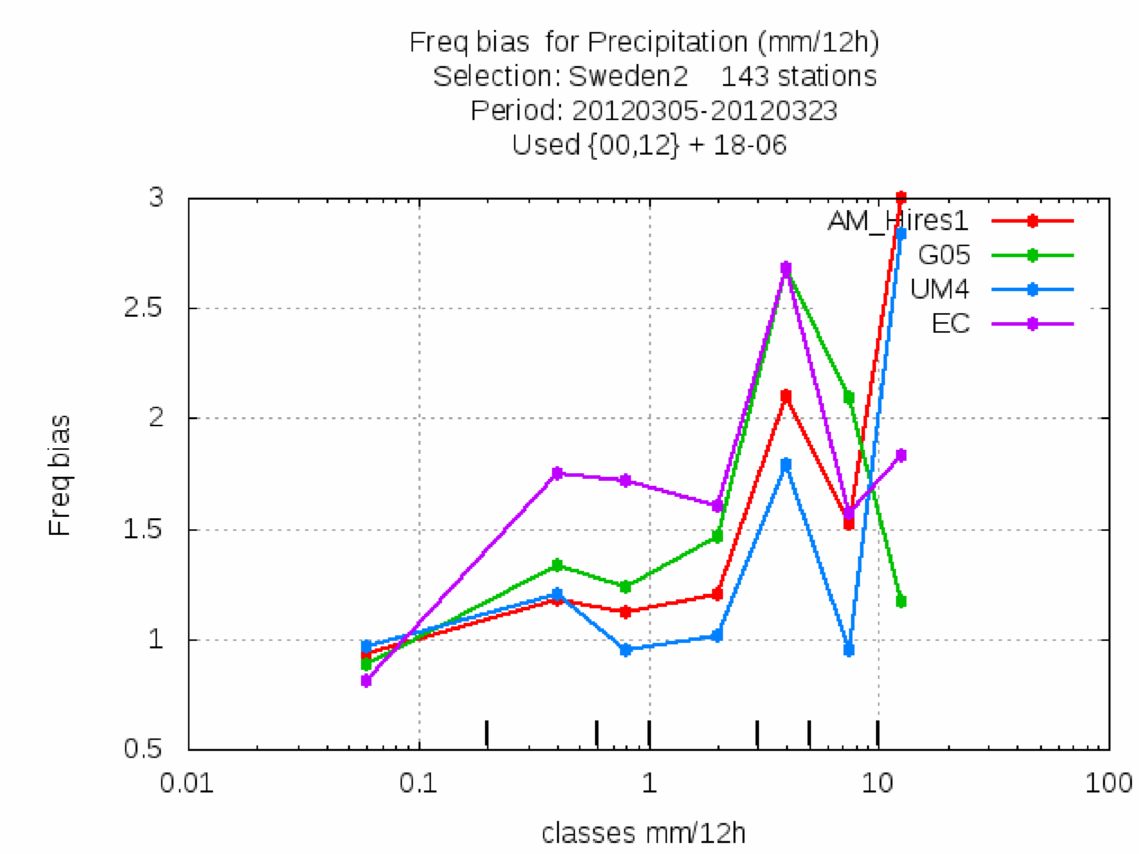


Figure 53: March 2012, Frequency BIAS (FB), 12hr precipitation for all Norwegian (top) and Swedish (bottom) stations.

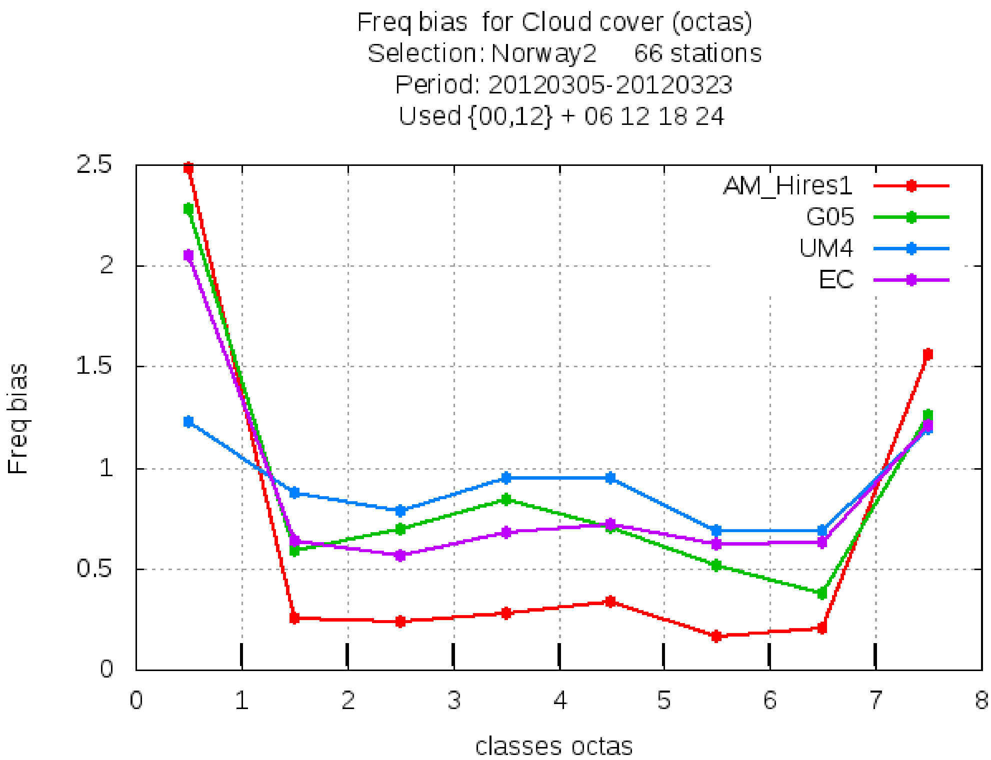
3.6.7 Verification results of total cloud cover (TCC)

On average AROME overestimates the cloud cover with 0.5–1.0 octas, while the three other models overestimate the cloud cover with 0–0.5 octas. In terms of RMSE UM4 is better than the other models, and AROME shows the largest RMSE.

ECMWF, AROME and G05 overestimate the occurrences of no-clouds, while UM4 underestimates no-clouds. All models overestimate the occurrences of total cloud cover, 15% by UM4, 25% by G05 and ECMWF and 60% by AROME. In general UM4 is in best agreement with the observed frequency. There are however, large differences between the model performances in Sweden and Norway. In Norway ECMWF, G05 and AROME forecast cloud free conditions more than twice as often as observed, and they also over-forecast cloudy conditions. In Sweden, AROME underestimates all conditions except for cloudy conditions, while UM4 has too many occurrences of partly cloudy conditions and underestimates the extremes.

Based on ETS, G05 is the better model on the low cloud amounts, while UM4 is the better model on cloudy conditions. AROME has problems compared to the other models for these weather parameters.

For the parameter total cloud cover it is very difficult to have a fair comparison between models with different resolutions.



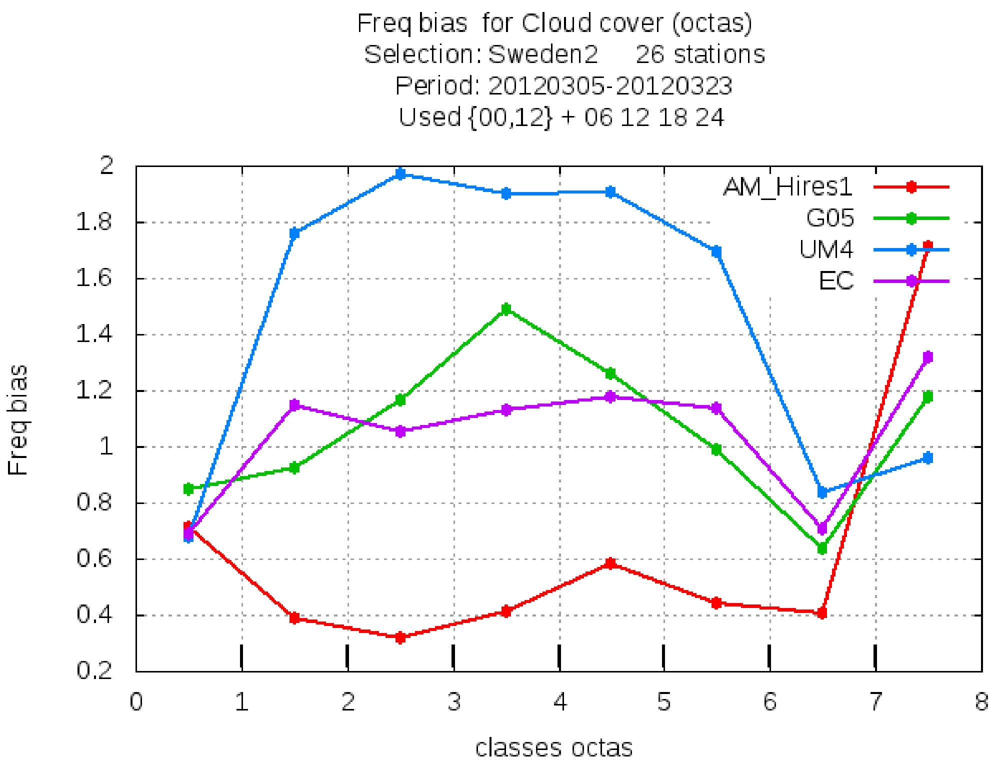


Figure 54: March 2012, Frequency BIAS (FB), Cloud Cover, Norway (top), Sweden (bottom).

3.7 Verification study of May 2012

In this study the daily runs from MetCoOp of HARMONIE AROME (2,5 km) are compared with daily runs of HIRLAM (G05) from SMHI and UM (4 km) from *met.no* and the ECMWF forecasts.

The verification results following in this chapter are divided into the main weather elements that are most relevant for users and give a quite good overview of the weather situation. The figures show different skill scores for the four models and are shown in order to give a picture of the behaviour of the different models, how well they fit the observations and how the models compare with each other.

3.7.1 Meteorological description of the period

The period in this study lasts one month from the 1 – 31 May. Low pressure activity dominated most of the month with north-westerly winds, sometimes interrupted by southerly winds. The low pressure systems were mainly centred in the North Sea or southern parts of Norwegian Sea and caused periods with chilly conditions in western and northern Norway. South-easterly parts of Norway experienced foehn wind and therefore the temperatures were mostly close to normal or above.

On May 22th a high pressure system established and gave warm and sunny conditions over major parts of Scandinavia for several days, especially southern Scandinavia. Temperature peaked 30 degrees Celsius in both Norway and Sweden, which is rather abnormal at this time of year. The high pressure system retrograded westward May 28th while a deep low pressure system propagated southward from the Barents Sea, which gave a big drop in temperature and windy conditions.

3.7.2 Verification results of mean sea level pressure (MSLP)

AROME is better than G05 and UM4, but not as good as ECMWF with respect to RMSE. AROME also shows a small systematic error increasing with lead time (less than 0.5hPa after 36h), while ECMWF show no systematic errors averaged over all stations. G05 and UM4 underestimate in general MSLP. These findings are in general valid for all regions, but the model differences are smaller at high elevations where there are also fewer stations.

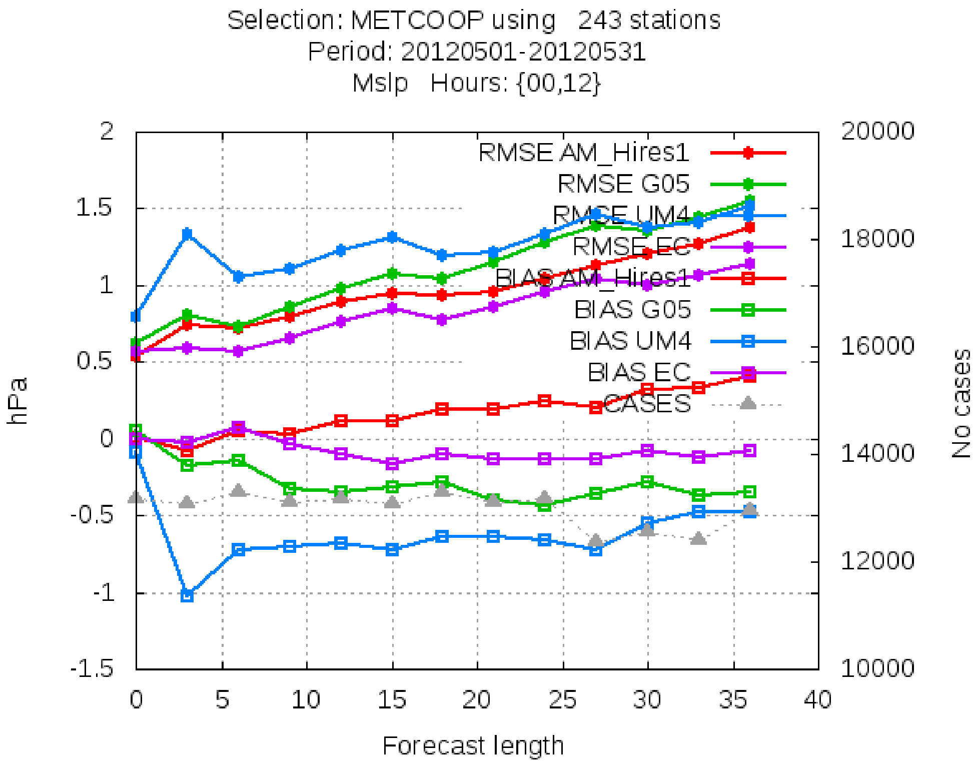


Figure 55: May2012, RMSE/BIAS, MSLP averaged over all stations.

3.7.3 Verification results of 10 m wind direction

Both in Sweden and in Norway, ECMWF is in general better than the other models in terms of smaller RMSE followed by AROME, G05 and UM4 (Figure 56). The gap between ECMWF and AROME is less pronounced inland and as the elevation of the stations increase. Why a better representation of topography and other local effects in the fine scale models does *not* result in smaller errors than the coarse resolution of ECMWF is unclear (but can possibly be due to sea breeze). This result has not been seen in the other periods that have been studied.

On average there are small systematic errors in the wind direction, but a small positive BIAS (5 degrees) is found with increasing resolution.

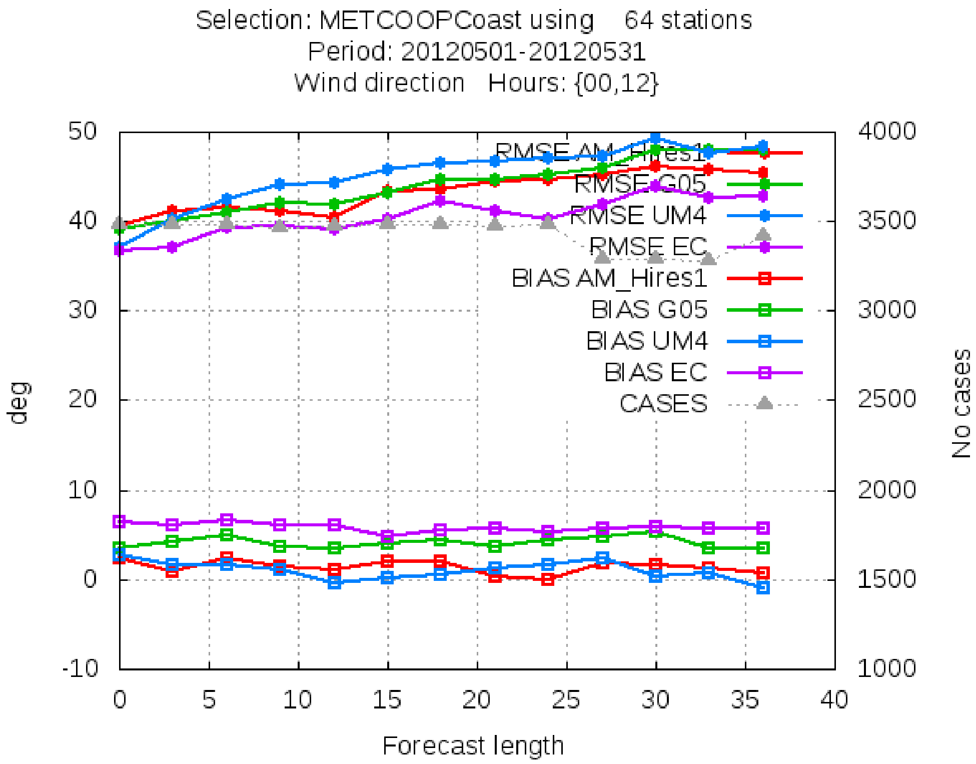


Figure 56: May 2012, RMSE/BIAS, 10 m wind direction at coastal stations.

3.7.4 Verification results of 10 m wind speed.

In terms of RMSE there are very small differences between the four models. ECMWF and UM4 averaged over all stations don't have any systematic errors, while AROME (or G05) show a small negative (or positive for G05) BIAS. The systematic underestimation of wind speed in AROME is most pronounced in Sweden on stations with elevation between 200 and 1000 meter above sea level and in the northern part of the domain. This negative BIAS leads to an increase in RMSE. In the mountains AROME shows a low positive BIAS. It is worth noticing that at the coast (or mountains), ECMWF shows smaller (or larger) RMSE than the other models.

Based on ETS for all stations (Figure 57) AROME is better than the three other models for wind up to 15 m/s, UM4 is better than G05, which again is better than ECMWF. However, for the strongest winds (>15 m/s) UM4 is best, followed by ECMWF. This latter behaviour is especially seen at Norwegian coast stations (in Sweden AROME and UM4 are best). Compared to earlier evaluated periods AROME shows an underestimation of occurrences with high wind speeds (Frequency BIAS (FB) well below 1). However, even though the May period is longer than the winter period there are very few occurrences of the exceedance of 15 m/s in May compared to the winter period. The wind forecast from AROME in the largest cities has been discussed as a problem area in some of the other periods. For this period, however, AROME shows good verification result with respect to ETS, see the figure below.

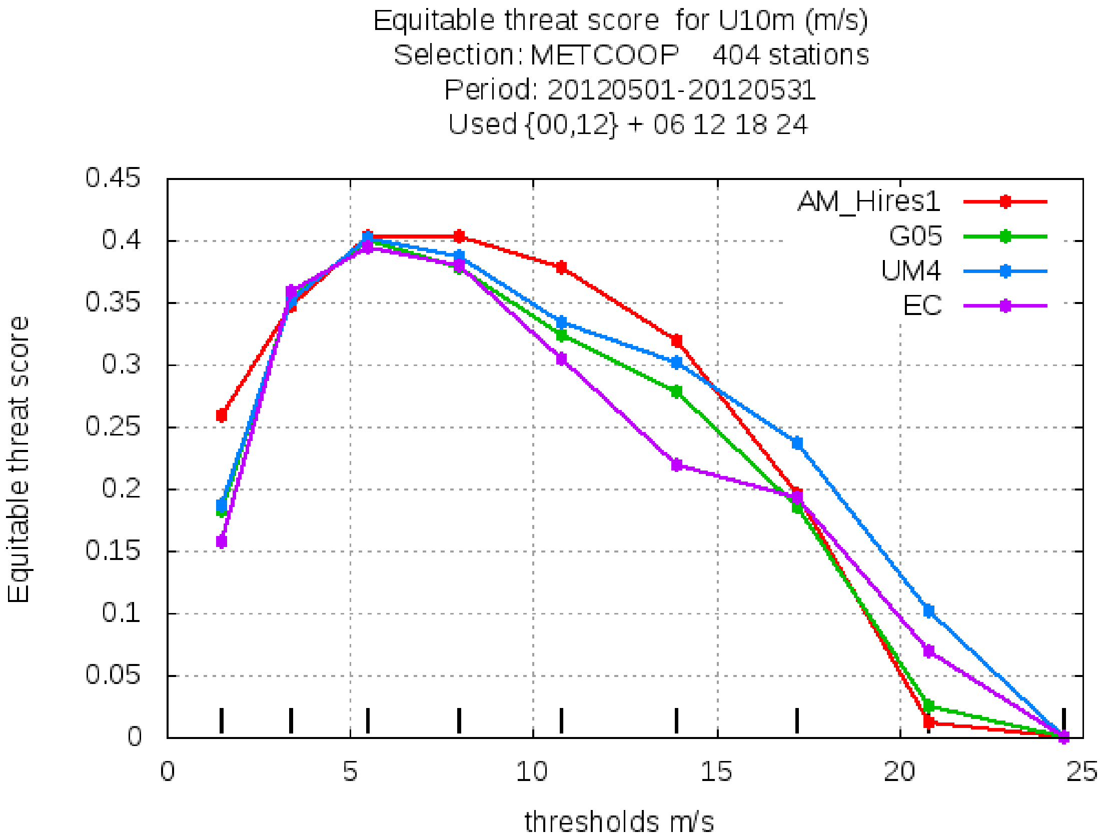


Figure 57: May 2012, ETS, 10 m wind speed averaged over all stations.

3.7.5 Verification results of near surface temperature (T2m)

All models with the exception of G05 show a negative BIAS (ECMWF 1°C, AROME 0.5–1.0°C, UM4 0.5°C). G05 is therefore the best model in terms of RMSE, while ECMWF and AROME show similar RMSE. The underestimation of temperature in AROME is present for all regions, but is more pronounced in Norway than in Sweden. It is more pronounced at mountain stations (-1.0°C) and coastal stations (-1.5 °C) (Figure 58). Based on the scatter plot (not shown) it is evident that it is particularly the high temperatures that the model(s) find difficult to forecast properly. The time series plot (Figure 59) reveals that there is a general underestimation of the temperatures and its diurnal amplitude. This latter is especially evident during the warm period in late May.

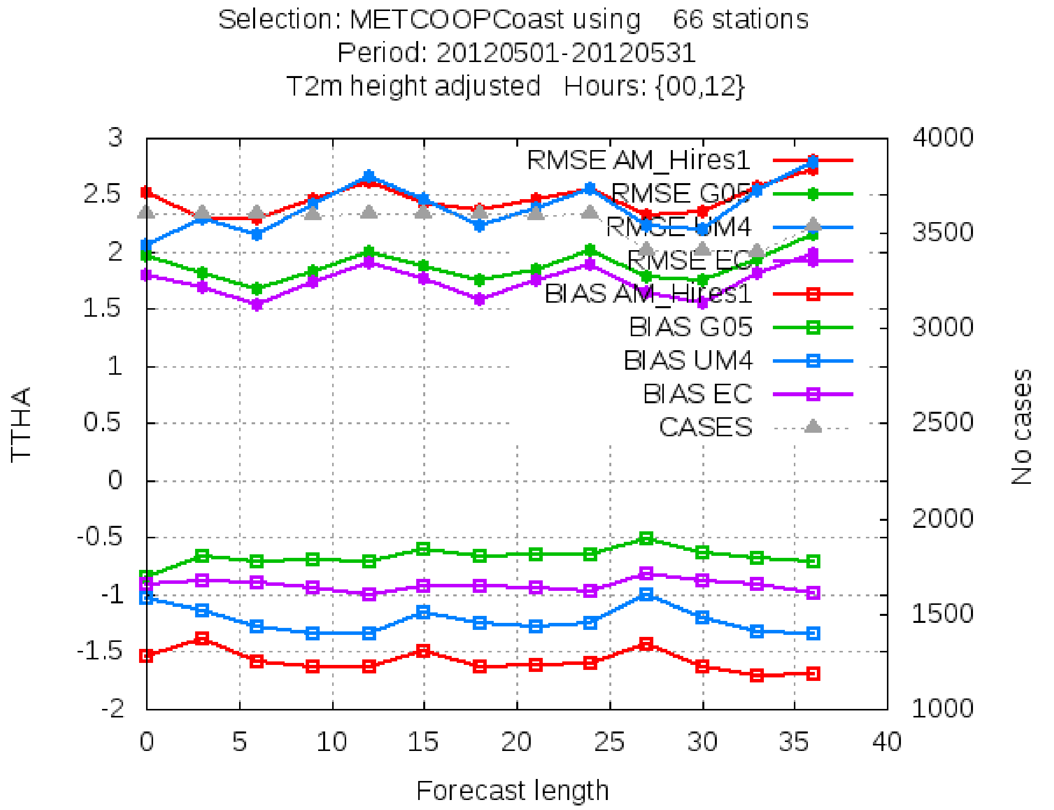


Figure 58: May 2012, RMSE/BIAS, averaged over coast stations.

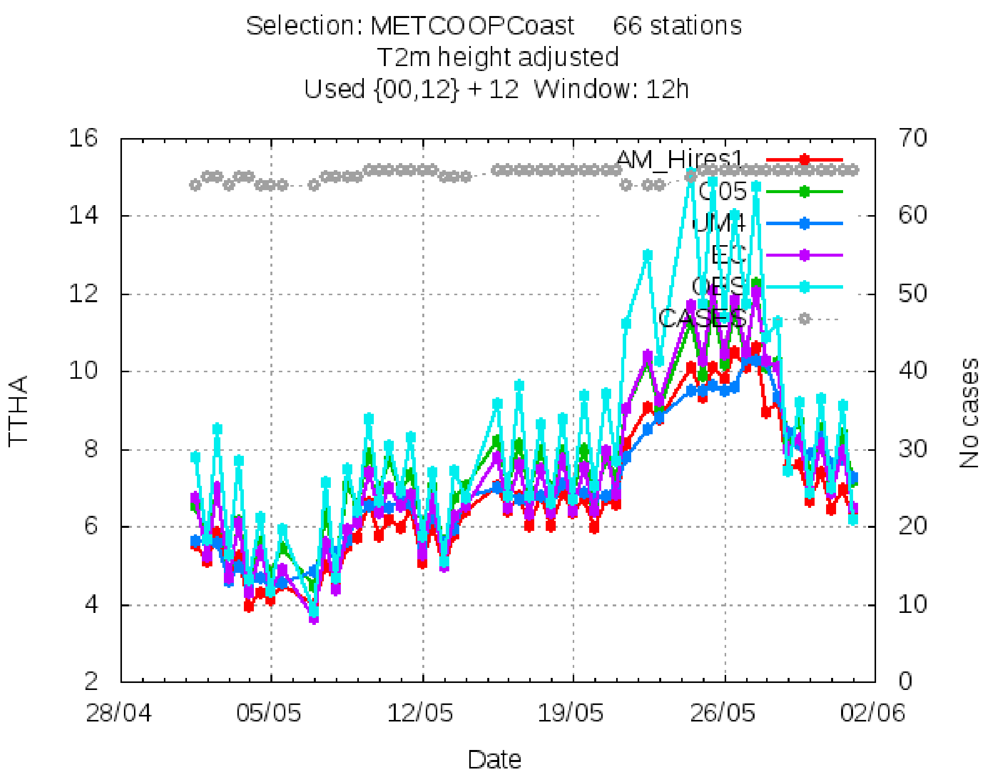


Figure 59: May 2012. Time series of model and observed temperatures averaged over coastal stations

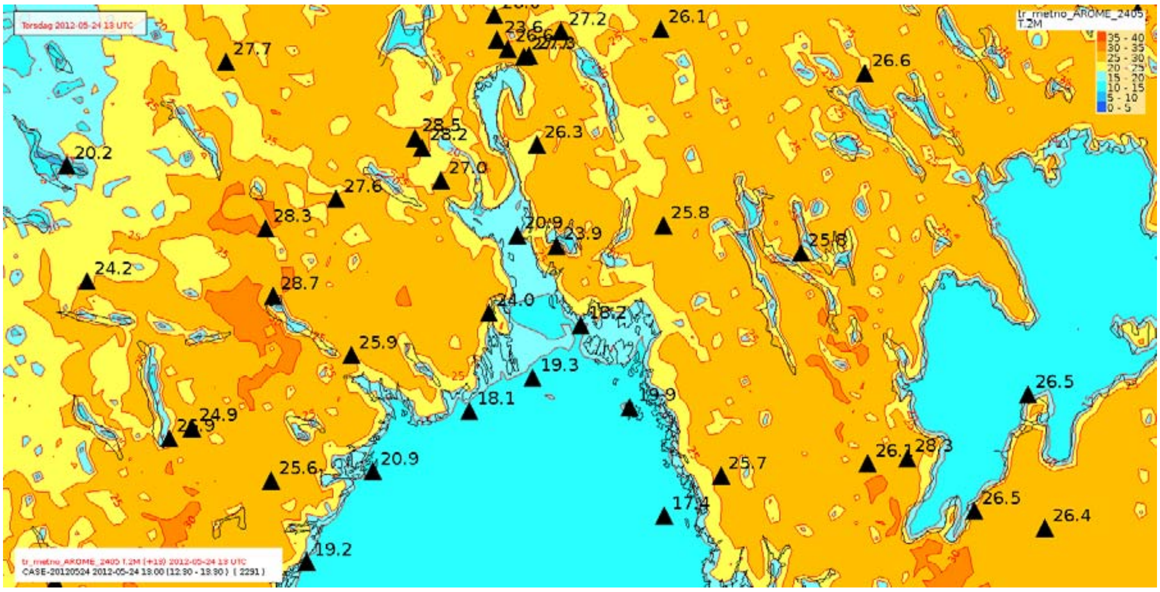


Figure 60: Snapshot of a +13hr forecast of T2m from AROME valid for 25/05/12 13UTC.

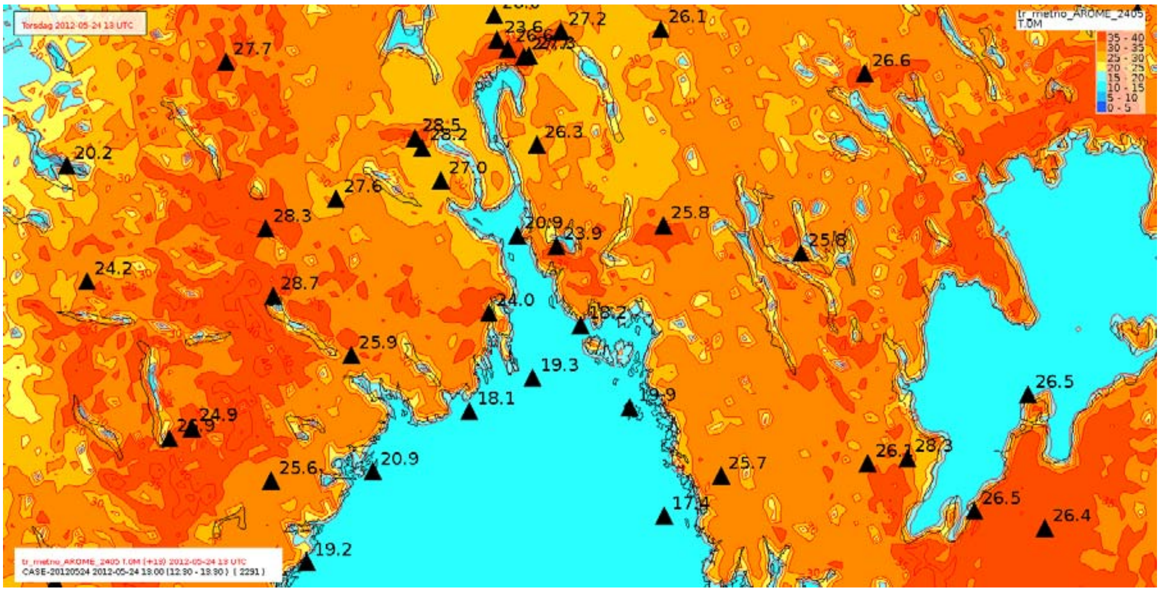


Figure 61: Snapshot of a +13hr forecast of T0m from AROME valid for 25/05/12 13UTC.

A snapshot of T2m and T0m (surface temperature) from the *met.no* version of the AROME model (figure above) illustrates how the model forecasts too cold temperatures at the stations at the outer part of the coast. This should be investigated further. Another issue which is illustrated in Figure 61, is the importance of the physiographic fields used in the model. We can see local spots with drastically colder temperatures than the surroundings. These differences are due to different surface characteristics and are not necessarily correct (as in the land-coast contrast example). Furthermore, this behaviour is more pronounced in very high resolution models than in coarse resolution models which to a certain degree smooth local differences.

3.7.6 Verification results for 12hr precipitation

All models show a positive BIAS averaged over all stations (UM4 0.3mm/12hr, AROME 0.35mm/hr, ECMWF 0.6mm/12hr and G05 0.75mm/12hr). Additionally ECMWF give the smallest RMSE, followed by AROME, UM4 and G05. In Norway AROME and ECMWF show similar RMSE score, but only slightly better than UM4, while in Sweden ECMWF is clearly better than the three other models.

The Frequency BIAS (FB) plot (Figure 62) reveals large errors in all models with respect to the occurrences of exceeding different precipitation thresholds. The overestimation of occurrences of small precipitation amounts is evident in UM4 with 10%, AROME with 20%, G05 with more than 50% and ECMWF with 70%. Towards larger precipitation amounts, the overestimations increase further for AROME, UM4 and G05, while ECMWF shows a decrease (but still overestimation).

In terms of ETS averaged over all stations, UM4 is better than AROME, and clearly better than ECMWF and G05 for all precipitation amounts (0–10 mm/12hr), see Figure 63. This is due to a much lower false alarm ratio in UM4 (and AROME) compared to ECMWF and G05, which more than compensates for the lower hit rate in AROME and UM (for less than 7mm/12hr). The same picture is seen when only looking at the Norwegian stations (Figure 63). However, for Swedish stations the ECMWF score is better than the other models for more than 3mm/12hr. Comparing the ETS from Sweden and Norway based on the best model for each threshold, reveals that the forecast in Sweden was much better than in Norway for small precipitation amounts.

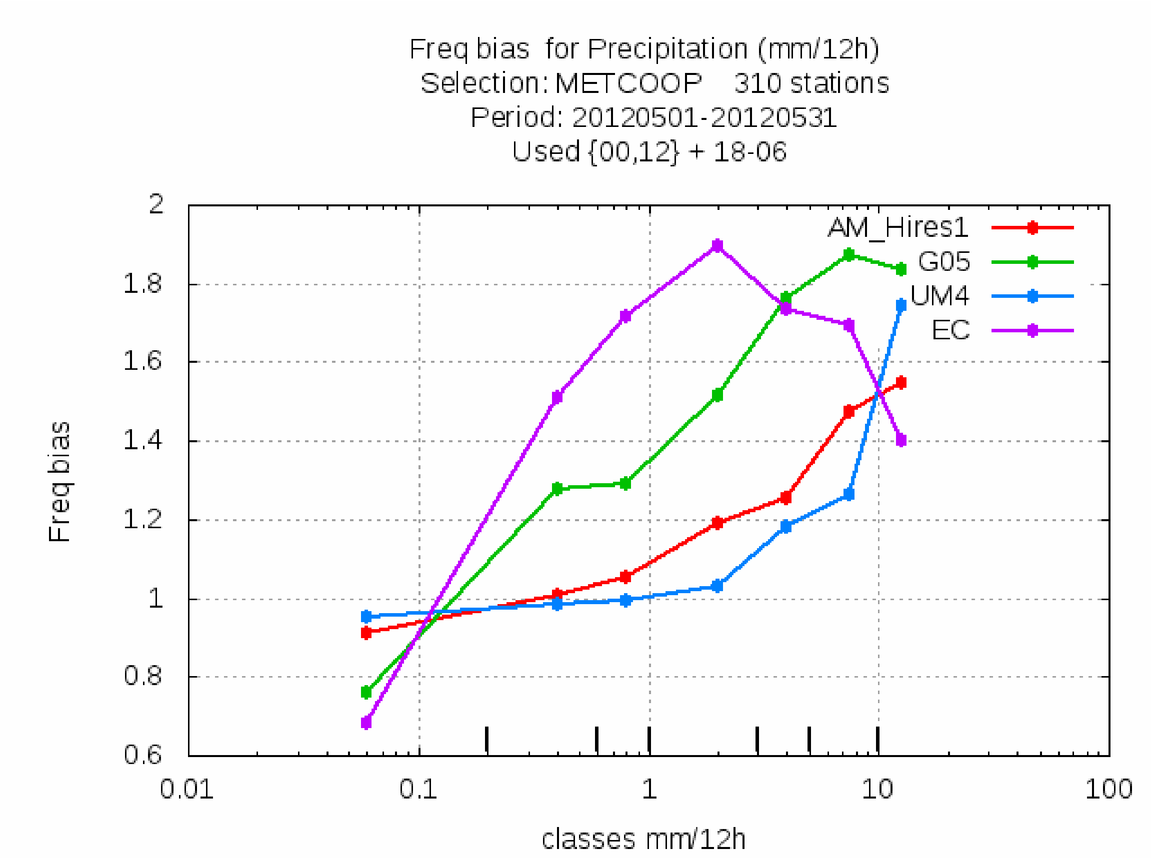
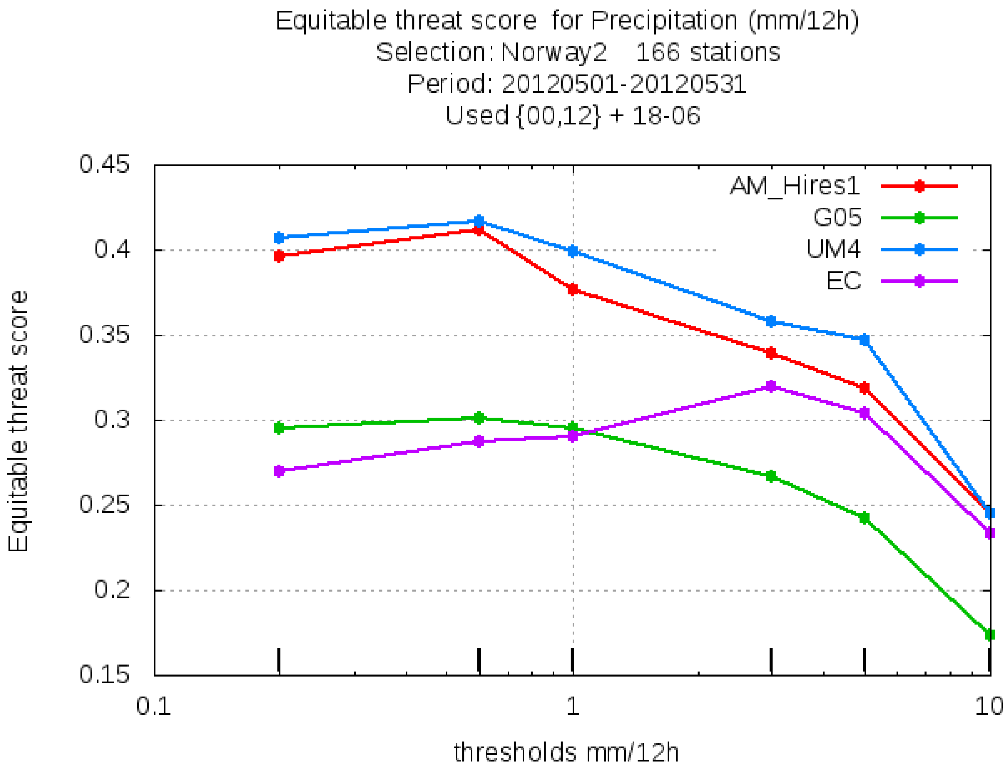


Figure 62: May2012, Frequency BIAS (FB), 12hr precipitation averaged over all stations.



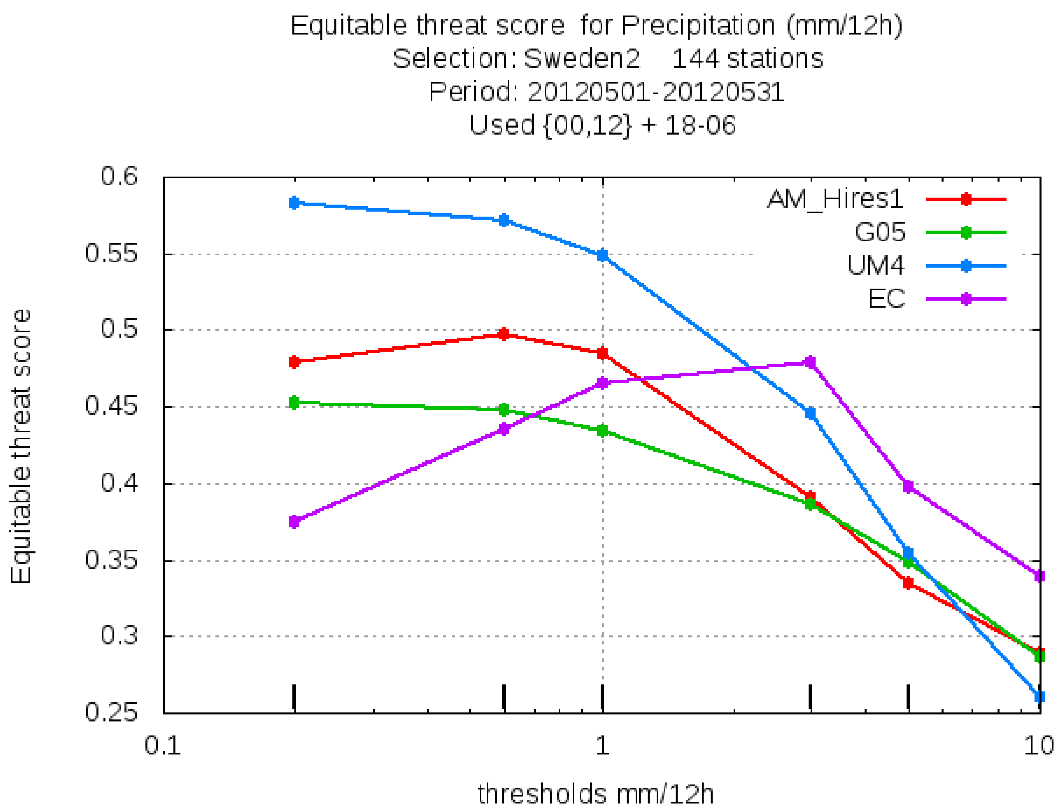
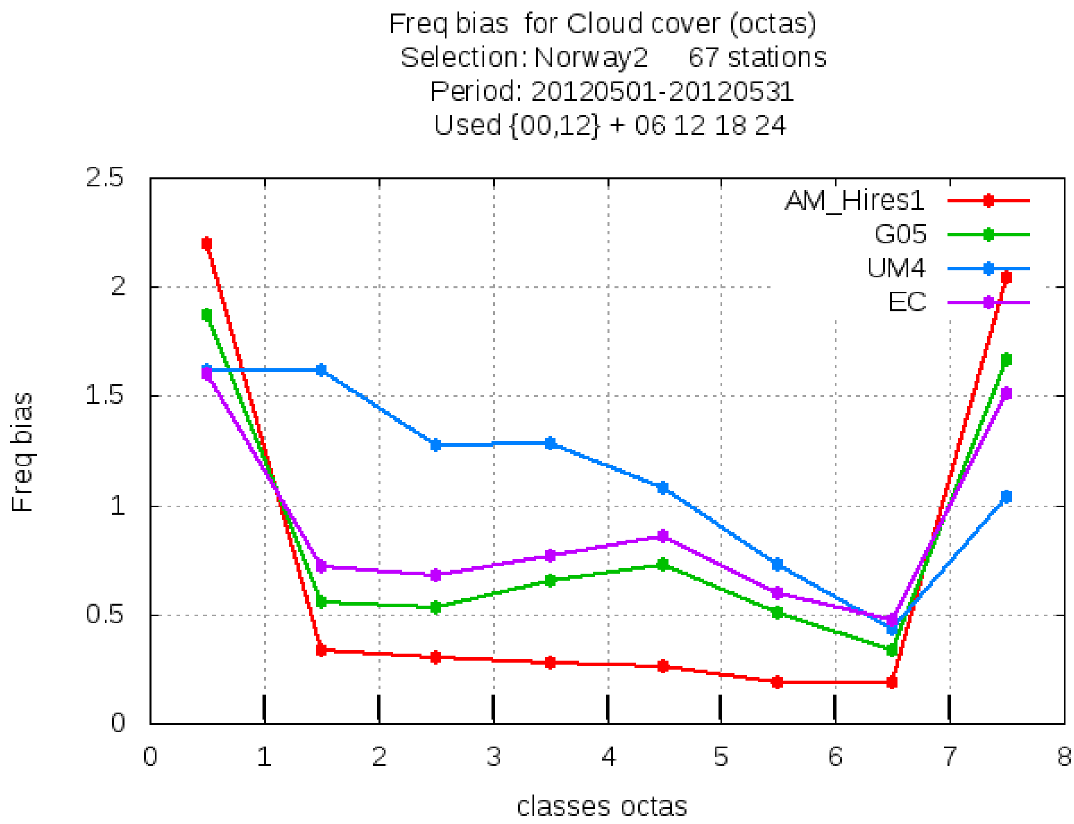


Figure 63: May 2012, ETS, 12hr precipitation Norway (top) and Sweden (bottom).

3.7.7 Verification results of total cloud cover (TCC)

Averaged over all stations, UM4 shows the smallest RMSE, but an underestimation of the total cloud cover (0.5octas). The three other models overestimate the cloud cover and AROME has the largest BIAS (0.5–1.0octas) and the largest RMSE.

The Frequency BIAS (FB) plot (Figure 64) reveals that all models forecast too many occasions of total cloud-free (ECMWF <10%, UM4 15%, G05 25% and AROME 40%). Also the occurrences for total cloud cover (8/8 octas) is heavily overestimated in AROME (>100%), G05 and ECMWF (>60%), while UM4 shows no BIAS. On the intermediate steps (1/8 – 7/8 octas) all models underestimate the occurences except for UM4 that shows a slight overestimation up to 5/8. There are however large differences between Norway and Sweden, which should be investigated further. For all stations, G05 and UM4 show best result with respect to ETS, while AROME shows least good result.



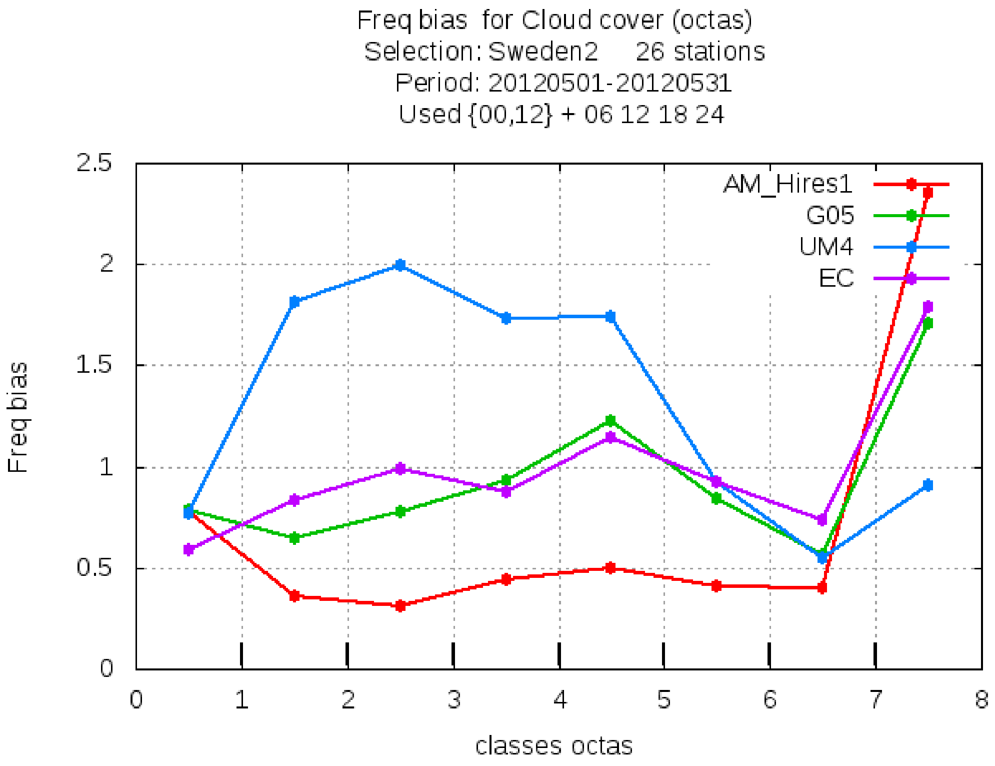


Figure 64: May 2012, Frequency BIAS (FB), total cloud cover, Norway (top) and Sweden (bottom).

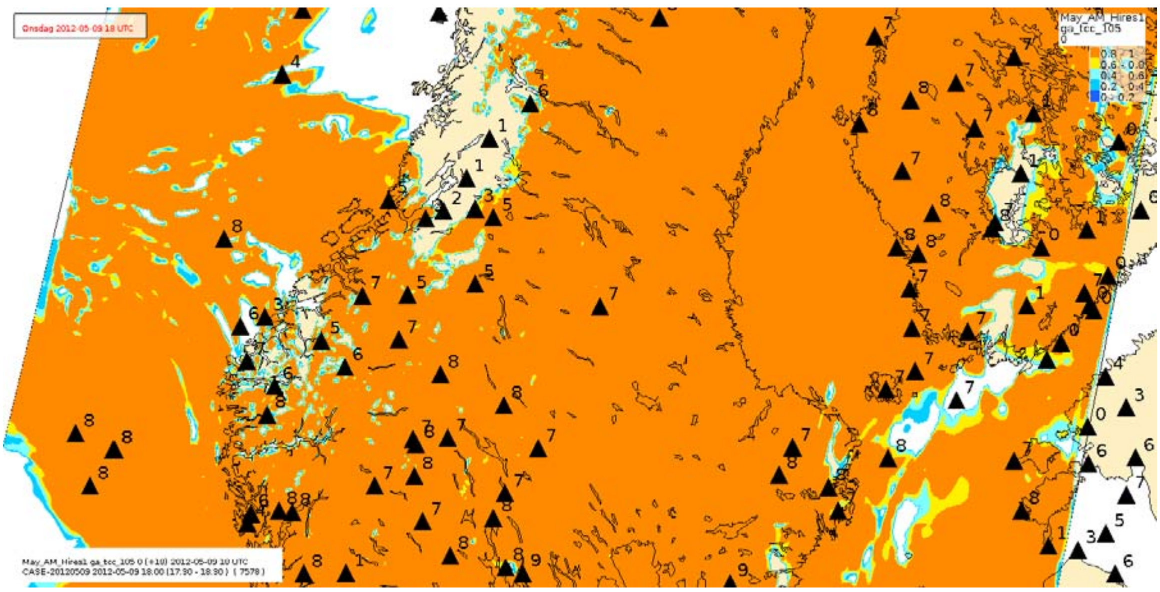


Figure 65: Snapshot of +18hr forecast of total cloud cover from AROME and observations, valid for 09/05/12 18UTC.

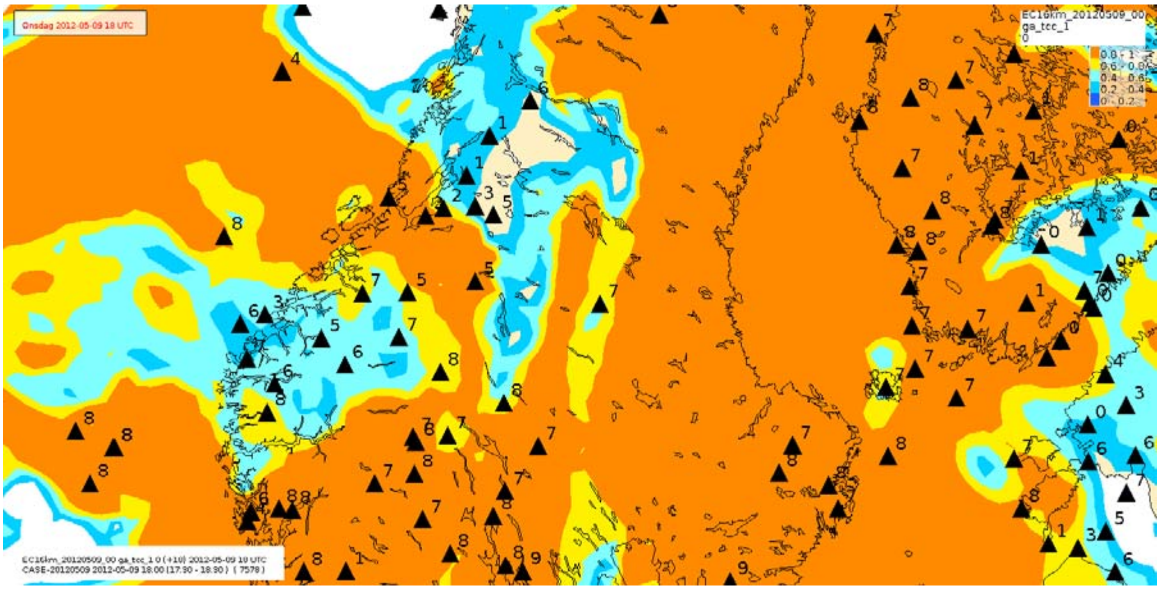


Figure 66: Snapshot of +18hr forecast of total cloud cover from ECMWF and observations, valid for 09/05/12 18UTC.

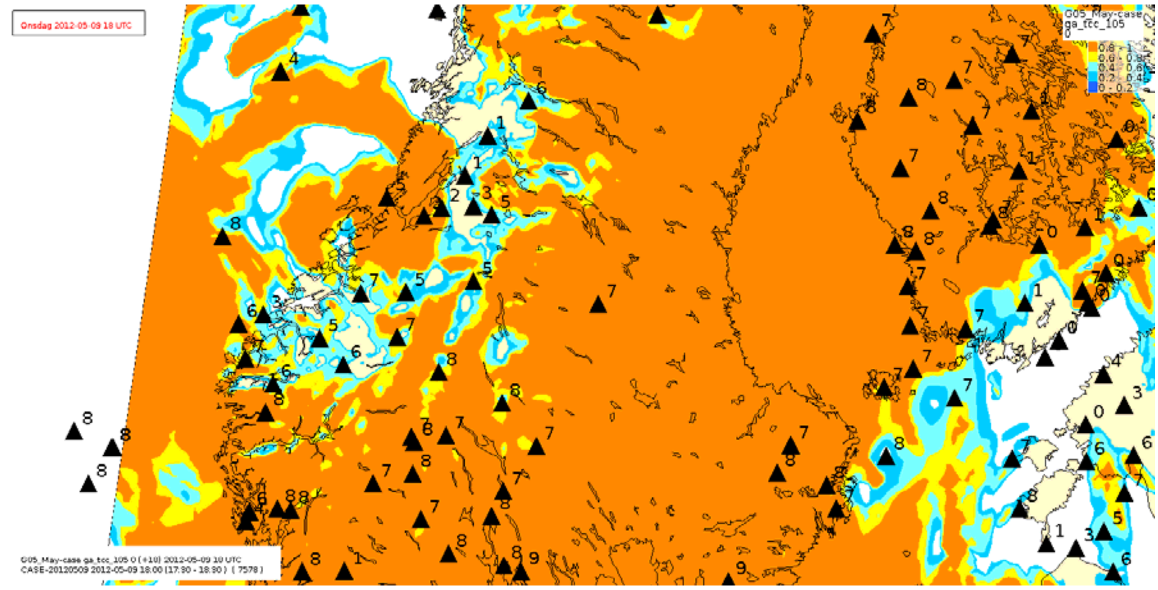


Figure 67: Snapshot of +18hr forecast of total cloud cover from G05 and observations, valid for 09/05/12 18UTC.

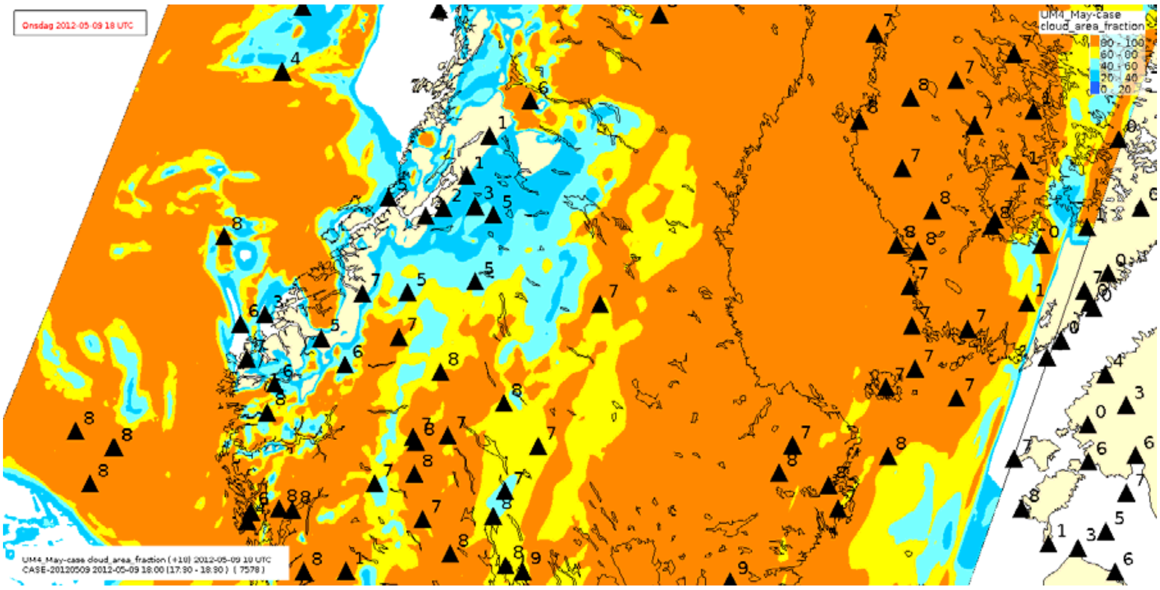


Figure 68: Snapshot of +18hr forecast of total cloud cover from UM4 and observations, valid for 09/05/12 18UTC.

4 Summary of test periods & score card.

The results are summarized in score cards inspired by the ECMWF score card shown in Newsletter 128 (2011). The score card in this study has different scores and station lists employed for the variables MSLP, 10 m wind speed, T2m, 12hr precipitation and total cloud cover. There are three score cards comparing AROME with ECMWF, G(M)05, and UM4, respectively and one “summary” score card giving the relative ranking of the AROME model amongst the other models as a number from 1 (the best model) to 4 (the worst model). The score cards are based on a subjective judgement of the relevant figures for the different test periods from the verification system.

Some members of the MetCoOp group have judged the skills of the models for the different scores shown in the figures and then filled in the score cards. Note that these score cards are not based on any significance test, but are subjectively judged by looking at the figures for the score shown for the four periods above. The combined result was stored in the cells in the tables using a sign to indicate the skill of AROME compared to the other models. The different experiments are rather short test periods (about two weeks) and many test results in the score-card are based on sub-samples of the observations used in the study, e.g. Norwegian coast. Standalone results may be vulnerable to random effects so one should not pay too much attention to standalone results but instead focus on the result as a whole.

Description of the tables:

Scores that have been used:

The scores are explained in detail in Chapter: 0.

RMSE= Root Mean Square Error

BIAS= systematic error

FB = Frequency BIAS (FB).

ETS= Equitable Threat Score
ETS10= ETS 10 m/s threshold, ETS15= ETS 15 m/s threshold,

ETS0.5= ETS 0.5 mm/12hr, ETS2.0= ETS 2.0 mm/12hr, ETS10.0 = ETS 10 mm/12hr.

The ETS is chosen as skill-score for the score cards. From a theoretical aspect, it is not necessarily the best of the scores described in 2.4, but it is used here since it is commonly used and thus assumed to be less unfamiliar to the readers of this paper. An advantage of this score is that it does not have any large tendency of favouring forecasts with a large positive or negative BIAS.

The different thresholds selected can be discussed, but the combination of the scores for each parameter is assumed to give a fairly complete picture of the skills of the models.

Symbols that have been used:

 indicates that AROME is clearly better than the compared model.

 indicates that AROME is better than the compared model.

 indicates that AROME is similar in quality to the compared model.

 indicates that AROME is worse than the compared model.

 indicates that AROME is clearly worse than the compared model.

4.1 Score card AROME versus ECMWF

AROME vs EC		**MetCoOp Domain**	**Norway**	**Sweden**	**Mountains**	**Coast**	**Inland 200– 600m**	**Comments**
MSLP	RMSE	▼	▼	▼	●	▼	▾	
10 m wind	RMSE	●	▾	▲	▴	▾	▾	1)
	FB	▲	▲	▲	▲	▲	▴	
	ETS10	▲	▲	▲	▲	▲	▲	
	ETS15	▲	▲	▲	▲	▲	▲	
Prec 12hr	BIAS	▲	▲	●	Few stations	▴	▲	2)
	FB	▲	▲	▴		▴	▲	
	ETS0.5	▲	▲	▲		▴	▲	
	ETS2.0	▴	▲	●		▴	●	
	ETS10.0	●	▾	●		▴	▾	
T2m	BIAS	▲	▲	▴	▾	▾	▲	
	RMSE	▴	▲	▾	▴	▾	●	
TCC	BIAS	▾	▾	▼	Few stations		●	
	FB	▼	▼	▾			▼	
	ETS	▼	▼	▾			▼	

1) Small differences, varies between periods. 2) High uncertainty in the observed amounts during winter.

Summary:
AROME is in general not as good as ECMWF on MSLP and total cloud cover (TCC), but is better than ECMWF on wind, precipitation (neutral on large precipitation amounts) and temperature. The conclusion for temperature is however dependent on region and type of verification score.

4.2 Score card AROME versus UM

AROME vs UM4		**MetCoOp Domain**	**Norway**	**Sweden**	**Mountains**	**Coast**	**Inland 200–600 m**	**Comments**
MSLP	RMSE	▲	▲	▲	▴	▲	▲	
10 m wind	RMSE	▴	●	▲	▲	▲	▼	1)
	FB	▴	▴	▴	▲	▾	▴	
	ETS10	▲	▲	▲	▲	▴	▲	
	ETS15	▴	▴	▴	▲	●	▲	
Prec 12hr	BIAS	▴	▲	▾	Few stations	●	▴	
	FB	▲	▲	▾		▲	▲	
	ETS0.5	▲	▲	●		▼	▲	
	ETS2.0	▼	▾	▼		▼	▾	
	ETS10.0	●	▾	▴		●	▼	
T2m	BIAS	▲	▲	▲	▲	▾	▲	
	RMSE	▲	▲	▲	▲	▲	▲	
TCC	BIAS	▼	●	▼	Few stations		▼	
	FB	▼	▼	●			▾	
	ETS	▼	▼	▼			▼	

1) largest difference in the march period before UM4 fix on wind speed was introduced.

Summary:
AROME is in general better than UM, but not on large precipitation amounts and total cloud cover (TCC).

4.3 Score card AROME versus HIRLAM

AROME vsG(M)05		**MetCoOp Domain**	**Norway**	**Sweden**	**Mountains**	**Coast**	**Inland 200–600 m**	**Comment**
MSLP	RMSE	▴	▴	▴	▴	▴	●	
10 m wind	RMSE	▾	▾	▲	▲	●	▼	
	FB	▲	▲	▲	▲	▲	▴	
	ETS10	▲	▲	▲	▲	▲	▲	
	ETS15	▲	▲	▲	▲	▴	▲	
Prec 12hr	BIAS	▲	▲	▴	Few stations	▲	▴	
	FB	▲	▲	▲		▲	▲	
	ETS0.5	▲	▲	▲		▾	▲	
	ETS2.0	▲	▲	▴		●	▲	
	ETS10.0	▲	▲	▴		●	▴	
T2m	BIAS	▼	▾	▼	▼	▾	▼	
	RMSE	▼	▾	▼	●	▼	▼	
TCC	BIAS	▼	▾	▼	Few stations		▼	
	FB	▼	▼	▼			▼	
	ETS	▼	▼	▼			▼	

Summary:
AROME is better than HIRLAM on MSLP, wind and precipitation. There are still some problems in AROME for the T2m forecast, but for this parameter HIRLAM is the clearly best model of the four models. HIRLAM is also better than AROME on the TCC, due to the problems that are seen in AROME for this weather parameter.

4.4 Ranking of AROME compared with ECMWF, UM and HIRLAM

A relative ranking of AROME (from 1 – the best to 4 the worst) compared with ECMWF, UM4 and G(M)05 (HIRLAM) for different scores and variable is given in the table below.

1: indicates that AROME shows better verification results than the three other models.

2: indicates that AROME shows better verification results than two other models and worse than one other model.

3: indicates that AROME shows better verification results than one other model, but worse than two other models.

4: indicates that AROME shows worse verification results than the three other models.

		MetCoOp	Norway	Sweden	Mountains	Coast	Inland 200–600 m	Comments
MSLP	RMSE	2	2	2	1-4	2	2-3	
10 m wind	RMSE	2-3	3-4	1	1	2-3	4	
	FB	1	1	1	1	2	1	
	ETS10	1	1	1	1	1	1	
	ETS15	1	1	1	1	1-2	1	
Prec 12hr	BIAS	1	1	1-4	Few stations	1	1	
	FB	1	1	2		1	1	
	ETS0.5	1	1	1-2		2-3	1	
	ETS2.0	2	2	3		3	2-3	
	ETS10.0	2	1-2	1-4		1-4	3-4	
T2m	BIAS	2	2	2	3	4	2	
	RMSE	2	2	3	1-2	2-3	2-3	
TCC	BIAS	4	3-4	4	Few stations		3-4	
	FB	4	4	4			4	
	ETS	4	4	4			4	

Summary:

HARMONIE AROME gives the best result of the four models on wind and precipitation (but not the largest amounts). AROME is the second best model choice for temperature and mean sea level pressure. Different models beat the HARMONIE AROME system on these parameters, so it is not one particular model which is better than AROME. AROME shows the worst result of the four models on total cloud cover.

Overall HARMONIE AROME seems like a good model choice in comparison with the HIRLAM, UM and ECMWF model systems.

4.5 Validity and reliability

In this verification study three months in total where included, and this is not enough to give significant conclusions for all weather types. However four different seasons with different weather types of special interest for SMHI and met.no are studied.

The verification of the model results for the different periods run in this study does not show same results, indicating weather dependent model quality. The different models show strengths and weaknesses in different situations.

The scorecard summarizes only the models performances in the four analyzed periods. Several of the summarised scores is in accordance with what is noted elsewhere (in example the HIRLAM community, met.no and SMHI), but a generalization of the results should be done with care.

HARMONIE performs in many aspects well, but reveals some deficiencies. There are also some issues that need further evaluation and investigation as listed below.

The models can in other setups (i.e. versions/domains/resolution) score differently when compared to each other. But the aim of this report is to compare HARMONIE with operational available models at met.no and SMHI. Comparison with other models and versions (i.e. ALARO and UM from MetOffice) will be performed and reported later on.

5 Twelve issues for further investigations and improvements

The results from the 4 periods constitute approximately three months of model simulations. This is not enough to draw firm conclusions but should give an indication of the behaviour of the AROME (HARMONIE), G(M)05 (HIRLAM), UM4 (Unified Model) and ECMWF-model. Differences found might be due to the relatively short periods, but might also give valuable insight into real differences between the models and further needs for development in the AROME model. As all experiments are done with version cycle 36h1.4 of AROME there is a possibility that revealed deficiencies already are improved in the existing version c37 (see https://HIRLAM.org/trac/wiki/ReleaseNotes/HARMONIE-37h1.1). With this in mind we list 12 issues that should be investigated further for a better understanding and for possible improvements of the AROME model. We have not prioritized the different issues, but obviously some of them are important for increasing the accuracy of forecasts within MetCoOp.

Issue 1) Differences between scores in Sweden and Norway

The relative ranking of the four models in terms of verification scores differ for many periods and parameters between Sweden and Norway. For cooperation between Norway and Sweden this should be explored further and better understood. One example is the RMSE for 10m wind speed; In Norway we rank AROME as equal to UM4, worse than ECMWF and far worse than G(M)05. However, in Sweden we rank AROME far better than all the three other models with respect to RMSE. For several of the precipitation measures we rank AROME as clearly better than ECMWF (BIAS and ETS for 2.0 mm/12hr) and UM4 (BIAS and ETS 0.5 mm/12hr) in Norway, while in Sweden we find the quality of the models quite equal. Additionally there are large differences in total cloud cover score and behaviour between Norway and Sweden which we discuss separately as issue 2.

Issue 2) Differences in cloud cover forecasts in Sweden and Norway

In particular the accuracy and characteristics of the cloud forecasts differ between Sweden and Norway. This difference is at least partly due to manual cloud observations in Norway and automatic measurements in Sweden which is not recommended to mix in model verification (WMO 2012 [9]). The consequences of the different characteristics of observations in Norway and Sweden should be investigated further.

Issue 3) Cloud cover

The cloud forecasts from AROME are poor compared with the other models and the Frequency BIAS (FB) reveals too many extreme occurrences (cloud free or totally cloudy). The cloud cover characteristics are elaborated further in a separate report. Low clouds and fog seems to occur too often in AROME, especially at night and in early morning. Fog may be reduced by assuming a somewhat higher sedimentation of cloud droplets at the lowest model level.

Issue 4) Coastal temperatures

The AROME temperatures in coastal areas show systematic errors that indicate an unrealistically high influence by the sea surface temperature. This is especially clear in spring. This behaviour is present in many models, but in this comparison most pronounced in AROME and may be masked in models with coarser resolution and a less pronounced land sea mask.

Issue 5) Extreme cold temperatures

All models show deficiencies in forecasting extreme cold temperatures. However, for the periods evaluated here this is most pronounced in AROME. Even though two periods with cold temperatures are included none of these are extremely cold and it is possible that the problem is more pronounced than revealed in our current four test periods. However, there may be issues regarding the cold start of AROME before the august and December/January test-periods. In C36h.1.4 a spin-up time is needed to obtain more correct values for the soil parameters. The handling of cold starts is improved in later cycles. To avoid problems connected to cold starts we omitted the first few days of the runs in the evaluation. A systematic improvement of the day-to-day score of the models was not seen, which suggests that this was not the problem.

The cold temperature problem is an issue that is already addressed by changes in the surface analysis with increased weight of T2m observations in AROME (Lindskog et.al [4]). The experiments indicate improvements, but the issue should be followed closely in experiments and pre-operational phase of MetCoOp.

Issue 6) MSLP and boundary frequency

The MSLP verification is in general better in ECMWF than in the nested fine scale models. There is in general a good correlation between the RMSE in ECMWF and in the fine scale models (i.e. the fine scale model errors are slightly larger than ECMWF and increase and decrease from day to day in a similar manner as ECMWF). However, on certain days a distinct peak is seen in AROME which is not in ECMWF (26/12, 27/12, 5/1, 9/3, 28/5). The 26–27/12 and 5/1 are the days of the extreme weather Dagmar, Cato and Emil and an hypothesis is that the regional models need more frequent lateral boundary updates than every 3rd hour when severe and rapidly evolving systems enter the integration domain. The hypothesis is further strengthen since a similar behaviour is found for GM05 (LBC coupling every 3rd hour), but to a smaller degree for UM (LBC coupling every hour from a HIRLAM 8km model). This should be investigated further.

Issue 7) Local wind storms

There are larger errors in the wind forecast from AROME for some of the largest cities in Sweden and Norway. The most severe case is a forecast for Trondheim in March 2012. The study of the Trondheim case reveals that the particular problem is not a city grid point problem, but a problem of heavily overestimating the wind speed over a larger area. Verification from Arlanda (large positive bias) and Landvetter (large negative bias) in March reveals that this is not necessarily a homogeneous problem for cities in general, but different problems for different locations with different explanations. However, such spurious peaks in error should be investigated further.

Issue 8) Wind speed and physiography

A clear underestimation of wind strength in AROME has been found for all periods for stations located between 200 and 600 m.a.s.l. This behaviour is also reported by the system manager of HARMONIE (Ulf Andrae), and should be investigated further. A hypothesis is that the AROME model underestimates the wind speed for forested areas.

Issue 9) Wind direction at the coast

The RMSE of wind direction at the coast is larger in the fine scale models than in the ECMWF model for three of the four periods (December/January, March and May) even though the bias is less pronounced in the fine scale models. This behaviour may be due to finer details and higher variability in the fine scale models, but higher resolution should in general give a better local description of the wind direction as seen on the inland stations.

Issue 10) Verification of high resolution precipitation

As illustrated with the plots of forecast and observed precipitation from the 15th August there are large spatial variations in observed precipitation amounts. The precipitation pattern from ECMWF is smoother, while the fine scale models show more spatial patterns but the maxima and minima are not necessarily in correct positions. Point verification, as done in this report, is therefore an advantage for models with smooth patterns and does not necessarily evaluate the total amount of information present in the models (for example there is no advantage for more realistic spatial variations). To evaluate the precipitation forecasts properly and explore the potential information they contain, spatial methods should be used.

Issue 11) Upper level verification

This report has focused on the surface fields and some properties of clouds. Additional investigations of features important for, for example, aviation meteorology should be performed and vertical profiles should be made available to duty forecasters in a similar manner as the surface parameters are today.

Issue 12) Detailed physiographic description

It is necessary to perform a quality check on the physiographic fields in the AROME model. Finer horizontal resolution requires higher accuracy in the description of the surface and especially if the model data is to be used for detailed local weather forecasts for the public.

6 References

[1] T. Driesenaar (2009) *General description of the HARMONIE model*
http://www.HIRLAM.org/index.php?option=com_content&view=article&id=65&Itemid=102

[2] J. Dudhia, S. Tibaldi, W. Zwieflhofer (2010) *Evaluation of the HIRLAM-A Programme.*
http://HIRLAM.org

[3] Kristiansen, J., Sørland, S.L., Iversen, T., Bjørge, D. and Køltzow, M. Ø. (2011).
High-resolution ensemble prediction of a polar low development.
Tellus 63A, 585-604.

[4] M. Lindskog, L. Bengtsson, U. Andrae, B. Stensen and K.-I. Ivarsson (2012) *IMPROVED HARMONIE SURFACE TEMPERATURE FORECASTS FOR NORDIC WINTER CONDITIONS.* Poster at All Staff Meeting, HIRLAM 2012,
http://www.cnrm.meteo.fr/aladin/spip.php?article237

[5] P. J. Roebber, Schultz, Colle, Stensrud (2004) *FORECASTER'S FORUM Toward Improved Prediction: High-Resolution and Ensemble Modeling Systems in Operations.* WEATHER AND FORECASTING, Volum 19

[6] Y. Seity, P. Brousseau, S.Malardel, G. Hello, P. Be ´ nard, F. Bouttier, C. Lac, and V. Masson . (2010). *The AROME-France Convective-Scale Operational Model*. MONTHLY WEATHER REVIEW, Volum 139.

[7] S. H. van der Veen (2012) *MSG cloud mask initialisation in hydrostatic and non-hydrostatic NWP Models.* Presentation at EMS (EMS2012-228) (Will be published. Will be implemented in HARMONIE)

[8] P. Undén et. al. (2002) *HIRLAM-5 Scientific Documentation*,
http://HIRLAM.org/index.php?option=com_docman&task=doc_download&gid=308&Itemid=70

[9] WMO, WWRP 2012 – 1, Recommended Methods for Evaluating Cloud and Related Parameters. WWRP/WGNE Joint Working Group on Forecast Verification Research (JWGFVR).
http://www.wmo.int/pages/prog/arep/wwrp/new/documents/WWRP_2012_1_web.pdf

7 List of figures

Norwegian
Meteorological Institute
met.no

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

SMHI

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

ISSN 1893-7519