Improving NWP Model Cloud Forecasts Using Meteosat Second-Generation Imagery

SIEBE H. VAN DER Veen
Royal Netherlands Meteorological Institute, De Bilt, Netherlands

(Manuscript received 24 January 2012, in final form 27 August 2012)

ABSTRACT

The cloud mask of the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) is a nowcasting Satellite Application Facility (SAF) that is used to improve initial cloudiness in the High-Resolution Limited-Area Model (HIRLAM). This cloud mask is based on images from the Meteorological Satellite (Meteosat) Second Generation (MSG) satellite. The quality of the SAF cloud mask appeared to be better than initial HIRLAM clouds in 84% of the cases. Forecasts have been performed for about a week in each of the four seasons during 2009 and 2010. Better initial clouds in HIRLAM always lead to better cloud predictions. Verification of forecasts showed that the positive impact is still present after 24 h in 59% of the cases. This is remarkable, because initial dynamics was kept unchanged. The magnitude of the positive impact on cloud predictions is more or less proportional to the initial cloud improvement, and it decreases with forecast length. Also, forecast 2-m temperatures are affected by initial clouds. The generally positive bias of the 2-m temperature errors becomes a few tenths of a degree larger during the night but it decreases a comparable amount during daylight, because MSG tends to increase the cloud amounts in HIRLAM. The standard deviation of the errors often improves slightly in the first part of the forecast, indicating that forecast temperatures correlate better with observations when MSG is used for initialization. For longer lead times, however, standard deviations deteriorate a few tenths of a degree in seven of the eight verification periods, which all had a length of about a week.

1. Introduction

In the past, various attempts have been made to assimilate or initialize clouds in numerical weather prediction (NWP) models. Bayler et al. (2000) changed cloud water mixing ratios such that the model clouds better match the Geostationary Operational Environmental Satellite (GOES) clouds, and they see positive impact on cloud forecasts for the first 6 h. Precipitation forecasts had improved as well. Yucel et al. (2002) have inserted clouds in a nonhydrostatic NWP model by introducing them directly in the model. Cloud forecasts improved in the first 4 h. Yucel et al. (2003) have inserted clouds in the fifth-generation Pennsylvania State University–National Center for Atmospheric Research (PSU–NCAR) Mesoscale Model (MM5) by a nudging assimilation technique. Cloud-top heights were derived from GOES infrared (IR) images, and cloud-base heights were inferred including GOES visible (VIS) images. Also the vertical structure of the model clouds before ingestion was used. They find that predictability of cloud cover and precipitation is a little enhanced for the first 3 h of the forecast. Also, diagnosis of surface radiation had improved. Vukicevic et al. (2004) have applied four-dimensional variational data assimilation (4DVAR) in the analysis of a nonhydrostatic cloud resolving model using IR and VIS images from the GOES-9 satellite, and they were successful in better defining a low-level stratus deck. The mean error in brightness temperatures of clouds in the analysis clearly improved; also +3-h forecasts improved, but they had a lower quality than the analysis. Vukicevic et al. (2006) showed that the short-term evolution of ice cloud mass can be clearly improved using GOES IR images in 4DVAR. Benedetti and Janisková (2008) have used Moderate Resolution Imaging Spectroradiometer (MODIS) cloud optical depth in a 4DVAR analysis of the European Centre for Medium-Range Weather Forecasts (ECMWF) model. This analysis includes moist physics schemes. The improvement of analyses and forecasts is small, yet the authors consider the procedure promising.
et al. (2010) have assimilated precipitation, and also clouds (stratocumulus) using cloud tops from the Meteorological Satellite (Meteosat) Second Generation (MSG) satellite and synoptic observations from cloud base. Recently, data assimilation procedures other than the three-dimensional variational data assimilation (3DVAR) or 4DVAR have also been applied in NWP models using “imagery from geostationary satellites,” albeit that both “observations” and the “truth” were then generated by the same or different NWP models. For example, Otkin (2010) has used an ensemble Kalman filter for an hourly assimilation of IR brightness temperatures in the Weather Research and Forecasting Model (WRF). Analyses showed better cloudy brightness temperatures and also better vertical profiles of cloud condensate. Another example is Zupanski et al. (2011) who applied a “maximum likelihood ensemble filter” in WRF. They found that assimilation of IR images from the future (to be launched) GOES-R satellite indicated that especially cloud ice and snow will have the largest positive impact on the quality of the cloud analyses. Furthermore, Otkin (2012) has assimilated IR radiances in WRF applying an ensemble Kalman filter and artificial observations that included geostationary IR images. Though not all model fields had improved, results showed a positive impact on brightness temperatures of clouds in the analyses; in “ensemble mean” forecasts this positive impact lasted for some 5 h. Improved liquid water and ice water paths lasted for only 1 h though.

It is generally believed that improving the forecasting of clouds requires advanced data assimilation techniques (see e.g., Auligné et al. 2011), and it is often stated (e.g., Yucel et al. 2003), that the rather weak and short-lasting impact of improved initial cloudiness in NWP models must be caused by a mismatch of the dynamic fields in the model and the newly introduced clouds. This may be true sometimes, but in this paper it is demonstrated that changing initial clouds can have a positive impact that lasts for at least a day, also when initial dynamic fields are left unchanged. Such a long impact on cloud-cover forecasts was not seen in earlier studies.

2. Cloud initialization in HIRLAM

a. Quality of SAF cloud mask

Initialization of clouds in a NWP model by applying satellite observations is ideally based on geostationary satellites, because a high time resolution is preferred when cycling the model in a production mode. A cloud mask based on a geostationary satellite is described in Feijt et al. (2000); this cloud detection system uses the “old Meteosat” (Meteosat-7 or earlier versions), see the website of the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT; http://www.eumetsat.int). In this study, however, we apply the newer nowcasting Satellite Application Facility (SAF) cloud mask provided by EUMETSAT, which is based on MSG. This cloud mask is described in Derrien and Le Gleau (2005), and it is based on a multispectral thresholding technique. For cloud detection dynamic thresholds are applied to different combinations of channels. Also geographical information such as surface height and land–sea mask is needed, and temperatures and water vapor content from NWP models are used for determining some thresholds. All 12 MSG channels are used during the day, but only the IR channels are used at night.

Before looking at the quality of cloud forecasts, the quality of this MSG “cloud mask” is investigated at 0000 and 1200 UTC. In Figs. 1a,b the bias and standard deviation of the errors are shown for the following four periods:

- Spring: 0000 UTC 2 April–1200 UTC 10 April 2010.
- Summer: 0000 UTC 22 July–1200 UTC 29 July 2010.

The errors are computed by comparing observations of synoptic (total) cloud cover to the amount of cloud cover in a rectangular “cloud mask area” in which 3 by 5 pixels are sampled such that the central pixel covers the synoptic station. The rectangular area comprises five pixels in the east–west direction and three pixels in the north–south direction. Over the Netherlands the pixel size is about 3.7 km (east–west) by 6.4 km (north–south). The amount of cloud cover is defined as the number of cloudy pixels in the area divided by 15. For verification only synoptic stations between 45° and 60°N and between 10°W and 20°E have been used. See Fig. 2 for an overview of the locations of the synoptic stations. The synoptic observations used for the verification are both by human observers (HO) and by cloud sensors (CS). The CS observations prevail in most countries, but the HO is still contributing for a substantial part in the verification area. See Boers et al. (2010) for more information on the accuracy of both these observations. The HO and CS observations have different accuracy characteristics. The quality of HO cloud observations is subject to a diurnal cycle: at night more clouds will be missed than during daytime. The accuracy is estimated to be around 1 okta, except for complete overcast or...
completely clear skies during daylight, then the accuracy will be higher. The HO is supposed to watch all clouds in the full hemisphere. The CS usually only looks at zenith, and time series of observations are translated to “hemispheric” observations. Thus, CS observations overestimate zero and eight okta occurrences, especially in case of slow cloud advection. Unlike HO observations, the quality of CS observations is not subject to a diurnal cycle. The estimated accuracy of CS observations is somewhat above 1 okta.

The integer numbers on the x axes in Figs. 1a,b refer to 1200 UTC, and the half numbers to 0000 UTC. Figure 1a shows that in the months of April, July, and October the SAF cloud detection slightly overestimates cloudiness. The bias in January is very close to zero, both during daylight and night time, and it shows a diurnal cycle and is highest during daylight. The standard deviation of the errors in January is also rather small with values between 1.5 and 2 oktas, except on one night (see Fig. 1b). The diurnal cycle shows anticorrelation between mean errors and standard deviation of the errors (cf. Figs. 1a,b). This phenomenon is also seen in the other months: lower biases (which are closer to zero) correspond often to higher standard deviations of the errors. It appeared also that the rms errors of the SAF clouds were larger at night than during daylight (not shown here), despite the smaller bias of the errors at night. In Fig. 1c domain-averaged synoptic cloud cover is shown for the four periods as well. It seems that higher cloud cover is correlated with smaller standard deviation of the cloud detection errors, but this is not always the case. We may conclude that generally the cloud mask performs well, especially during daylight. The (on the average) slightly positive bias of the cloud mask from the nowcasting SAF is in agreement with findings from Derrien and Le Gleau (2005), who showed that detection of clouds in clear areas happens more often than missing clouds in cloudy areas.

### b. Transfer of MSG clouds to HIRLAM clouds

The High-Resolution Limited-Area Model (HIRLAM) is run with data assimilation cycles in a configuration as
described in section 3a. In both the reference runs and the “MSG-synop” (in the following part of the paper referred to as “MS”) runs clouds are initialized using 4DVAR. In this assimilation humidity in the new analysis is partly determined by prior information from the previous run, and partly by new observations such as radiosondes, synoptic observations, and Advanced Microwave Sounding Unit (AMSU) data. These latter observations provide vertical profiles of humidity and temperature.

After making the analysis digital filtering is applied in order to balance wind fields and pressure fields such that generation of gravity waves is avoided. In the MS run, clouds are further improved after this analysis and filtering, in a next step.

In this next step it is aimed to convert the 2D cloud mask to 3D clouds in the hydrostatic limited area model HIRLAM [see Undén et al. (2002) and the HIRLAM website (http://hirlam.org)]. This is possible when cloud-top heights and cloud-base heights are known. A constant amount of cloud cover is assumed in each vertical column within a cloud. The amount of cloud cover derived from the MSG cloud mask is computed from the 15 pixel rectangular area as described in the previous section. Cloud-top heights are assigned to the highest HIRLAM level where the temperature equals the temperature measured by MSG with the 10.8-μm channel. Cloud-base heights are derived from synoptic observations. As these lack at most model grid points, cloud-base heights are horizontally interpolated, using all synoptic observations with a weighting factor proportional to the minus sixth power of the distance. In case of more than one cloud layer in the synoptic observations, the height of the lowest layer is taken, except when this layer has a cloud cover of 2 oktas or less and the layer above has a cover of 3 oktas or more; then the cloud-base height of the second lowest layer is taken as the cloud base for initialization. Thus, only one initial cloud base is defined. As a consequence, cloudiness will often be overestimated in case more than one cloud layer is present in reality.

In HIRLAM 7.3 the cloud cover \( N \) is primarily a function of the relative humidity \( (q_m/q_s) \). Apart from this it also weakly depends on temperature \( T \). The \( q_s \) is the saturation specific humidity and \( q_m \) is the water vapor specific humidity. We simplified the relation by only taking into account the dependence on relative humidity, such that the relation between \( N \) and \( q_m \) is a reasonable approximation for the one valid in the Kain–Fritsch version of HIRLAM, see Eq. (1):

\[
q_m = q_s \left( 1 - C \right)^\sqrt{N} + C \quad (N > 0) \quad \text{and} \\
q_m = \min(q_m, C q_s) \quad (N = 0),
\]  

(1)

with

\[
C = \frac{\text{rh}_{\text{max}} - (\text{rh}_{\text{max}} - \text{rh}_{\text{min}}) \sin \left( \frac{\pi p}{\text{Ps}} \right)}{\text{rh}_{\text{max}} - \text{rh}_{\text{min}}},
\]  

(2)

In Eq. (2), \( p \) is the pressure and \( \text{Ps} \) is the surface pressure. This formula expresses that the critical relative humidity \( C \) above which clouds can start to form is lowest in the midtroposphere \( (\text{rh}_{\text{min}}) \), and highest near the earth’s surface \( (\text{rh}_{\text{max}}) \) and near the tropopause. The lowest value at around 500 mb reflects the relatively high value of moisture variability at this level (see e.g., Molod 2011). For convective cloud cover in HIRLAM additional expressions apply; these were not taken into account. Convective clouds are not simply depending on humidity, but primarily on updraft mass fluxes, which are determined by convective available potential energy (CAPE). Convection can be triggered by temperature perturbations which depend on, for example, variance in vertical velocities, but also boundary layer turbulence can trigger convection. See Kain (2004) for a description of the convection scheme applied in HIRLAM. It is expected that the current relatively simple procedure for creation or removal of clouds can probably also improve the amount of convective cloud cover, because CAPE will often increase if humidity is added, or decrease if it is made dryer.

Boundary layer clouds are not treated differently during the initialization procedure. In HIRLAM these clouds are assumed to exist in an environment with relatively low moisture variability due to mixing. This implies a rather high critical relative humidity above which subgrid clouds can form. This phenomenon is accounted for because \( C \) in Eq. (2) has a relative high value near the earth’s surface. See Undén et al. (2002) and Cuxart et al. (2000) for a description of the turbulence scheme used in HIRLAM.

Cloud initialization is performed using Eq. (1) and Eq. (2), and \( N \) is defined to be zero for \( T \) colder than at cloud top and below the (interpolated) cloud base. After this cloud initialization, which is executed in HIRLAM after the analysis and model initialization, 3D temperatures are adjusted such that the virtual temperature \( T_v \) remains unchanged. This virtual temperature is computed as \( T_v = T(1 + 0.61 q_m - q_l) \) with \( q_l \) the liquid water specific humidity [see also Monteiro and Torlaschi (2007)]. Keeping the virtual temperature constant ensures that hydrostatic dynamics in the model is not influenced initially, though later on this may still happen. A side effect of keeping the virtual temperature constant is that \( q_l \) slightly changes, implying that \( q_m \) should be computed again, given a prescribed \( N \) from Meteosat. An iterative loop of five or six repeats is usually
sufficient to reach convergence to a state where $q_m$ corresponds to exactly the right cloud cover from Meteosat, without having changed the initial virtual temperature $T_v$. One experiment carried out without this "temperature correction" (not shown here) exhibited slightly worse verification results for the forecast cloud cover, indicating that keeping the virtual temperature constant does make sense.

c. Quality of initial clouds

In Fig. 3 the errors in the MSG cloud mask and in the initial HIRLAM cloudiness for both the +0-h forecast of the "reference run" and the +0-h forecast of the "MS run" are displayed for the four periods defined in section 2a. In the four corresponding panels of the figure errors of initial clouds at 0000 and 1200 UTC analysis times are pictured consecutively in each curve.

The four different panels in Fig. 3 show that the bias of errors of the nowcasting SAF clouds is generally between 0 and 1 okta, except in January (top left panel) where it is quite close to zero. The HIRLAM reference runs have on most days quite a negative bias at initialization time, especially during the day. A clear diurnal cycle in the bias is always present, except during a few days in January. The average bias of the MS runs is around zero, though in July it is a little higher. The bias of the MS runs exhibits the same diurnal cycle as that of the reference runs (except in January), but the amplitude is smaller. The standard deviation of the initial errors of HIRLAM reference runs is nearly always larger than the standard deviation of the errors of the SAF cloud mask. This is quite obvious in January (top left panel), but it is also true on most days in the other periods. The standard deviation of initial errors in MS runs is often a little higher than the standard deviation of errors in the SAF cloud mask, but on most days it is still smaller than the standard deviation of initial errors in the HIRLAM reference runs.

An interesting conclusion that can be drawn from inspection of all the initial errors in Fig. 3 is that the bias of initial HIRLAM clouds often shows a diurnal cycle with maximum values during the night. The initial bias in MS runs exhibits smaller amplitudes, however, possibly because the SAF clouds also have a diurnal cycle in

![Fig. 3. Bias and standard deviation of errors in initial clouds, both at 0000 and 1200 UTC. (top left) January, (top right) April, (bottom left) July, and (bottom right) October. Color scheme: bias reference HIRLAM: gray, standard deviation reference HIRLAM: black, bias MS HIRLAM: dark blue, standard deviation MS HIRLAM: cyan, bias SAF cloud mask: orange, and standard deviation SAF cloud mask: purple. The large tick marks on the x axes refer to 1200 UTC, the small tick marks refer to 0000 UTC.]
the bias, but this is anticorrelated with the initial HIRLAM runs. Maybe synoptic observations miss more clouds at night, which could (partly) explain the more positive bias of initial HIRLAM clouds. Verification results of cloud forecasts indicate this may indeed be the case (e.g., see Fig. 9a). In this figure it is obvious that biases are higher for forecasts valid at times during the night (beginning and end of the forecast) than for forecasts valid during daylight. This would also mean that the diurnal cycle in the bias of the SAF clouds is in fact even larger than shown in Fig. 1a. In addition it might be that initial HIRLAM runs contain too few convective clouds during daylight. The relatively large amplitude of the initial biases in the HIRLAM reference runs in July (see bottom left panel in Fig. 3), would confirm such a conclusion.

The results presented in this section lead to the conclusion that the SAF cloud mask is nearly always able to improve initial cloudiness in HIRLAM to the quality level that the SAF clouds have themselves. This is certainly true for the bias, and in most cases also for the standard deviation (and the rms) of the errors.

3. Cloud forecasts

a. Setup of numerical experiments

Both reference runs and runs with MS cloud initialization have been performed for the four periods defined in section 2a. The cycling frequency was once every 6 h, but the model runs used MS clouds only twice a day (i.e., at 0000 and 1200 UTC). The MS initialization is not applied in case the SAF clouds had a lower quality than initial clouds in the reference run. This happened in a few cases.

The MS clouds were inserted in HIRLAM (version 7.3) after the 4DVAR analysis and initialization by changing the specific humidity (see section 2b for more details). The model resolution is 22 km and the number of grid points is 406 × 324. The domain covers the whole of Europe, the North Atlantic Ocean, and the northeast of Canada (see orange area in Fig. 4). This area is considerably larger than the area where Meteosat clouds have been inserted, see the blue area in Fig. 4. Outside this blue area initial HIRLAM clouds are always left unchanged. Figure 5 illustrates the forecast cloud cover with HIRLAM over a central part of western Europe, with (Figs. 5b,e) and without (Figs. 5a,d) MS initialization. The +1- and +12-h forecasts are shown. Black areas refer to clear sky, and (lighter) gray shades to (more) clouds. It is remarkable to see that even after 12 h (Figs. 5d,e) the influence of MS cloud initialization is still clearly visible. See the gray circles for an indication where the difference between the runs is most apparent. In Figs. 5c,f the SAF cloud masks are displayed for comparison with the HIRLAM forecasts. It is clearly seen that the +1-h forecast with MS initialization resembles this cloud mask a lot better than the +1-h forecast of the reference run. For the +12-h forecasts it is less obvious which is best, though the increased cloudiness in the MS forecast is better at matching the SAF cloud mask.

b. Verification of cloud forecasts

We will now investigate the influence of better initial clouds in HIRLAM on the cloud forecasts. Verification is performed by comparison with synoptic observations for the same area as mentioned in section 2a.

In Figs. 6a,b the bias and standard deviation of the errors of cloud forecasts are shown for the verification week in January. Figure 6a shows the errors of all HIRLAM runs that started at 0000 UTC. This picture indicates that the reference run has a strong negative bias between −1 and −2 oktas during the whole forecast. The MS run improves this considerably, though it is becoming more and more negative in the course of time. The figure also shows the standard deviation of the errors for MS and reference runs. It is seen that the MS run has a much smaller standard deviation in the errors than the reference run, indicating that MSG and synoptic observations insert valuable cloud information in the HIRLAM model. The difference with the reference run is diminishing in the course of time, but it remains during the whole forecast. The differences between rms errors are even larger (not shown here), obviously because both bias and standard deviation of the errors have improved due to MS initialization.

In Fig. 6b the same results are visible as in Fig. 6a, except that now all runs have a start time at 1200 UTC. Similar conclusions apply for these “1200 UTC runs,” except that most errors are somewhat smaller. The smaller errors of the MS run are probably due to the higher quality of the cloud mask from the nowcasting SAF during daylight, as is shown in this paper and also by Derrien and Le Gléau (2005).
Fig. 5. (a) Total cloud cover at 1200 UTC + 1 h 16 Jan 2010 (HIRLAM REF). (b) Total cloud cover at 1200 UTC + 1 h 16 Jan 2010 (HIRLAM MS). (c) Total cloud cover at 1300 UTC 16 Jan 2010 (SAF cloud mask). (d) Total cloud cover at 1200 UTC + 12 h 16/17 Jan 2010 (HIRLAM REF). (e) Total cloud cover at 1200 UTC + 12 h 16/17 Jan 2010 (HIRLAM MS). (f) Total cloud cover at 0000 UTC 17 Jan 2010 (SAF cloud mask). (a)–(f) See color bar for cloud cover in oktas.
In Fig. 7a the errors of forecasts that started at 0000 UTC in April 2010 are shown. Again standard deviation and bias of the errors are shown. The MS initialization was able to improve the standard deviation of the cloudiness errors for lead times up to 17 h, after that the impact is neutral. The bias of the errors of the MS run is somewhat improved, except at the beginning and the end of the forecast, where it is too high. Forecasts that start at 1200 UTC (Fig. 7b) show an amelioration of the clouds for the whole forecast. The bias of the error is always better for the MS run, and the standard deviation is slightly better or neutral.

Figures 8a,b show errors for July 2010. For both starting times MS runs have a bias close to zero, while reference runs predict generally not enough clouds, especially those starting at 1200 UTC. For runs starting at 0000 UTC the improvement of initial cloudiness is modest, but the standard deviation of the errors is always smaller (sometimes neutral) than in the reference run (see Fig. 8a). It is furthermore shown in Fig. 8b that MS runs starting at 1200 UTC always have smaller standard deviations than the reference runs for the full length of the 24-h forecasts.

Finally we have a look at runs that start in October (Fig. 9a for 0000 UTC, Fig. 9b for 1200 UTC). In Fig. 9a the standard deviation of the error is slightly better than in the reference run, except for the +4- until +8-h forecasts. The bias of the MS run is better than or equal to that of the reference run, except for the first 2 h. For the 1200 UTC runs the bias of the MS run is always better. The standard deviation is better until the +15-h forecast, and again at the end of the forecast.

4. Impact of improved cloudiness on 2-m temperatures

It is expected that improving initial cloudiness will not only impact forecast cloudiness, but also 2-m temperatures, convection, and precipitation. In addition to cloudiness we have verified the 2-m temperatures in the Netherlands at 28 different stations. These temperatures
are assumed to be affected by radiation changes due to the changed cloudiness. In Figs. 10a,b the impact of improved cloudiness on 2-m temperatures is shown for January. Both bias and standard deviation of the errors are indicated. For all reference runs in Figs. 10a,b temperature biases are positive. The bias becomes even larger in the MS run that starts at 0000 UTC, but in the 1200 UTC run it becomes better (lower) for the first 4 h. The reason for this is probably that increased cloudiness due to MSG tends to lower temperatures during daylight, but at night temperatures are increased because of stronger IR radiation sent downward by the clouds. The standard deviation of the errors decreases for both MS runs: it is better during the first 16 h in the 0000 UTC run and it is better during the first 18 h in the 1200 UTC run. These results indicate that patterns of forecast temperatures better match patterns of observed temperatures in the case of the MS cloud initialization. But the forecast temperatures are generally too high, and augmented cloudiness at night nearly always increases the already positive temperature bias. During daytime the positive bias of the temperature is often reduced, again because of the (on the average) increased cloudiness.

In Fig. 11a errors are shown for runs that start at 0000 UTC in April 2010. Reference runs and MS runs have nearly equal standard deviations of the errors, but the bias of the 2-m temperature errors improves for longer lead times of the MS run, probably because more clouds have a cooling effect during daylight. Runs starting at 1200 UTC (Fig. 11b) show a somewhat stronger influence of MS clouds, but the bias has only improved for the shorter forecast lengths and at the end of the forecast. For lead times in between the bias becomes more positive than in the reference run, probably because more clouds warm the earth’s surface at night. The standard deviation of the errors is initially unaffected, but MS clouds worsen these errors a little for longer lead times.

Furthermore, look at Figs. 12a,b where verification results for July 2010 are pictured. Biases have clearly decreased during daylight and have increased during the night. This is consistent with the higher average cloudiness in July due to MS initialization, as shown in Figs. 8a,b.
For the 0000 UTC runs the MS biases are better than those of the reference runs. Standard deviations of the temperature are more or less equal for MS runs and reference runs, with some small deviations.

In October 2009 the influence of MS clouds on the bias of the temperatures is quite small for runs that start at 0000 UTC (see Fig. 13a). The standard deviation slightly improves when MS clouds are used in HIRLAM. In Fig. 13b the bias of MS runs (starting times at 1200 UTC) becomes larger during the night, like we have seen in previous runs. Standard deviations of forecast 2-m temperatures slightly degrade for MS runs that start at 1200 UTC, but only for longer forecast times.

From the results presented in this section we conclude that the standard deviation of errors in 2-m temperatures often improves slightly during the first part of the forecast, in January 2010 even the first 16 and 18 h. The reason for this is probably that in this winter period MS initialization had improved the cloud forecasts considerably more than in the other seasons. A pattern often seen is that MS clouds cause the 2-m temperature to become lower during the day and higher at night. This often makes biases worse at night and better during daylight. Changes are in the order of a few tenths of a degree.

5. Sensitivity analysis for different initialization options

From the experiments described in this paper it appeared that initial cloudiness was quite often improved by MS initialization. It was investigated whether a limited change of initial cloudiness could be beneficial as well, because the cloud initialization procedures presented in this paper tend to overestimate initial cloud amounts in cases where more than one cloud layer is present in the atmosphere. For this purpose three different experiments were set up for the week in January: 1) a “reference” experiment in which MS initialization changes the HIRLAM clouds as described in section 2b (for Kain–Fritsch); 2) the same experiment, but now
with the restriction that the specific humidity $q_m$ is not allowed to change more than 30% compared to a run without MS initialization, and 3) an experiment like 2) but now changes are limited to 10%. In Fig. 14 the results of these three experiments are depicted for the standard deviation of the errors of forecast cloud cover. It can be concluded that the effect of both limits is to deteriorate the forecasts in the first 4–10 h. But for longer lead times (11–24 h) the limits for initial $q_m$ changes appear beneficial for the forecast, especially the 30% limit. The 10% limit (gray line) gives deterioration or suboptimal improvement for all forecast lengths. Averaged over all forecast lengths, the 30% limit appears to be the best choice. After 24 h the three curves have converged again, probably because the cloud initialization has largely lost its impact.

It was also investigated in what way the magnitude of improvements in initial cloudiness influences the (average) decrease of rms errors during the cloud forecasts. It appeared that decreases in forecast errors were more or less equal to the initial cloud improvement during the first few hours of the forecast. For longer lead times the magnitude of improvements became smaller, at a rate decreasing with forecast length. The period after which the initial improvement was halved varied between 5 and 9 h. This period is supposed to be also dependent on model resolution because numerical diffusion, which is partly responsible for the fading of the influence of initial cloudiness, depends on grid size.

6. Conclusions

In this paper it is demonstrated that the cloud mask of the nowcasting SAF for MSG is in most cases (about 84%) better than initial clouds in the HIRLAM model. This result is obtained for the Kain–Fritsch version of HIRLAM. Verification has been carried out using synoptic observations, both “man made” and “automatic.” Approximately one week was verified in each of the four seasons. The magnitude of the improvements was quite variable, but nearly always positive, and it appeared to be largest in the verification period in January 2010.
A general feature seen is that the quality of the nowcasting cloud mask is higher during daylight than at night. Biases in the cloud mask errors appeared to be lower (often closer to zero) at night than during daytime.

When the SAF cloud mask is introduced in HIRLAM, together with synoptic cloud-base heights and MSG cloud-top temperatures, the quality of the initial HIRLAM clouds is improved to the level of the SAF cloud mask. As a consequence the cloud forecasts ameliorate accordingly.

It was surprising to see that nearly all initial HIRLAM clouds showed a diurnal cycle in the bias of the errors (i.e., during daylight this bias was lower than at night). This cycle is anticorrelated with that of the SAF clouds. The diurnal cycle in the initial HIRLAM cloud amounts could be (partly) explained by the possibility that synoptic observations at night miss more clouds than during daylight. If this is true then the diurnal cycle in the SAF cloud amounts would be even stronger than shown in this paper. Apart from this the HIRLAM reference runs contained often too few clouds initially (and further in the forecast). This negative bias is also seen in different studies (see e.g., Nielsen 2009). The negative bias largely disappeared when MSG clouds were inserted.

We have performed forecasts with HIRLAM with a lead time of 24 h, with and without MS cloud initialization, and we have verified the runs for which MS in HIRLAM was initially better than the reference HIRLAM. In all experiments runs had a cycle frequency of once every 6 h, but MS clouds were only introduced at 0000 and 1200 UTC. It is shown in a little more than 60 runs with the Kain–Fritsch version of HIRLAM that improving the initial cloudiness always leads to improved cloud-cover forecasts. In 59% of the forecasts the domain-averaged impact appeared still positive after 24 h, when looking at rms errors. This result is important because earlier studies never found such long impact, and it indicates that changing initial dynamics may be less important for improving cloud cover than was previously thought.

It is likely that the quality of initial cloudiness in HIRLAM could be further improved, because adjusted water vapor fields imposed by the MS clouds were only a reasonable approximation and they could be further refined. Furthermore, additional observations could be used, for example, multiple cloud layers from synoptic observations, other SAF products like cloud type, vertically integrated liquid water paths, and GPS information on humidity. It would be worthwhile to try to improve initial cloudiness further and investigate to what extent forecasts would then improve as well.

For practical use of cloud initialization it is important to note that the quality of initial clouds, which determines the quality of the forecast, is in principle already known at analysis time. This quality can be assessed by comparison to synoptic observations. This implies that it is possible to skip cloud initialization during the cycling of an operational model, if initial clouds would appear to have a too low quality. A general conclusion is that MSG cloud insertion in HIRLAM can be a powerful tool in cloud nowcasting.

We have also verified the 2-m temperature forecasts for all the HIRLAM runs we have performed, but the verification area was now confined to the Netherlands. In the reference runs (without MS clouds) the forecast temperatures were too high in most cases (positive bias in the order of 1 K). The MS initialization often increased the cloudiness, as a consequence forecast temperatures became even higher at night (which was worse), but became lower during daylight (which was often an improvement). In addition it was seen that the standard deviation of errors in forecast 2-m temperatures remained the same or improved in the first part of the forecast, sometimes until the first 18 h. But at the end of the forecast they deteriorated somewhat. This deterioration may be related to a too strong reaction of the radiation module in HIRLAM to the model clouds, but this should be investigated further. It may also be that initial cloud depths are sometimes too large because synoptic cloud-base and cloud-top height from MSG are combined to form one cloud layer, also when clear layers are present in between in the real atmosphere. Furthermore, it could be that initial cloud tops are sometimes too high in case only boundary layer clouds are present in reality. Boundary layer clouds have often a cloud top right under the inversion, as is pointed out in Golding (1998). Under such circumstances it is likely that at some height above the inversion the same temperature is present as at the boundary layer cloud top. This leads then to a too high cloud top using the
initialization procedures described in this paper. It is considered necessary to solve the “bias problem” of the 2-m temperatures before the MS initialization in HIRLAM would be suited to be implemented in an operational mode.

In the study presented here the insertion of MS clouds in HIRLAM was only performed twice a day. It is anticipated that the influence of MS clouds could be further optimized if clouds were added hourly. This is confirmed by recent experimentation in a Rapid Update Cycle with HIRLAM [see also Benjamin et al. (2004)].

Improving initial cloudiness and humidity will also have an impact on stratiform precipitation and the intensity and triggering of convection. These subjects will be part of a follow-up paper. Apart from this, it may be useful to integrate the MSG cloud mask in 3DVAR or 4DVAR in HIRLAM, possibly still in combination with overwriting the original HIRLAM clouds with MS clouds after the analysis, as was done in this study. An advantage could then be that not only humidity fields but also dynamic and pressure fields are improved.

It is also intended to study the application of the MSG cloud mask in the nonhydrostatic HIRLAM–Aire Limitée Adaptation Dynamique Développement International (ALADIN) Research on Mesoscale Operational NWP in Europe and Mediterranean area (EuroMed) (HARMONIE) model. See the HIRLAM website for more information on this model (http://hirlam.org). It is expected that insertion of clouds in HARMONIE could improve the model’s ability to warn for severe weather occurrences, as these are often connected to deep convection, which in its turn can be triggered by the presence of sufficient humidity (clouds and water vapor) in the model atmosphere. A further step in HARMONIE could be to insert not only clouds, but also precipitation particles. Ideally this should be done with dual-polarized radars, which can differentiate between the various types of hydrometeors.

Acknowledgments. The author thanks EUMETSAT for the use of the nowcasting SAF and MSG data.

REFERENCES


Nielsen, K. P., 2009: Verification of Hirlam/Aladin cloud forecasts with MSG cloud physical products. Oral presentation, HIRLAM All Staff Meeting, Utrecht, Netherlands, HIRLAM. [Available online at http://hirlam.org.]


