Seventh Framework Programme

Theme 6 [SPACE]

Project: 607193 UERRA

Full project title:

Uncertainties in **E**nsembles of **R**egional **R**e-**A**nalyses

Deliverable D2.10

MESAN cloud analysis

WP no:	2
WP leader:	SMHI
Lead beneficiary for deliverable :	SMHI
Name of author/contributors:	Tomas Landelius, Jelena Bojarova
Nature:	Report
Dissemination level:	PU
Deliverable month:	24
Submission date:	October 31, 2016

1. Introduction

Clouds play an important role for solar and terrestrial radiation. As a consequence, clouds have a strong impact on both the energy budget and global climate. The monitoring of the evolution of cloudiness through the years is important for the local climates and in itself an important climate indicator. Surface stations are sparse and with varying density and practices over time. Space based observations are necessary and satellite retrieved cloud products have revolutionised this research. Especially with the number of channels and resolution being available on AVHRR (1978-) and SEVIRI (2004-) instruments the potential for reliable satellite based cloud products is great. Data from earlier geostationary instruments, like MVIRI from EUMETSAT, make it possible to extend the geostationary record back to 1983.

Good quality cloud data sets have been produced by the EUMETSAT Satellite Application Facility on Climate Monitoring (CM SAF) for the geostationary Meteosat and AVHRR polar platforms. They complement each other over the European area but an optimally gridded data set covering all of Europe is needed for climate studies, validation of models and solar energy potential. One of the tasks in UERRA is to produce such a pan-European analysis of hourly cloud fraction at 5.5 km resolution for the years 1983-2013.

This report describes the data and methods used to produce a first version of the European cloud fraction analysis covering the years 2004-2009 when the CLAAS-A1 (SEVIRI) and CLARA-A1 (AVHRR) datasets overlap. A good quality cloud product is soon to be released by the CM SAF also for the entire Meteosat era, 1991 (later on 1983) onwards. The next version of the cloud analysis is intended to be based on this dataset together with the new CLARA-A2 dataset from the polar orbiters to cover the intended 30 years.

2. Data

The proposed cloud cover analysis should combine the best available data for every given point and time. The data used consist of binary information about cloud cover from the CM SAF datasets CLAAS-A1 (SEVIRI) and CLARA-A1 (AVHRR) together with meta-data about position, time and processing status. When no satellite data is available, cloud fraction from the EURO4M HIRLAM reanalysis is used as a gap filler. Synoptic observations (SYNOP) are used to estimate error covariances and for validation. In this section the different datasets are described in more detail.

CLAAS-1

The first edition of the SEVIRI cloud property dataset (CLAAS-1) was released by the CM SAF in 2013 (Stengel et al., 2013). The SEVIRI instrument has a spatial resolution of 3 km × 3 km at the nadir and a complete image of the Earth's full disk consists of 3712 × 3712 pixels. Input to the retrieval schemes were inter-calibrated effective radiances of Meteosat-8 and 9, allowing the the dataset to homogeneously span 8 years from 2004 to 2011. The cloud mask is described in the Algorithm Theoretical Basis Document (CM SAF, 2013a) from which the following information is obtained.

SEVIRI on board MSG represents a significant step ahead compared with the earlier MVIRI instrument, with a higher temporal sampling (quarter hourly instead of half hourly), and an increased spatial resolution (3 km at sub-satellite for all channels except 1 km for the high resolution visible (HRV) channel, compared with 5 km and 2.5 km for the infra-red and visible channels of Meteosat). But the major

improvement is its enhanced spectral characteristics (12 spectral bands against only three for Meteosat) which allows an accurate cloud cover analysis even at night-time (due to its 3.9 µm channel).

The cloud mask is derived directly from results of a cloud screening or cloud masking method and comprises six categories: Cloud filled, cloud-free, cloud contaminated and non-processed, snow/ice contaminated, undefined. The central aim of the cloud mask processing is to delineate all cloud-free pixels in a satellite scene with a high confidence. In addition, the product provides information on the presence of snow/sea ice, dust clouds and volcanic ash plumes.

Bit number	Flag name	Description
0	Land	Pixel is over land.
1	Coast	Pixel is located in the coastal region.
2	Night	It is night where the pixel is located.
3	Twilight	The pixel is located in the twilight zone.
4	Sunglint	Sunglint is likely.
5	High terrain	Pixel is located over high terrain.
6	Inversion	Low level inversion is present according to NWP.
7	NWP present	NWP information have been used.
8	AVHRR channel missing	Some AVHRR channel is missing.
9	Low quality	One or more of the features of the decisive tests were close to one of its thresholds indicating low quality.
10	Reclassification	Pixel was re-classified in the filtering of isolated pixels.
11	Contaminated	The state before spatial smoothing (the filtering above) was cloud contaminated.
12	Cloudy	The pixel was cloudy before spatial smoothing.
13	External sea ice information	External sea-ice information used.
14	Internal sea ice information	Ice information derived from NWP.
15	Sea Ice	Sea-ice present according to sea ice maps.

Table 1: Description of the bits in the PPS processing flag.

Information about the processing is given in a 11 bit quality flag. Here, three bits define illumination and viewing conditions (0: Undefined, 1: Night, 2: Twilight, 3: Day, 4 Sunglint), two bits describe the NWP input data (0: Undefined, 1: All NWP parameters available and no low level inversion, 2: All NWP parameters available with a low level inversion, 3: At least one NWP parameter missing), two bits describe the SEVIRI input data (0: Undefined, 1: All useful SEVIRI channels available, 2: At least one useful SEVIRI channel missing, 3: A least one mandatory SEVIRI channel missing), two bits describe the quality of the processing itself (0: Non processed; containing no data or corrupted data, 1: Good quality, 2: Poor

quality, 3: Reclassified after spatial smoothing, i.e. very low confidence), a single bit for temporal processing indicator (significant for cloud-free pixels; 0: Not performed 1: Performed), and finally one bit for HRV processing indicator (significant for cloud-free pixels; 0: Not performed 1: Performed).

UNIX tar files with the cloud mask product were retrieved from the CM SAF ECFS archive at ECMWF.

CLARA-A1

The CLARA-A1 dataset is a global dataset of cloud, surface albedo and surface radiation products derived from measurements of the Advanced Very High Resolution Radiometer (AVHRR) onboard the polar orbiting NOAA and Metop satellites (Karlsson et al., 2012). The Cloud Mask is the most basic product of the SAF NWC Polar Platform System (PPS). The product is described in detail in a Algorithm Theoretical Basis Document (CM SAF, 2011) from which much of the following information has been collected.

The aim of the Cloud Mask product is to delineate all absolutely cloud-free pixels in a satellite scene with high confidence. In addition, it will identify cloud free snow or ice contaminated pixels when illumination allows and provides processing flags indicating processing conditions and estimated quality for each pixel. The instantaneous CLARA-A1 retrievals have a spatial resolution of 4 km × 4 km. The cloud mask from CLARA-A1 is reported using the same six classes as for CLAAS-1. The meta-data used for the cloud analysis consist of the instantaneous sun-satellite viewing geometry and the processing flags. The processing flag is a 16 bit number that should be interpreted according to Table 1.

Separate UNIX tar files with satellite swath information and cloud mask products were retrieved from the CM SAF ECFS archive at ECMWF.

Synoptic observations

Independent data is needed for validation and statistical error analysis for the optimal interpolation method. For this purpose six-hourly surface data of total cloud cover, coded by the observers into the World Meteorological Organization (WMO) synoptic code (SYNOP), was retrieved as BUFR files from the MARS archive at ECMWF. Here, the cloud cover is reported at 00, 06, 12 and 18 UTC in nine categories (0, 10, 25, 40, 50, 60, 75, 90, and 100 %).

Numerical weather prediction data

A regional reanalysis covering Europe for the years 1979–2014 has been produced with the HIgh Resolution Limited-Area numerical weather prediction (NWP) Model (HIRLAM) forecast model and data assimilation system (Dahlgren et al., 2016). The reanalysis was done as part of the EU FP7 project “European Reanalysis and Observations for Monitoring” (EURO4M). Surface and upper-air variables were analysed at 0000, 0600, 1200 and 1800 UTC on a three-dimensional grid-mesh with 22 km spacing covering Europe. Cloud cover is available as forecast data with 3 hourly time steps and the total cloud fraction is reported with a number between 0 and 1.

For the present cloud analysis these forecast data are used for gap filling in order to produce a continuous data set in space and time for every hour during the years 2004-2009. When needed, the forecast with a valid time closest to the analysis hour is used.

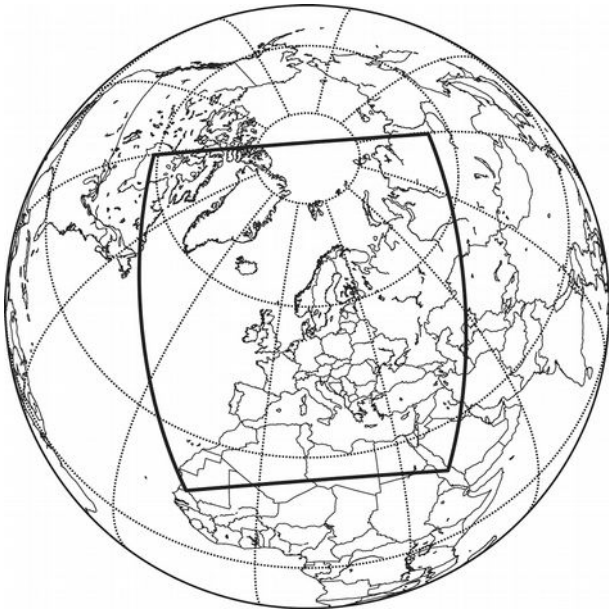

Figure 1: Geographical domain

3. Method

In this section the geometry of the analysis domain is presented along with information about the analysis method, how the observations are prepared for the analysis and how the necessary error statistics is estimated.

Geometry

In EURO4M, the HIRLAM reanalysis was further downscaled and used as input for a 2D-analysis of a number of surface-related parameters (Landelius et al., 2016). To be compatible with these data the present cloud cover analysis is using the same grid and geometry. The horizontal grid is expressed in a rotated latitude and longitude geometry defined by shifting the South Pole to latitude 29° S and longitude 15° E. The grid mesh has 1273 × 1213 grid points with 0.05° resolution (5.5 km) and its geographical coverage of the domain is shown in Figure 1.

Super-observations

The cloud mask from CLAAS-1 and CLARA-A1 is a categorical variable (the six classes previously described). In order to arrive at a variable compatible with traditional cloud cover estimates such as SYNOP data the six classes are first reduced to three – cloudy (pixels classified “cloud-filled” or “cloud-contaminated”, cloud free (“cloud-free” or “snow/ice-contaminated) and missing data (“non-processed” or “undefined”). Then a weighted fractional cloud cover is calculated for subregions with an area of 16.5 km x 16.5 km (3 x 3 analysis grid boxes) using the non-missing data:

$$C_{tot} = \sum w_i C_i / \sum w_i \ .$$

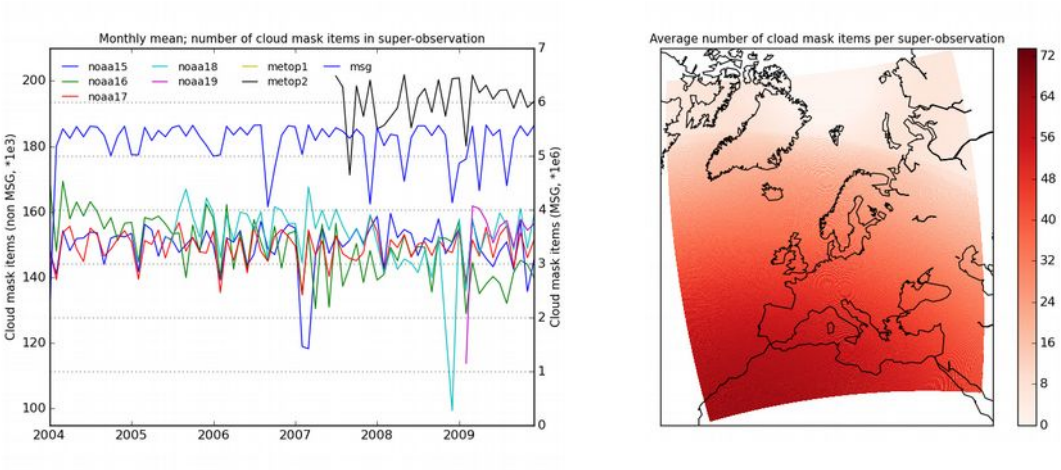

Figure 2: Average number of cloud mask items used for the hourly super-observations. Left: per satellite. Right: per grid box.

The calculations are done using all satellite data acquired within a time window of about 60 minutes, centred at the analysis hour and with overlapping subregions for every point in the analysis grid. This procedure is similar to the one introduced by Lorenc (1981) where single observations are combined into super-observations to avoid numerical instabilities in the analysis when observations are very close together in space. Using a sub-region with an area of 16.5 km x 16.5 km (3 x 3 analysis grid boxes) for the computation of total cloud cover is in line with what has previously been done when comparing CM SAF data to SYNOP observations (Derrien and LeGleau, 2005, Werkmeister et al., 2015).

With a time window of just below 60 minutes there will (when no MSG scenes are missing) be three SEVIRI scenes available for every analysis hour (HH) since these data are acquired every 15 minutes (HH-15, HH+00, HH+15). The number of available swaths from the polar orbiting satellites will depend on the number of operating satellites at the date in question as well as the time of day.

The left panel in Figure 2 shows the monthly average, per satellite, of the number of cloud mask items used for the calculation of one super-observation field. The average number of cloud mask items per a single grid box is shown in the right panel of the same figure. Note that the number of data from the geostationary satellite outnumbers the data from the polar-orbiting satellites to the extent that their contribution to the northern area is hardly visible.

The weights used when calculating the fractional cloud cover depend on three things: 1) time difference between analysis time and data time, 2) scanning geometry, and 3) the SAF processing flags. For the time difference a simple relation is used to account for typical time scales in cloudiness. Note that the weighting drops substantially between 18 and 30 minutes resulting in an effective time window of about 60 minutes:

$$w_{tdiff} = 1/2 + 1/2 \tanh(7 - \delta t \cdot 0.3),$$

Here δt is the time difference in minutes. Distortion due to the viewing geometry is divided into three factors: scan angle, satellite – Earth distance, and surface inclination:

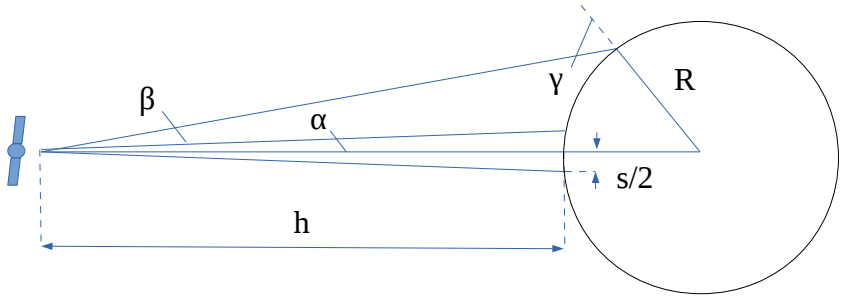

Figure 3: Satellite viewing geometry

$$
\begin{aligned}
\alpha &= 2\,atan(s/2/h) \\
w_{scan} &= h(\tan(\beta)-\tan(\beta-\alpha)) \\
w_{dist} &= 1+\frac{R}{h}(1-\cos(\gamma)) \\
w_{incl} &= 1/\cos(\gamma) \\
w_{geom} &= s/(w_{scan}\,w_{dist}\,w_{incl})
\end{aligned}
\quad ,
$$

Here, w_{geom} is the total weight, α is the apex angle (nadir field of view), β is the satellite scan angle, γ is the Earth zenith angle. The field of view at the sub satellite point is given by *s* and the orbit height and Earth radius by *h* and *R* respectively, see Figure 3.

Information from the SAF processing flags are also used to influence the weighting of the cloud mask observations. Initially the weight for the quality flag, w_{qc}, is set to 1. In case of a AVHRR swath from the polar orbiters the weight is reduced with a factor of 0.5 if any of the following bits are set in the quality flag (see also Table 1): Sunglint (4), Inversion (6), Low quality (9), or Internal sea ice information (14). The weight is set to zero in case of Reclassification (10),. For the SEVIRI scene the weight for the quality is reduced with a factor of 0.5 when “Poor quality” is reported and set to zero if the flag indicates that the observation is “Non processed” or “Reclassified”.

The total weight used in the calculation of the cloud fraction is given by the multiplication of the weights for the time difference, viewing geometry and quality information:

$$w = w_{tdiff}\,w_{geom}\,w_{qc} \quad .$$

The relations for calculating weights based on time differences, viewing conditions and quality meta data were derived in collaboration with SAF personnel (Scheirer, personal communication, 2015).

Optimal interpolation analysis

Originally the idea was to use the MESAN (Häggmark et al., 2000) or MESCAN (Cornel et al., 2016) system for the cloud cover analysis. However, it turned out to be too computationally demanding due to the facts that the number of observations is more or less equal to the number of grid points in the first guess field (NWP cloud cover from HIRLAM) and that the overlapping super-observations are correlated. On the other hand, assuming that the error covariance for the first guess and the super-observations are homogeneous, the analysis can be done efficiently in the Fourier domain.

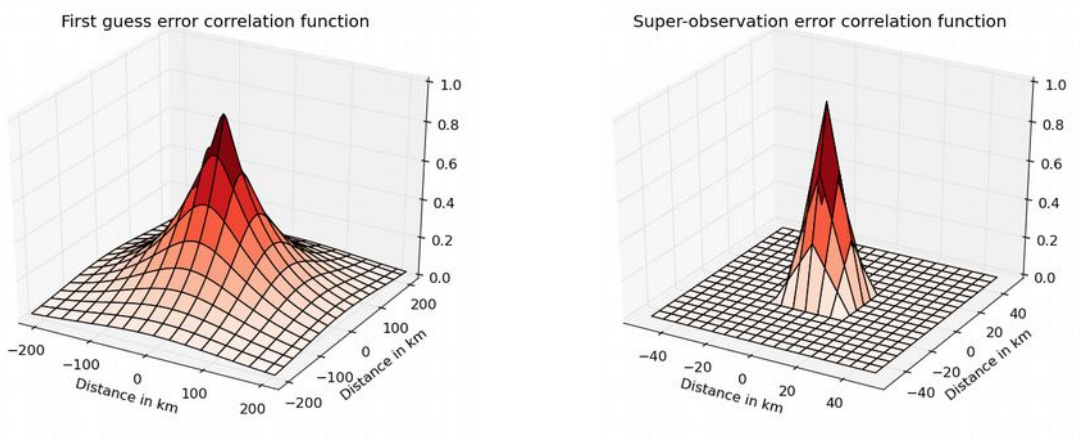

Figure 4: Error correlation functions for the first guess (left) and super-observations (right). Note the difference in horizontal scales – the correlation for super-observations is narrower.

The optimal interpolation (OI) method (Daley, 1991) that is employed by MESCAN/MESAN is a generalisation of the least squares method to multivariate systems and combines the observations and a first guess to produce an analysis. The first guess vector state, x_b , often referred to as the background state, is usually a forecast generated from a NWP model. To obtain the analysis, x_a , an analysis increment, δx_a , is added to the background state. The analysis increment is obtained by calculating the weighted difference between the observational vector data, y , and an observation operator applied to the background, $H(x_b)$. This quantity is also known as the innovation vector, $y - H(x_b)$. In this application the observation operator is the identity matrix since there are observations corresponding to every grind point in the first guess (how this is obtained is described later in this section). Given that H equals the identity operator, the analysis is given by the following relation:

$$x_a = K\left(y - x_b\right),$$

where K is the gain matrix (or weight matrix), determined from the estimated statistical error covariances of the forecast, B, and the observations, R. The expression of the analysis increment, may be further written as follows:

$$\delta x_a = B\left(B+R\right)^{-1}\left(y - x_b\right).$$

Hence, in the OI method, the spatial structure of the innovation is incorporated into the correlation functions. Now with the number of grid points being of the order $O(10^6)$ the number of elements in the covariance matrices are $O(10^{12})$. However, if the error covariance is homogeneous (the same for every point in the grid), the corresponding matrix will have a symmetric Toepliz form and be approximately circulant (true if assuming periodicity of the field) and possible to decompose as (Golub and Van Loan, 1996):

$$C = F^T D F$$

where D is a diagonal matrix representing the Fourier transform of the first row in C. This corresponds to the Fourier transform of the covariance field for any single point in the grid (save a phase shift). Now if we Fourier transform the first guess (or observation) error vectors and calculate their error covariances in Fourier space the resulting matrices will be diagonal (x_t denotes the truth):

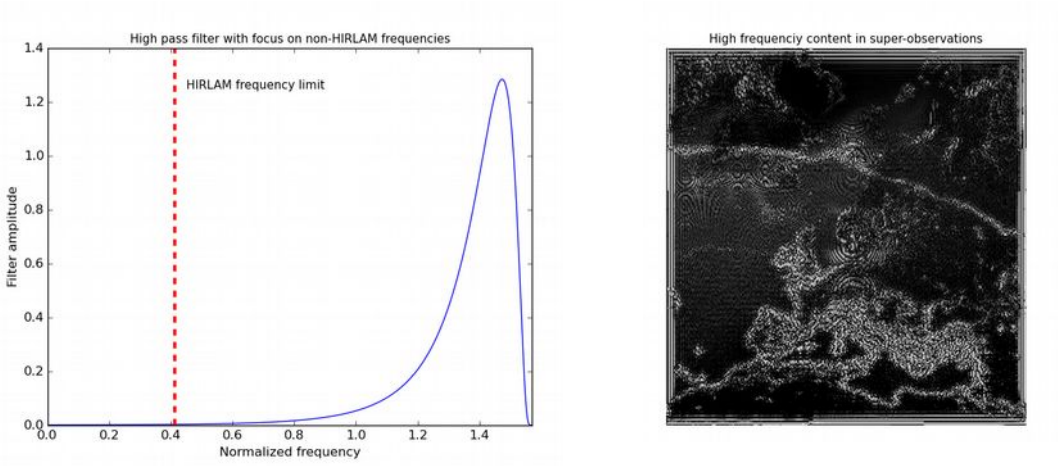

Figure 5: High frequency contents in the super-observations. Left: Amplitude of the used high-pass filter. Right: High-pass filtered mean cloud cover from the super-observations.

$$E\left\{F\left(x_b-x_t\right)\left(F\left(x_b-x_t\right)\right)^T\right\}=F\left\{\epsilon_b\epsilon_b^T\right\}F^T=FBF^T=D_B$$

since

$$F^T D_B F=B \quad .$$

Using this fact the expression for the analysis increment can be rewritten as:

$$\begin{aligned}\delta x_a &= F^T D_B F\left(F^T D_B F+F^T D_R F\right)^{-1}\left(y-x_b\right)\\ &= F^T D_B\left(D_B+D_R\right)^{-1}F\left(y-x_b\right)\end{aligned} \quad .$$

This makes the analysis computationally tractable since all entities are now of order $O(10^6)$ and the two-dimensional Fourier transform, F, as well as its inverse, $F^T=F^{-1}$, can be evaluated efficiently. Note that since the error covariance matrices are assumed to be circulant they equal the error correlation matrix times a constant. Hence it is only the ratio between the scalar covariances for the first guess and the observation error that determines the gain matrix ones the correlation matrices are established.

The error correlation matrix for the first guess from HIRLAM was estimated from an ensemble generated with the NMC method (Parrish and Derber 1992) for the years 1982-2010 based on Fourier transformed differences between 48 and 24 hour forecasts issued at 00, 06, 12, and 18 UTC.

For the correlation matrix corresponding to the observation error another method had to be employed. Here the assumption is that the original cloud mask observations are uncorrelated and that the formation of the super-observations is what brings in the spatial correlation. The calculation of cloud amount as a fraction of the number of cloudy points within a 3 x 3 neighbourhood corresponds to a convolution of the original data with a 3 x 3 box filter. Filtering uncorrelated Gaussian noise results in a signal where the autocorrelation is given by the self-convolution of the filter. In this case the autocorrelation becomes a 2D triangular function of size 6 x 6 grid points. The error correlation functions for the first guess and observations are illustrated in Figure 4.

In order to estimate the error covariances, the following three measures were calculated under the assumption that the first guess, the super-observations and the SYNOP observations are all uncorrelated:

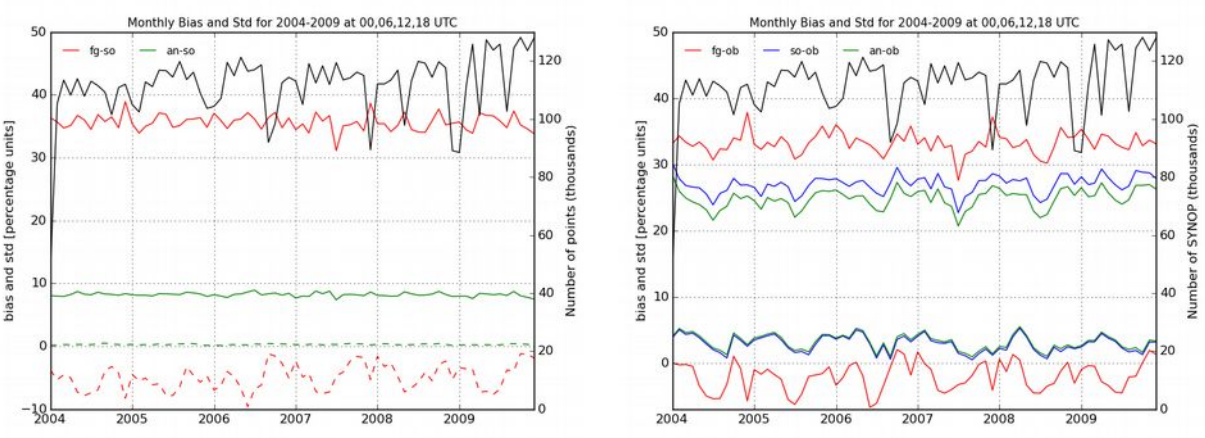

Figure 6: Left: monthly means of bias (dashed lines) and standard deviation (solid lines) for the first guess departure (red) and analysis increment (green). Right: monthly means of bias (lower lines) and standard deviation (upper lines) compared to SYNOP observations (red: first guess, blue: super-observations, green: analysis) . The number of data points used for comparison is shown with a solid black line (legend on right axis) in both panels.

$$
\begin{aligned}
E\{(x_b - y_{synop})^2\} &= E\{((x_b - t) - (y_{synop} - t))^2\} &= E\{\epsilon_b^2\} + E\{\epsilon_{synop}^2\} \\
E\{(x_b - y_{so})^2\} &= E\{((x_b - t) - (y_{so} - t))^2\} &= E\{\epsilon_b^2\} + E\{\epsilon_{so}^2\} \quad . \\
E\{(y_{synop} - y_{so})^2\} &= E\{((y_{synop} - t) - (y_{so} - t))^2\} &= E\{\epsilon_{synop}^2\} + E\{\epsilon_{so}^2\}
\end{aligned}
$$

This equation system was solved for ε_b , ε_{so} and ε_{synop} using statistics from the years 2004 – 2009. In order to reduce the influence of any common underlying seasonal and diurnal variations, separate values where calculated for each month and hour (00, 06, 12 and 18 UTC) and where then averaged to obtain a single value of the error standard deviation per entity: ε_b = 0.29 , ε_{so} = 0.21 and ε_{synop}= 0.17 respectively. The ratio between the error covariance for the first guess and the super-observations was then set to two (0.29^2 / 0.21^2 rounded to nearest integer).

In order to compute the Fourier transform of the field with super-observations it needs to be without missing data. Hence, areas with missing data are replaced with cloud cover from the HIRLAM EURO4M reanalysis. To reduce the gradients between NWP and satellite data a median filter (of size 7 x 7 pixels) is applied along the borders of the areas with missing data. Moreover the same median filter is afterwards also applied to areas with low quality and few cloud mask estimates in the super-observation to reduce the effect of moire patterns in such regions. In order to use the median filter in homogeneous areas a region growing operation is first applied to the mask defining problematic areas (less than 40 cloud mask data constituting the super-observation with a total weight lower than 0.75).

4. Results

To verify that there is more high-resolution information in the gap-filled and median filtered super-observations than in the first guess from HIRLAM, the mean cloud cover from these super-observations was high passed filtered focusing on frequencies not present in the first guess, see left panel in Figure 5. The

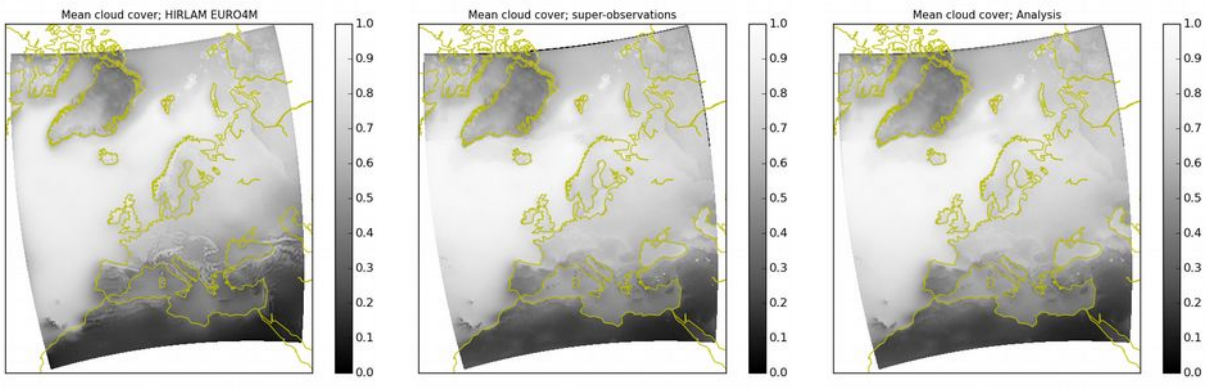

Figure 7: Mean cloud cover for the years 2004 – 2009 (left: EURO4M HIRLAM, middle: super-observations, right: analysis).

result is shown in the right panel of Figure 5 and it is evident that the high frequency information is not just noise since geographical features are clearly recognizable.

To check that the assimilation is working properly first guess and analysis departures (differences compared to super-observations) are calculated. Monthly means of the bias and standard deviations of these departures are illustrated in the left panel of Figure 6. The analysis works as desired and adjusts the first guess to the observations reducing the standard deviation and bringing the bias down to almost zero. However, the degree of this fit is to a large extent controlled by the prescribed first-guess and observation-error statistics. In order to review the performance of the analysis it should to be compared to independent data. The results from such a comparison with independent SYNOP data is shown in the right panel of Figure 6. Here the standard deviation of the analysis increments is lower than that of both the first guess and super-observation departures. The bias of the analysis increment is reduced (but changes sign) compared to the first guess but is almost identical to that of the super-observation increments.

The mean cloud cover for the years 2004 - 2009 is shown in Figure 7 for the first guess (EURO4M HIRLAM, the super-observations and the resulting analysis). Here it also clear that the analysis is close to the super-observations, which is the desired behaviour. Analysis increments (firs guess minus analysis in this case) for the winter and summer season as well as for day and night time are presented in Figure 8. The increment is lowest for the winter night time and largest for the summer noon. Since there is very few observation in the northernmost region the increment is almost zero there at all times. However, south of this region in the North Atlantic, during daytime in the winter. Over the Mediterranean the first guess presents too much clouds during the winter and too little during summer nights while it is neutral for the summer noon. On the other hand, at summer noon the first guess has too much clouds over most parts of the European continent. This is also true for the winter day time, but to a lesser degree.

5. Conclusion

As part of the UERRA project an hourly analysis of total cloud cover has been produced on a grid with a horizontal resolution of 5.5 km using the optimal interpolation method. Binary cloud mask data from the CM SAF CLAAS-1 and CLARA-A1 datasets where used as input to a process where super-observations of cloud cover was formed as a weighted average over a neighbourhood of 3 x 3 grid boxes. Cloud cover

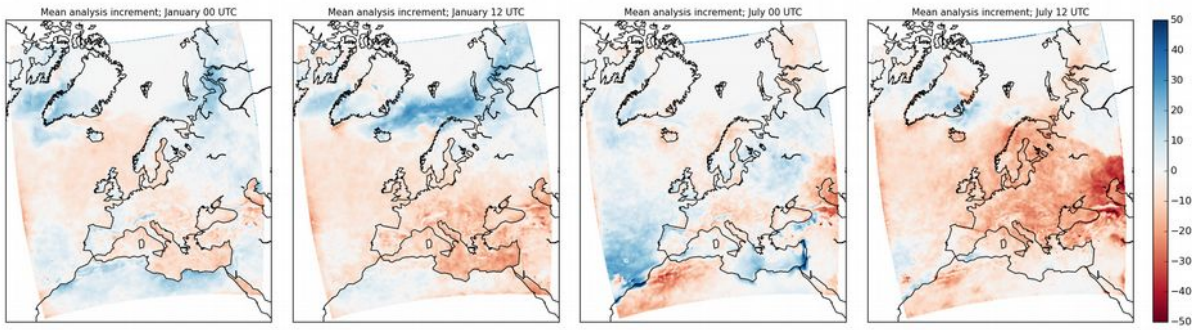

Figure 8: Mean analysis increments for different months (January and July) and times of day (00 and 12 UTC) for the years 2004 – 2009.

from the EURO4M HIRLAM reanalysis was used for the first guess in the analysis and also as a gap filler in case of missing data.

The analysis works as expected and draws the first guess close to the super-observations. It also verifies well when compared to independent SYNOP data. The super-observations have a positive bias compared to SYNOP observations. There could be several reasons for this bias (CM SAF, 2013). One example is the well known overestimation of cloudiness by geostationary satellites at high viewing zenith angles. The SYNOP measurements can also be inaccurate for the same reason with a more small scale geometry. The so-called scenery effect leads to the overestimation of cloudiness by SYNOP due to the obscuring of cloud-free spaces by convective clouds with high vertical extent.

A problem with the generation of super-observations is that it results in some moiré patterns. The problem persists even if the size of the neighbourhood used for binning the binary cloud mask is increased to 22 x 22 km. The solution used here was to apply a median filter to reduce the effect. However, even if the median filter smooths less than a low pass filter it still results in a loss of resolution in the filtered output. The reason behind the formation of the moiré pattern should be investigated in more detail prior to a run on the upcoming CM SAF data covering the years 1983 – 2014.

The estimation of the error covariances for the first guess and the super-observations was based on statistics from differences between first guess and the super-observations as well as differences between these two entities and SYNOP observations. The assumption was that the first guess, the super-observations as well as the SYNOP observations should be independent. However, a systematic variable bias that appears in any two of these data will result in a correlation and add an offset to both the estimate of the covariance error for the first guess and the super-observations. Since it is the ratio between these two that is of importance for the result it will depend on how large this offset will be. The method of Desroziers et al. (2005) was tested but resulted in unrealistic large length scales of the statistics for B + R. Again the reason for this is probably an underlying seasonal co-variation in the first guess and the super-observations. To reduce these effect, statistics for individual months and times of the day these effects where calculated and then averaged in order to arrive at a better estimate.

With the arrival of new data from the CM SAF in terms of a cloud cover product for the whole Meteosat period (1983 – 2015) and the new CLARA-A2 version (1982 – 2014) it should now be possible to apply the method presented in this report to these data and finally produce a combined analysis of total cloud cover for all of Europe.

References

CM SAF (2011), CM SAF Cloud, Albedo, Radiation dataset, AVHRR-based, Edition 1 (CLARA-A1) Cloud Products Algorithm Theoretical Basis Document. Available at http://www.cmsaf.eu/EN/Documentation/Documentation/ATBD/pdf/SAF_CM_DWD_ATBD_GAC_CLD_1.pdf?__blob=publicationFile&v=5

CM SAF (2013a), Algorithm Theoretical Basis Document. SEVIRI cloud products. Edition 1. Available at http://www.cmsaf.eu/EN/Documentation/Documentation/ATBD/pdf/SAF_CM_DWD_ATBD_SEV_CLD_1.pdf?__blob=publicationFile&v=5

CM SAF (2013b), Validation Report SEVIRI cloud products. Edition 1. Available at http://www.cmsaf.eu/EN/Documentation/Documentation/ValidationRep/pdf/SAF_CM_DWD_VAL_SEV_CLD_1.2.pdf?__blob=publicationFile&v=4

Soci, C. et al. (2016), High-resolution precipitation re-analysis system for climatological purposes. **Tellus A**, [S.l.], apr. 2016. ISSN 1600-0870, doi:http://dx.doi.org/10.3402/tellusa.v68.29879.

Dahlgren, P., Landelius, T., Kållberg, P. and Gollvik, S. (2016), A high-resolution regional reanalysis for Europe. Part 1: Three-dimensional reanalysis with the regional HIgh-Resolution Limited-Area Model (HIRLAM). Q.J.R. Meteorol. Soc., 142: 2119–2131. doi:10.1002/qj.2807

Daley, R., (1991), Atmospheric data analysis. Cambridge University Press. ISBN 0-521-38215-7.

Desroziers, G., Berre, L., Chapnik, B. and Poli, P. (2005), Diagnosis of observation, background and analysis-error statistics in observation space. Q.J.R. Meteorol. Soc., 131: 3385–3396. doi:10.1256/qj.05.108

Golub, G. H. and Van Loan, C. H. (1996), Matrix computations (third edition). Johns Hopkins University Press, Baltimore, USA

Häggmark, L., Ivarsson, K.-I., Gollvik, S. and Olofsson, P.-O. (2000). Mesan, an operational mesoscale analysis system. Tellus 52A, 2-20.

Karlsson, Karl-Göran; Riihelä, Aku; Müller, Richard; Meirink, Jan Fokke; Sedlar, Joseph; Stengel, Martin; Lockhoff, Maarit; Trentmann, Jörg; Kaspar, Frank; Hollmann, Rainer; Wolters, Erwin. (2012): CLARA-A1: CM SAF cLouds, Albedo and Radiation dataset from AVHRR data - Edition 1 - Monthly Means / Daily Means / Pentad Means / Monthly Histograms. Satellite Application Facility on Climate Monitoring. DOI:10.5676/EUM_SAF_CM/CLARA_AVHRR/V001. http://dx.doi.org/10.5676/EUM_SAF_CM/CLARA_AVHRR/V001

Landelius, T., Dahlgren, P., Gollvik, S., Jansson, A. and Olsson, E. (2016), A high-resolution regional reanalysis for Europe. Part 2: 2D analysis of surface temperature, precipitation and wind. Q.J.R. Meteorol. Soc., 142: 2132–2142. doi:10.1002/qj.2813

Lorenc, A. C. (1981), A Global Three-Dimensional Multivariate Statistical Interpolation Scheme. *Mon. Wea. Rev.*, **109**, 701–721, doi: 10.1175/1520-0493(1981)109<0701:AGTDMS>2.0.CO;2.

Parrish, D., and J. Derber, (1992), The National Meteorological Center's spectral statistical interpolation analysis system. Mon. Wea. Rev., 120, 1747-1763.

Stengel, Martin; Kniffka, Anke; Meirink, Jan Fokke; Riihelä, Aku; Trentmann, Jörg; Müller, Richard; Lockhoff, Maarit; Hollmann, Rainer. (2013): CLAAS: CM SAF CLoud property dAtAset using SEVIRI - Edition 1 - Hourly / Daily Means, Pentad Means, Monthly Means / Monthly Mean Diurnal Cycle / Monthly Histograms. Satellite Application Facility on Climate Monitoring. DOI:10.5676/EUM_SAF_CM/CLAAS/V001. http://dx.doi.org/10.5676/EUM_SAF_CM/CLAAS/V001

Werkmeister, A., Lockhoff, M., Schrempf, M., Tohsing, K., Liley, B., and Seckmeyer, G.: Comparing satellite-to ground-based automated and manual cloud coverage observations – a case study, Atmos. Meas. Tech., 8, 2001-2015, doi:10.5194/amt-8-2001-2015, 2015.