Better use of correlation information in AMV extraction scheme

Régis Borde, Arthur De Smet and Greg Dew
EUMETSAT, Am Kavalleriesand 31, 64295 Darmstadt, Germany

Abstract

The Height Assignment (HA) is currently the most difficult task in the Atmospheric Motion Vectors (AMV) extraction scheme. Several sources of error can be introduced at the height assignment step, but one of the main difficulties is to clearly identify the pixels that lead the tracking process in the tracer box, in order to select them for the HA calculation. A good pixel selection process should ensure a direct link is kept between the feature really tracked and the calculation of the height. The most common method sorts the coldest pixels in the target box and uses them to calculate the AMV height. However, recent work showed that some of the coldest pixels can have very small and/or negative contributions to the cross correlation process. Indeed, it is then proposed to use individual pixel contribution-to-cross-correlation-coefficient information in the pixel selection process, in order to get a closer link between the feature tracked and the HA. This paper will present and discuss the latest results that have been investigated at EUMETSAT, in order to improve this pixel selection process in AMV extraction algorithms.

INTRODUCTION

Atmospheric Motion Vectors (AMVs) are one of the most important products derived from all geostationary satellites, because they constitute a significant part of the observation data assimilated in Numerical Weather Prediction (NWP) models. Indeed, they are the only upper wind observations with good global coverage for the tropics and mid-latitudes, especially over the large ocean areas. AMVs are routinely extracted by a number of meteorological satellite operators (EUMETSAT, NOAA/NESDIS, Japan Meteorological Agency (JMA)).

For Meteosat Second Generation (MSG), the derivation of displacement vectors is realised by tracking clouds or water vapour features in consecutive Meteosat satellite images. The final hourly AMV product is an average of three vectors calculated from a sequence of four images. The basic elements of wind vector production are (Holmlund, 2000): (a) selecting a feature to track; (b) tracking the target in a time sequence of images to obtain a relative motion; (c) assigning a pressure (altitude) to the vector; and (d) assessing the quality of the vector. The height assignment step, which estimates cloud top height in the case of cloud motion vectors, is still recognised to be the source of largest error in this AMV extraction process. In particular, the pressure associated with the AMV is sometimes smaller than the forecast pressure of best correspondence to the detected speed. This situation is especially frequent at high levels, in strong wind shear situations, and can have a negative impact on the forecast issued from NWP models. He operationally used height assignment schemes try to account for semi-transparency, broken cloud fields, multi-layered clouds, and low level targets. The sources of errors include the sensitivity of the methods to local atmospheric parameters (Borde and Dubuisson, 2007). Within all these schemes, the pixels that will be input to the height assignment need to be selected. However, this selection process is not currently done with a direct and clear link to the tracked feature.

At EUMETSAT, AMVs are derived e.g. from tracking clouds in the 10.8 µm IR channel of Meteosat, using a target box size of 24x24 pixels (72x72 km² at the sub-satellite point). Various types of clouds can be detected in such a target box, potentially moving at different speeds and altitudes. One of the critical issues is then to select the pixels within in the target box that should be used for the HA calculation. The EUMETSAT cross correlation scheme used for tracking the clouds is mainly based on the maximum contrast criterion, which intrinsically tends to favour the coldest cloudy pixels in the
infrared channels. The EUMETSAT height assignment scheme uses a sub-group of the cloudy pixels for the height assignment, considering the coldest peak of the 1-D histogram of cloud top pressures calculated on a pixel basis in the target box (ASD internal EUMETSAT document). However, this set of pixels selected does not necessarily correspond to those that mostly contribute to the tracking. In this paper the individual pixel contribution to the cross correlation coefficient defined by Büche et al. (2006) is used to illustrate the necessity for considering such information in the AMV HA.

INDIVIDUAL PIXEL CONTRIBUTION TO THE CROSS CORRELATION COEFFICIENT

In the following, the AMVs considered are derived from tracking clouds in the 10.8 µm IR channel of Meteosat-8, using 24x24 pixels target boxes. No filters and/or enhancement processes are used in this study. The first step of the AMV extraction scheme is to find a target at a selected grid point. For each location within a target search area of 48x48 pixels centred on the grid point, the local mean and standard deviation of 3x3 pixels are computed. The target location is selected to be the 24x24 target box which has the maximum contrast. Overlap of more than 50% is not allowed between adjacent targets. Initially the scheme will try to find a cloudy target, containing more than 50 pixels classed as cloudy, based on the MSG cloud mask product. If this is not found, it will try to find a clear sky target.

Thereafter the position that best corresponds to this target is located in the second image, taken 15 minutes later. The search is done in an 80x80 pixel box centred on the target location. A cross correlation method is used to do the matching. The matching compares only the individual pixel counts of the target box with all possible locations of the target box in the search area to find the best match. The degree of matching between pixel radiances/counts a and b between the two images A and B is classically given by the following two-dimensional cross-correlation coefficient:

$$\text{CC}(m,n) = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{a_{i+m,j+n} - \bar{a}(m,n)}{\sigma_a(m,n)} \frac{b_{i,j} - \bar{b}}{\sigma_b} = \sum_{i,j} \text{CC}_{ij}(m,n)$$

(1)

where m, n is the (lines, elements) displacement of the target box in image B from the initial position in the first image A. The correlation coefficient CC(m, n) is normalized to values between -1 (mirror structures) and +1 (identical structures). The symbols $\bar{a}$ and $\sigma_a$ represent the average and the standard deviation of the radiances/counts a in image A, respectively (correspondingly for b in image B). Values M and N correspond to the box size, MxN=24x24 for this study. According to Büche et al. (2006), the correlation coefficient can also be written following the third part of Eq. (1), where the symbol $\text{CC}_{ij}$ expresses how much the individual pair of pixels (i, j) in image1 and (i+m, j+n) in image 2, contributes to $\text{CC}(m, n)$.

![Figure 1](image1.png) Figure 1: Infrared counts within the target area are plotted against their individual pixel contribution (left) for the corresponding target area (right). AMV has been extracted using IR10.8 channel of SEVIRI (1st December 2006, 2:00 and 2:15 UTC images)
Figure 1 illustrates how the individual pairs of pixels taken from the 24x24 pixel target boxes between two consecutive Meteosat-8 images (1st December 2006, 2:00 and 2:15 UTC images), contribute to the maximisation of CC(m,n). Green dots correspond to clear sky pixels, red dots to cloudy pixels within the target area. The corresponding scene (SCE-CLA) and cloud top height (CLA-CTH) information are plotted on the right side. High level, mid level and low level clouds correspond respectively to clear blue, violet and grey colours. The correlation matching has been done using count values, but radiance values can also be used. Usually, the coldest and warmest pixels in the target box contribute the most to CC(m,n). In the case of a clear distinction between cold and warm scenes within the target box, the relative individual pixel contributions, CC
\[\text{ij}\]
, present a clear ‘C-shaped’ distribution, as shown in Figure 1. The distance between the two branches corresponds to the contrast of the structures within the target area. Several pixels have a negative CC
\[\text{ij}\]
, which generally correspond to pixels that have very different radiative properties but the same position within the two target boxes in the image 1 and image 2. Appearance and/or decay of clouds between images 1 and 2 generally induce such negative values for CC
\[\text{ij}\]. Pixels that contribute the most to CC(m,n) can be defined as those that have CC
\[\text{ij}\] > average CC
\[\text{ij}\], figured by the dashed blue line Figure 1.

ESTIMATION OF THE PRESSURE USING CC
\[\text{ij}\] INFORMATION

To estimate the AMV altitude, the satellite operators (EUMETSAT, NOAA/NESDIS, JMA) use several techniques to select the pixels that are used for the HA calculation inside the target box. But none of them currently includes direct information coming from the correlation process. The most common method selects the coldest pixels present in the target box to calculate the height. NOAA/NESDIS uses a fixed threshold of 25% coldest pixels for the GOES instrument (Daniels et al., 2002). EUMETSAT uses as a basis the cluster associated with the coldest peak of the dynamic histogram analysis of CLA-CTH parameter, calculated on a pixel basis inside the target area for MSG (ASD internal EUMETSAT document). These CLA-CTH estimations are corrected for semi-transparency effect. JMA uses the most frequent 1-D histogram of cloud top pressures calculated on a pixel basis in the target box for MTSAT (Oyama and Shimoji, 2008). These calculations are also corrected for semi-transparency effect. Once the pixels are selected, several methods are used to calculate the altitude of the AMV as function of the cloud type. Opaque cloud heights are estimated from the representative Equivalent Black Body Temperatures. The CO2 slicing (Menzel et al., 1983) and WV-IRW intercept methods (Schmetz et al., 1993) are used operationally to estimate the cloud top pressure of semi-transparent clouds; both require clear sky and cloudy formation present in the target box to estimate the HA. All methods are applied on the cluster or selected samples of the cluster. The representative or sampled cloudy radiance of the cluster generally corresponds to the average radiance of the cloudy pixels that are selected.

Instead of a simple radiance average, the representative clear sky and cloudy radiances can be calculated by weighting the pixel radiances by CC
\[\text{C}\] and used as input for the HA methods. Oyama et al. (2008) show some improvements applying this at JMA. However, in the framework of this study we have tested the method described by Borde and Oyama (2008), which uses the CLA-CTH parameters calculated on a pixels basis to estimate the pressure. That choice has been guided by the EUMETSAT intention to use its future CLP (also called OCA) product (Watts et al., 1998) to provide an estimate of the AMV pressure. This one will provide information on cloud top height, cloud fraction, cloud type, droplet size,…etc, for every cloudy pixel. As this product is currently under operational implementation and not yet available, the preliminary tests presented below have been done using the classical CLA-CTH product.

Figure 2 shows the CLA-CTH within the target area plotted as function of CC
\[\text{C}\] for the same case figured out in Figure 1. Triangle and diamonds correspond now to the weighted pressure calculated from CLA-CTH, balanced by their corresponding CC
\[\text{C}\]. The dark blue triangle represents the result for the pixels that contribute more than <CC
\[\text{C}\] to the correlation coefficient for the coldest branch (cf. Figure 1), while the clear blue diamond is the result for the whole cold branch (CC
\[\text{C}\] > 0). The orange triangle and greenish diamond correspond to the same calculations for the warm branch. Error bars represent the corresponding weighted standard deviation, balanced by CC
\[\text{C}\]. The horizontal dashed line corresponds to the AMV pressure estimated using the 25% coldest pixels. The solid red line corresponds to the CLA-CTH histogram. Some advantages of this technique are described in Borde and Oyama (2008), and it is beyond the scope of the current study to describe them in more detail.
Figure 2: AMV pressure and standard deviation are estimated for cold and warm branches as a weighted mean of the CLA-CTH parameters, balanced by the individual pixel contribution CC_ij to correlation.

PRELIMINARY COMPARISON

To estimate the impact of the method described above, some comparisons have been made against radiosonde observation and forecast fields. However, for technical reasons and operational chain availability, 3 changes have been tested at the same time. On the ‘OPE’ chain the current operational algorithm was used, in order to make comparisons. On the ‘LTV’ chain, the image enhancement procedure used on OPE for the infrared channel has been switched off; the CLA-CTH has been calculated using the CO2 slicing method (channel 12.0 and 13.4 μm) instead of a similar method using channels 6.2 and 10.8 μm that is employed on OPE, and the AMV pressure has been calculated following the method described above, instead of the dynamic clustering histogram analysis and method combination described in the previous section, used on OPE.

The AMVs obtained on the LTV chain have a general better quality than those obtained on OPE (results not shown). The increase of AMVs having a quality index greater than 80% on the LTV chain is nearly 15 to 20% compared to OPE. The forecast consistency, vector consistency, speed consistency and direction consistency tests give all better performances on the LTV chain than on OPE.

The AMVs extracted on OPE and LTV chains have been compared against the nearest radiosonde observations in the period from 20 August to 9 September 2008. The statistics are presented in Tables 1 and 2 respectively. Following the usual CGMS standard format, several criteria must be satisfied to consider a good co-location between AMV and R/S: horizontal distance < 150 km; vertical distance < 25 HPa; speed difference < 30 m/s; direction difference < 60 deg and AMV speed > 2.5 m/s. Specific statistics are also done filtering only the excellent AMVs that have a quality index greater than 80% (noted EXL in the Tables 1 and 2). Results give information on the mean speed, the mean vector difference, the NRMS, the number of collocation, the RMS and the speed bias. In addition to global statistics, specific information is given for high level AMVs (above 400 hPa), low level AMVs (below 700 hPa) and medium level AMVs (between 400 and 700 hPa), and for 3 geographical areas: North Hemisphere, South Hemisphere and Tropics.
<table>
<thead>
<tr>
<th></th>
<th>GLOBAL</th>
<th>NORTH</th>
<th>SOUTH</th>
<th>TROPICS</th>
<th>GLOBAL EXL</th>
<th>NORTH EXL</th>
<th>SOUTH EXL</th>
<th>TROPICS EXL</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL LEVELS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN OBS. SPEED</td>
<td>12.67</td>
<td>12.25</td>
<td>21.83</td>
<td>10.30</td>
<td>14.08</td>
<td>13.48</td>
<td>22.95</td>
<td>11.00</td>
</tr>
<tr>
<td>MEAN VECT. DIFF</td>
<td>6.59</td>
<td>0.57</td>
<td>0.55</td>
<td>0.66</td>
<td>0.38</td>
<td>0.34</td>
<td>0.36</td>
<td>0.45</td>
</tr>
<tr>
<td>MEAN OBS. SPEED</td>
<td>4.30</td>
<td>1.91</td>
<td>1.18</td>
<td>0.69</td>
<td>0.50</td>
<td>0.63</td>
<td>0.66</td>
<td>0.46</td>
</tr>
<tr>
<td>MEAN VECT. DIFF</td>
<td>3.03</td>
<td>0.41</td>
<td>0.39</td>
<td>0.42</td>
<td>0.30</td>
<td>0.34</td>
<td>0.37</td>
<td>0.41</td>
</tr>
<tr>
<td>MEAN OBS. SPEED</td>
<td>4.23</td>
<td>4.93</td>
<td>5.28</td>
<td>4.18</td>
<td>3.35</td>
<td>3.35</td>
<td>3.35</td>
<td>3.35</td>
</tr>
<tr>
<td>MEAN VECT. DIFF</td>
<td>4.67</td>
<td>4.93</td>
<td>5.28</td>
<td>4.18</td>
<td>3.35</td>
<td>3.35</td>
<td>3.35</td>
<td>3.35</td>
</tr>
<tr>
<td>LOW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN OBS. SPEED</td>
<td>4.30</td>
<td>1.91</td>
<td>1.18</td>
<td>0.69</td>
<td>0.50</td>
<td>0.63</td>
<td>0.66</td>
<td>0.46</td>
</tr>
<tr>
<td>MEAN VECT. DIFF</td>
<td>3.03</td>
<td>0.41</td>
<td>0.39</td>
<td>0.42</td>
<td>0.30</td>
<td>0.34</td>
<td>0.37</td>
<td>0.41</td>
</tr>
<tr>
<td>MED</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN OBS. SPEED</td>
<td>12.67</td>
<td>12.25</td>
<td>21.83</td>
<td>10.30</td>
<td>14.08</td>
<td>13.48</td>
<td>22.95</td>
<td>11.00</td>
</tr>
<tr>
<td>MEAN VECT. DIFF</td>
<td>6.59</td>
<td>0.57</td>
<td>0.55</td>
<td>0.66</td>
<td>0.38</td>
<td>0.34</td>
<td>0.36</td>
<td>0.45</td>
</tr>
<tr>
<td>MEAN OBS. SPEED</td>
<td>4.30</td>
<td>1.91</td>
<td>1.18</td>
<td>0.69</td>
<td>0.50</td>
<td>0.63</td>
<td>0.66</td>
<td>0.46</td>
</tr>
<tr>
<td>MEAN VECT. DIFF</td>
<td>3.03</td>
<td>0.41</td>
<td>0.39</td>
<td>0.42</td>
<td>0.30</td>
<td>0.34</td>
<td>0.37</td>
<td>0.41</td>
</tr>
</tbody>
</table>

**AMV R / S CO-LOCATION STATISTICS**

**AMV R / S CO-LOCATION STATISTICS**
Comparison of Tables 1 and 2 shows a general increase in the number of co-locations on the LTV chain, which means that the AMV derived characteristics on LTV agree with R/S observations more frequently, having comparable speed, direction and altitude (within 25 hPa) in the same time. This increase is more important for high level and mid level AMVs with respectively +12% and +53 % more global co-locations with R/S than results obtained from the OPE chain. These percentages are respectively +30% and +88% considering only the global excellent (EXL) AMVs that have a quality index greater than 80%. There are then more excellent AMVs on the LTV chain, and they are statistically more frequently in good agreement with R/S observations. However, the speed biases are unfortunately slightly degraded on LTV, especially at high level (~1m/s worse than on OPE). But, like biases are not estimated on the exact same sets of data on LTV and OPE chains, hence a direct statistical comparison of this criterion is very tricky.

At low level, both biases and RMS are similar on LTV and OPE chains. The number of co-locations is larger on OPE than on LTV for the global set of data, -17% on LTV, but it is the opposite considering only excellent AMVs, +5.5% on LTV. It should be noted that the amount of good co-locations is reduced on LTV for the Northern Hemisphere, -9%, and is increased for the Southern Hemisphere, +25%, and for Tropics, +13%. It can be noted that biases are generally worse in the Southern Hemisphere for both OPE and LTV.

Another classical test to check the quality of AMVs consists of estimating the level of best fit against forecast fields. Technically, that test compares the AMV speed and direction against the local forecast wind speed and direction profiles, and searches for the best agreement. The altitude determined by the level of best fit is then compared to the altitude set to the AMV. Interpretation of the best fit comparisons requires however some care, due to possible tricky situations where there is not a unique well defined minimum vector difference. Best-fit results of two different tests are presented in Figure 3, one of them considering very strict criteria of comparison (solid line) ensuring a very well defined minimum vector difference between AMV and forecast fields, the other one considering weak criteria of comparison (dashed line).

![Figure 3: Best-fit bias between AMV pressure and best-fit level from forecast field. Red curves correspond to the use of weak (dashed line) and strict (solid line) criteria to estimate the best-fit pressure on OPE chain. Green curves represent the same for LTV chain.](image)

The results are presented only for high level AMVs, between 150 and 400 hPa, and split by pressure layers (y_axis) for which the corresponding mean best_fit bias is calculated (x_axis). The best-fit pressure bias between AMVs and forecast fields is reduced on the LTV (green curves) chain.
compared to the OPE chain (red curves) for both weak (dashed lines) and strict (solid lines) criteria. The amount of good best-fit estimates for the whole period (20 August – 9 September) is also larger on the LTV chain than on OPE: 959 good estimates on OPE and 1302 on LTV (35% more than on OPE). Considering strong best-fit criteria, the amounts on OPE and LTV are respectively 187 and 271, which represents an increase of 45% on LTV.

**DISCUSSION**

The technique tested above (Borde et Oyama, 2008) uses the existing CLA-CTH parameter in conjunction with CC\_ij information to set AMV height. This technique should improve the overall consistency of the AMV algorithms, preserving the essential link between altitude estimation and feature tracked.

Preliminary results of a comparison with the current AMV algorithm at EUMETSAT show:
- An increase in the amount of AMVs having QI > 80
- A slightly better general agreement with forecast fields
- A degradation in the high level slow bias against radiosonde observation, but an increase in the number of collocations. At mid level and low level, biases are comparable to operational outputs, but the number of good collocations with radiosonde observations is increased.

The current results are obtained for only 3 weeks of comparison, from 20 August to 9 September 2008, which is not long enough to draw strong conclusions. However some positive tendencies clearly appear, even if the slow high level bias against radiosonde is slightly degraded. As 3 changes have been tested at the same time, further investigations are needed to determine the real impact of each change on the results.

**ACKNOWLEDGEMENTS**

The authors would like to gratefully thank J.B. Gustaffson for help and interesting discussions.

**REFERENCES**

ASD; ‘MSG Meteorological Products Extraction Facility Algorithm Specification Document’ edited by EUMETSAT. Reference: EUM.MSG.SPE.022


