A NEW NESTED TRACKING APPROACH FOR REDUCING THE SLOW SPEED BIAS ASSOCIATED WITH ATMOSPHERIC MOTION VECTORS (AMVS)

Jaime Daniels[1] and Wayne Bresky[2]

1. NOAA/NESDIS Office of Research and Applications, NOAA Science Center, Camp Springs, Maryland 20746, U.S.A.
2. I.M. Systems Group (IMSG), Rockville, Maryland 20852, U.S.A.

Abstract

The GOES-R Algorithm Working Group (AWG) Winds Team is working on the development and validation of algorithms for the generation of Atmospheric Motion Vectors (AMVs) from the future GOES-R Advanced Baseline Imager (ABI). Meteosat SEVIRI imagery is serving as a GOES-R ABI proxy data source for the development, testing, and validation of the GOES-R AMV algorithm. The findings discussed in this paper are a direct result of these testing and validation activities.

Statistics comparing satellite-derived motion estimates to collocated radiosonde observations often show a pronounced slow speed bias at mid and upper levels of the atmosphere. One possible explanation for this slow bias is a poorly assigned height (too high). Recent work by Sohn and Borde (2008), however, suggested a link between the size of the target box used and the magnitude of the slow bias. Specifically, they found that a smaller target box leads to both a faster wind estimate and a lower height assignment. Both of these factors will contribute to a smaller slow bias. Independent tests performed by the authors of this paper that involved varying target size (5 to 21 pixels) and temporal intervals (5 to 30 minutes) have confirmed these earlier findings. This testing, as well as subsequent analysis of individual case studies, have led the authors to develop a new approach to tracking that relies on a smaller target box "nested" within a larger one to derive a field of vectors for each pixel location in the larger window. The resulting field of vectors is analyzed with a cluster analysis program (DBSCAN) to identify the dominant motion contained within the target scene and a final average wind is computed from the sample of vectors comprising the largest cluster. Additionally, the cloud top heights of the points making up the largest cluster are used to assign a representative height to the AMV that directly links the height assignment step to the tracking step. A comparison of AMVs derived via the new approaches with AMVs derived conventionally show a significant improvement in the overall quality of the derived AMVs characterized by significant reductions in the slow speed bias as well as the RMSE.

1. INTRODUCTION

Satellite-derived AMVs have long exhibited a pronounced slow speed bias at mid (400-700mb) and upper levels (100-400 mb) of the atmosphere when compared to radiosonde profiles and other wind measurements. This is particularly true in the extratropical winter season (Forsythe and Saunders, 2008). Figure 1 shows a typical time series plot of the speed bias for the GOES-12 infrared winds at upper levels. Note the consistent 1-2 m/s slow bias. The slow bias is often attributed to a combination of factors that include an improper height assignment (too high), clouds not behaving as passive tracers and the possibility that the target scene contains multiple scales of motion leading to the derivation of an estimate that is a spatial average. Recent work by Sohn and Borde (2008) has shown a direct link between the size of the target box and the magnitude of the slow bias. Specifically, the authors noted an increase in speed and a lowering of height as the box size is reduced, both of which will contribute to a smaller slow bias. Unfortunately, using a smaller target box produces "noisier" winds (e.g., RMS increases as box size decreases). Complicating matters further is the observation of Borde and Oyama (2008) that the pixels having the greatest influence on the tracking solution are not necessarily the same pixels used to assign a height to the tracer. Most producers of satellite-derived

winds use the coldest pixels from the target scene (Genkova et. al., 2008) to assign height regardless of their importance in the tracking process. The implied disconnect between the two processing steps often leads to large verification errors.

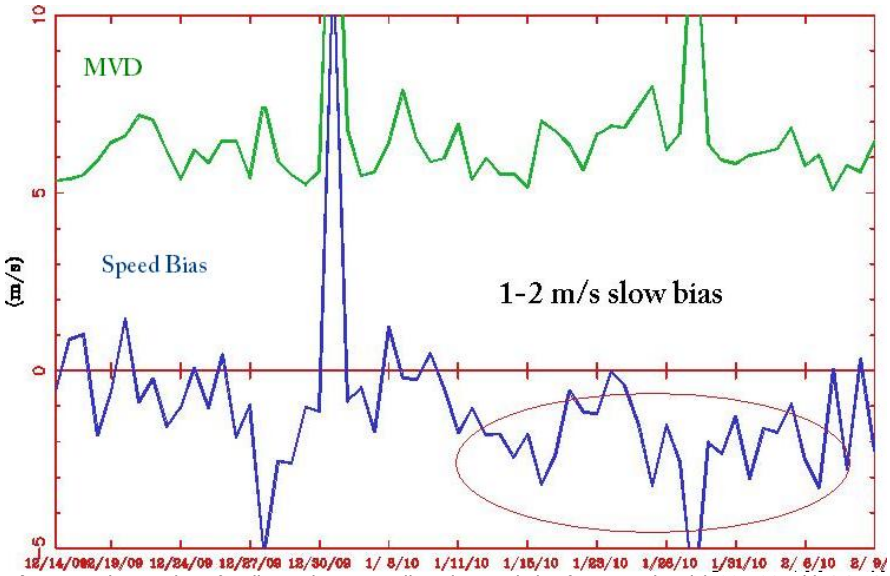

Figure 1: Time series of radiosonde-AMV collocation statistics for upper-level (100-400 mb) GOES-12 infrared winds. The time series covers the period from December 14, 2009 to February 9, 2010.

The authors of this paper have extended the results of earlier testing by studying the impact of varying loop intervals (5 to 30 minutes), as well as much smaller target box sizes (5 to 21 pixels), on the AMV verification statistics. The new results, discussed in Section 2, confirm the earlier findings of Sohn and Borde. At the same time they suggest a potential solution to the slow speed bias problem that combines the advantages of a smaller target box without sacrificing the higher overall quality afforded by using a larger target box.

The rest of the paper is divided as follows: Section 2 summarizes the testing performed with varying time intervals and target box sizes. Section 3 discusses the nested tracking approach in greater detail and presents specific case studies highlighting its application. Section 4 discusses the impact of the new approaches on the quality of the derived winds. Lastly, Section 5 summarizes the most important points made in this paper and discusses new possibilities yet to be explored.

2. TIME INTERVAL AND TARGET BOX SIZE TESTING

Earlier studies by Sohn and Borde (2008) have shown that a direct link exists between the target box size and the magnitude of the AMV slow speed bias. The authors showed that a smaller target box produces lower height assignments and faster winds, with both factors yielding a reduction in the slow speed bias when compared to a control run utilizing a much larger box size.

The above study was extended in this paper to include varying time intervals (5, 10, 15 and 30 minutes) in addition to varying the target box sizes (5x5, 9x9, 15x15 and 21x21 pixels). The data used in this study consisted of sequences of infrared imagery from Meteosat-8 centered at 00 and 12Z for the period June 1-8, 2008. In total, 16 processing configurations were tested (4 box sizes and 4 time intervals). Quality statistics were computed for each configuration by collocating the satellite winds to wind profiles from the radiosonde network. It is important to note that the current test was designed to be consistent with Sohn and Borde who kept the target locations fixed while they expanded the target box size incrementally about the same central point. A 6-hr forecast from the NCEP Global Forecast System (GFS) was used to position the search box in the subsequent image while the following constraint was used to determine the size of the search box,

$$(u-u_g) \leq \frac{(L-2)x}{2T} \qquad (1)$$

where **L** is the size of the correlation surface in pixels, **T** is the time interval in minutes, **u** is the AMV u-component, $\mathbf{u_g}$ is the forecast u-component, and **x** is the resolution in km. Equation 1 is a constraint on the maximum departure permitted from the forecast by the bounds of the search region. If the constraint is held constant then the size of search box can be computed. Here we use a loose constraint of 30 m/s. As a consequence of holding this constraint constant, the size of the search region was allowed to expand with increasing time interval. Testing was done on Infrared imagery from Meteosat-8 for the period June 1-8, 2008. Wind datasets were produced at 00 and 12Z and were subsequently compared to radiosonde profiles following recommendations of the Coordination Group for Meteorological Satellites (CGMS) (e.g., Velden and Holmlund 1998).

Figure 1 shows the impact of the various time interval and box size configurations on the speed bias relative to a control run using a time interval of 15 minutes and a box size of 15 pixels (same settings used operationally). This figure shows clearly that one way to minimize the slow speed bias is to minimize the target box size. The reduction is a result of slightly faster winds and lower heights (not shown). These trends are very consistent with the ones reported by Sohn and Borde. It is interesting to note that a larger time interval has a similar effect on minimizing the slow bias.

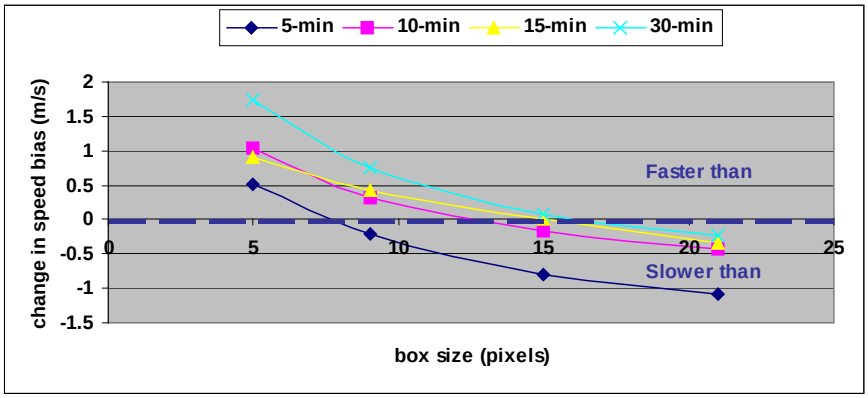

Figure 1: Change in speed bias relative to control for winds derived from Meteosat-8 infrared imagery valid at 00 and 12Z for the period June 1-8, 2008.

Figure 2 shows the impact of each processing configuration on the RMSE. This figure makes it quite clear that a smaller target box yields a noisier vector field. The impact is most pronounced with box sizes smaller than 9x9 pixels.

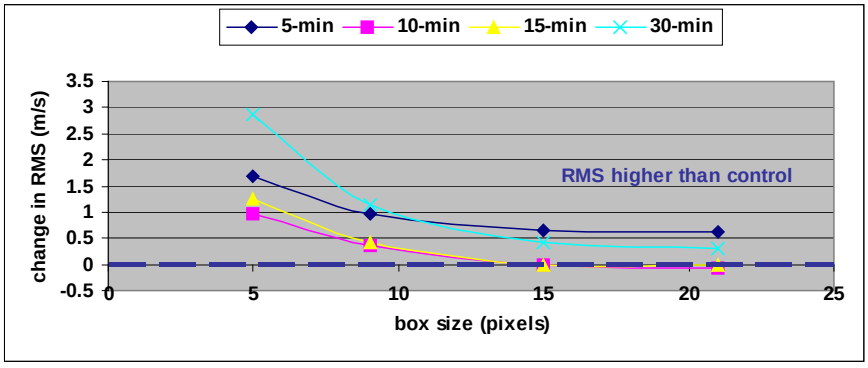

Figure 2: Change in RMSE relative to control for winds derived from Meteosat-8 infrared imagery valid at 00 and 12Z for the period June 1-8, 2008.

The preceding results imply that the speed bias problem cannot be addressed simply reducing the size of the target box. Ideally, one would like to reduce the bias without increasing the noise. In the next section we discuss an approach that accomplishes this goal.

3. NEW APPROACHES

3.1 NESTED TRACKING

The method proposed to minimize the slow bias involves nesting a smaller (5x5 pixels) box within the larger target scene so that a field of local motion vectors is derived over the interior pixels. A schematic of the approach is shown in Figure 3 alongside one example of the vector field produced by the approach. Differences in orientation and magnitude can arise between the local motion vectors if more than one cloud layer is being tracked or if multiple scales of motion are being detected. Outliers – those vectors that differ greatly from most of the sample – can result if the cloud is evolving or if the smaller box is insufficiently large to resolve the true motion, the so called aperture effect discussed at length in the field of computer vision (Trucco and Verri, 1998).

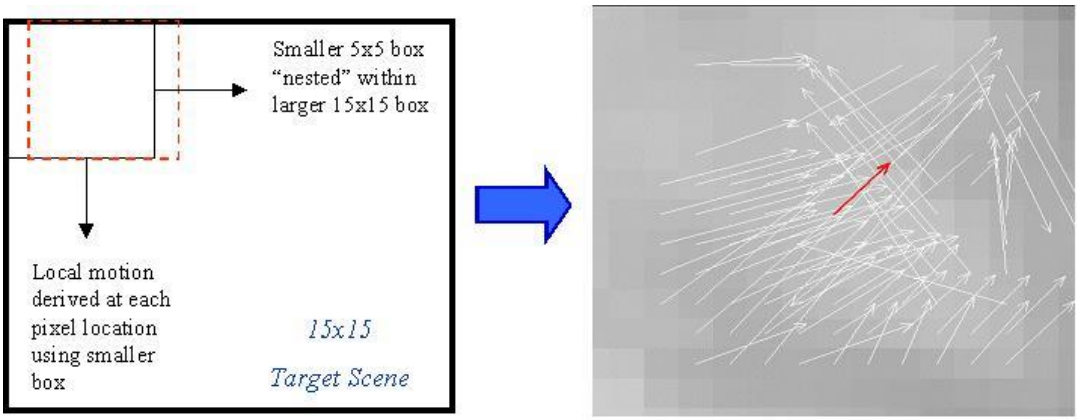

Figure 3: The nested tracking approach. The white vectors show the local motion derived with the 5x5 box centered on the pixel location. The average vector is shown in red.

Not surprisingly, the average local motion vector (in red) is a very close match to the vector derived by tracking the entire target scene (i.e., the larger 15x15 box). This close correspondence is highlighted in Figure 4 and confirms that the motion estimate from the larger box is an average of local motion within the scene. The differences that do exist in this figure may be caused by the lack of local motion vectors near the boundary, where a 2-pixel offset prevents a vector from being computed.

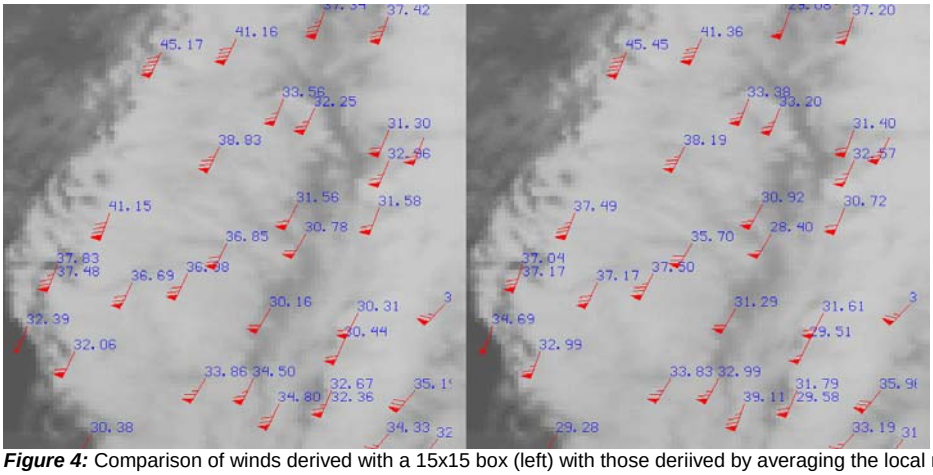

Figure 4: Comparison of winds derived with a 15x15 box (left) with those deriived by averaging the local motion. Speeds are plotted in blue (m/s).

Having shown the motion derived from the larger box is an average of local motion within the scene, it follows that varying motions can, and often do, produce a slower average wind. One way to prevent this from happening is to use a cluster analysis method to isolate similar motion then compute the average vector from the isolated cluster. A method for doing this is discussed in the next section.

3.2 CLUSTER ANALYSIS WITH DBSCAN

The justification for using a cluster analysis algorithm to analyze the local motion field is twofold. First, as was discussed in the previous section, the local motion field can be quite noisy. The field of vectors often reveals motion associated with two or more cloud layers and/or spatial scales. Removing noise and separating the sample into coherent motion clusters can prevent the excessive averaging that contributes to a slow speed bias. Second, identifying clusters in the local motion field provides a means for directly linking the tracking step with the height assignment step. In other words, the pixels belonging to the coherent clusters allow us to limit the sample used for height assignment to only those pixels important for tracking. This will be discussed further in Section 5.

In this paper, we selected a cluster analysis algorithm called DBSCAN (Ester et. al., 1996), a density-based algorithm for identifying clusters in spatial databases with noise. It was selected because it is very effective at identifying clusters of varying shapes and, unlike other methods such as k-means, does not require the user to specify, apriori, the number of clusters to find.

The first step in deriving the motion is to compute the local motion for the entire scene with the nested tracking approach. The initial sample of vectors is limited to those vectors having a high correlation score (0.8 or above). This was done to eliminate gross mismatches from the sample. Second, the initial sample of vectors (specifically the displacements) is analysed by DBSCAN and the coherent clusters are identified. The last step is to compute an average vector using only the displacements associated with the largest cluster. One example of output from DBSCAN is shown in Figure 5.

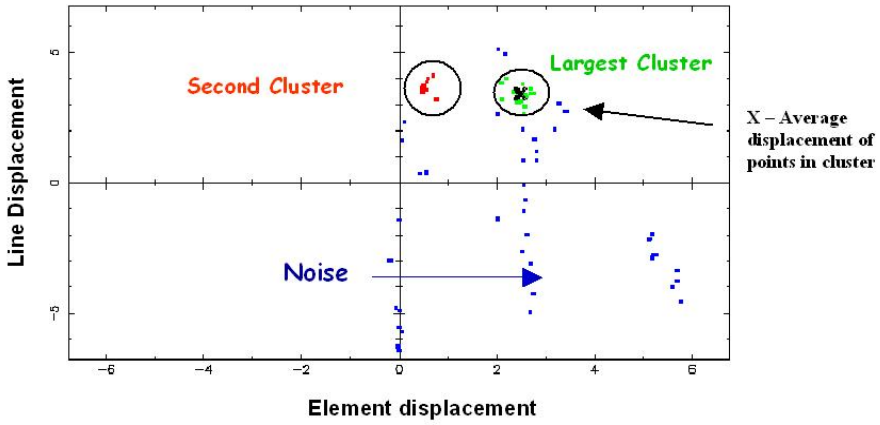

***Figure 5*:** Clusters identified by DBSCAN.

3.3 EXAMPLES

In this section we present examples utilizing the new techniques. The first example (Figure 6) is the accompanying vector plot for the DBSCAN example shown above. It highlights a fairly complex and evolving cloud scene at upper levels. Only the vectors associated with the largest cluster are plotted after the analysis is performed. Many of the local motion vectors have been correctly classified as “noise” and have been removed from the final output. There is also evidence to suggest that a second smaller cluster (shown in red in Figure 5) may actually be tracking the larger-scale motion of the entire cloud system. Clearly, the DBSCAN analysis has removed many of the vectors that “cancelled” one another in the initial sample. By doing so, it has sped up the initial average vector (in red) from 25 m/s to nearly 40 m/s (in green). Note that the magnitude of the new vector is in much better agreement with the forecast than the initial vector magnitude.

.

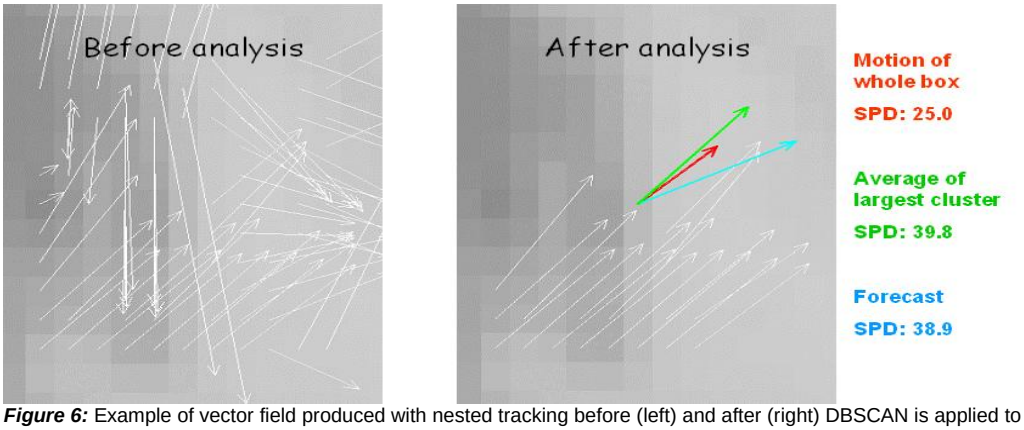

Figure 6: Example of vector field produced with nested tracking before (left) and after (right) DBSCAN is applied to find the largest cluster. The forecast vector (blue) is shown for comparison.

The first example was meant to show the effectiveness of DBSCAN in an extreme scenario. A much more common example is shown in Figure 7. Once again DBSCAN has correctly eliminated most of the noise, producing a new vector magnitude that is slightly faster than the initial magnitude.

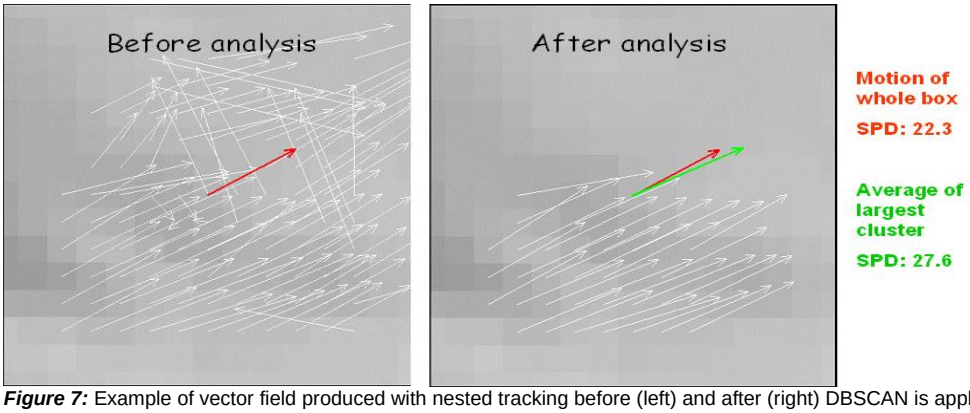

Figure 7: Example of vector field produced with nested tracking before (left) and after (right) DBSCAN is applied to find the largest cluster.

Both of the above examples highlight an important point – the points in the final sample are usually close to one another and are often found in regions of enhanced structure. This implies that the cloud heights in these regions vary significantly. Conversely, the coldest cloud tops are somewhat removed from the final sample locations. These observations argue against using the cold sample to assign a height to these tracers. The next section discusses a new approach to height assignment that exploits these observations.

3.4 LINKING HEIGHT ASSIGNMENT WITH TRACKING

In Section 3.2 we noted that another benefit of combining a cluster analysis algorithm with nested tracking was the ability to identify those pixels having the greatest impact on the tracking solution. This allows us, in a consistent way, to assign a height using only those pixels in the largest cluster, rather than using an arbitrary cold sample that may not be having a large influence on the tracking solution. The strategy followed in this paper was to assign a representative height to the AMV by taking the median cloud-top pressure of all points in the largest cluster. As one would expect, the median of the largest cluster usually produces a height assignment substantially lower in the atmosphere than a height computed from the cold sample of the entire target scene. A typical distribution of cloud-top pressure values is shown in Figure 8. As you can see from this figure, the pressure values of the largest cluster (green histogram) are spread across a fairly broad layer (approximately 75 mb) centered on 360 mb while the coldest pixels are confined to a narrower range

of approximately 25 mb centered at 310 mb. This is consistent with the observations made in Section 3.3 that the final sample of points is often located in a region of enhanced cloud structure and may not coincide with the coldest pixels.

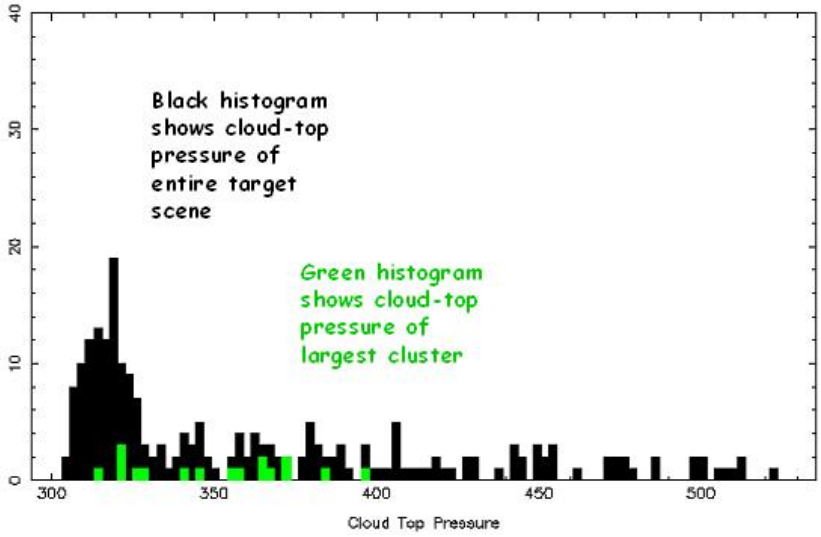

Figure 8: Cloud-top pressure distribution for a single target scene.

The impact of the new approach on the AMV height distribution is shown in Figure 9. In general, the new heights are significantly lower in the atmosphere than the old (cold sample) heights. This agrees with the observations made in Section 3.3 and is consistent with the individual target scene shown in Figure 8.

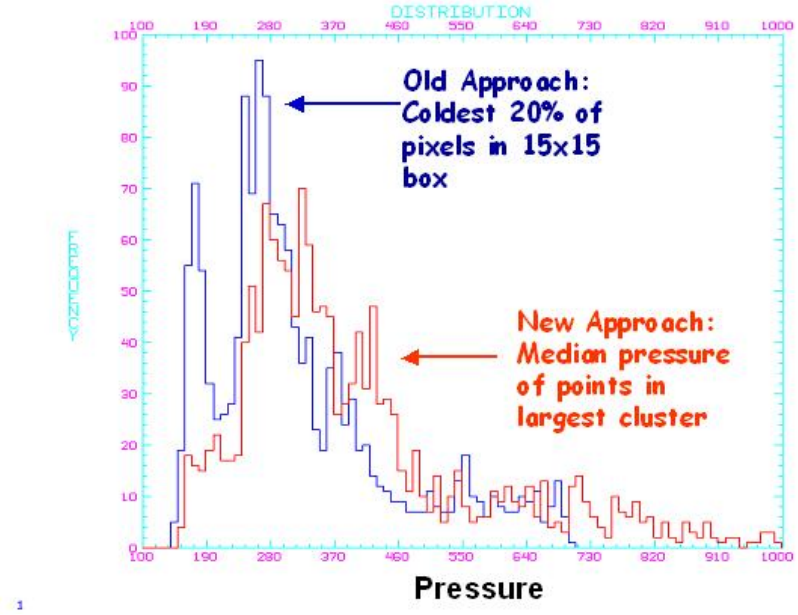

Figure 9: AMV height distribution using cold sample (blue) and the new approach (red).

4. IMPACT OF NEW APPROACHES ON QUALITY

To measure the impact of the new approaches on the quality of the AMVs the new methods were applied to a 2-month dataset of Meteosat full-disk infrared imagery and the winds were subsequently collocated with radiosonde profiles at 00 and 12Z. The two months processed were August 2006 and February 2007. A scatter plot of radiosonde speed to AMV speed for August 2006 (Figure 10) reveals that the test wind is a much better fit to the radiosonde wind.

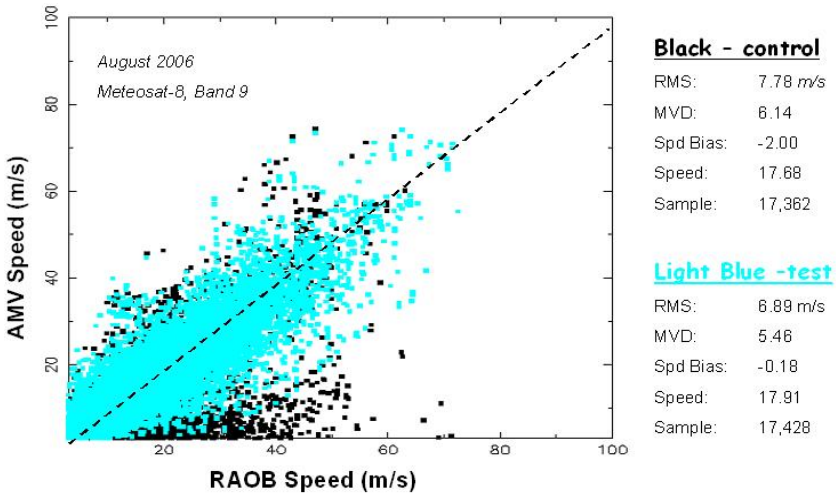

Figure 10: Scatter plot of radiosonde and AMV speed for August 2006 for the control (black) and test (light blue)

Statistics for February 2007 are displayed in Table 1.

	Control 15x15 box	**Test** New Approaches
RMSE	**7.53**	**6.63**
Average Difference	5.95	5.28
Speed Bias	**-1.97**	**0.06**
Speed	17.46	17.71
Sample	14548	14553

Table 1: AMV-radiosonde collocation statistics for February 2007.

5. SUMMARY AND FUTURE PLANS

Two new methods have been proposed to combat the slow speed bias commonly found in derived satellite AMVs at mid and upper levels of the atmosphere. Strictly speaking, one method is concerned with tracking while the other is concerned with height assignment, though the two are closely tied to one another. Nested tracking, the first new method proposed, involves nesting a small target box within a larger target scene to produce a field of local motion vectors derived over the interior domain of the larger target box. The field of vectors is analyzed with a cluster analysis algorithm (DBSCAN) to remove outlying motions and to isolate distinct motion clusters. The local motion vectors of the points in the largest cluster are then averaged to produce a final estimate. Case studies were shown to highlight the new tracking method. It is important to note that, although not shown here, some speed adjustments produce a slower AMV. Moreover, most speed adjustments, whether they are positive or negative, are small (on the order of 5 m/s). Still, in some extreme cases, the adjustments may exceed 10-15 m/s. The extreme cases were more likely to be positive adjustments (i.e., faster wind) than negative adjustments.

The second new method proposed in this paper relies upon the success of the cluster analysis program in identifying the largest motion cluster. The AMV is then assigned the median cloud-top pressure of the largest cluster. In general, the AMV height from the new approach is significantly lower in the atmosphere (roughly 40 mb) than the height produced by cold sampling. Combined, the two methods provide a consistent means of directly linking the tracking and height assignment processes together.

Comparison statistics based on two months of testing using Meteosat-8 SEVIRI imagery (one summer month and one winter month) show that the new methods substantially reduce, if not

eliminate, the AMV slow speed bias. The slow speed bias is minimized through a combination of faster winds and lower height assignments. The end result is a set of higher quality AMVs.

Plans for the future include implementing the new techniques into the current GOES processing at NESDIS, a potential NWP forecast impact study and plans to take a closer look at the information contained in the second and third largest clusters to see if cases exist where these clusters yield useful, even better, information than the largest cluster.

REFERENCES

Borde, R. and R. Oyama (2008): A Direct Link Between Feature Tracking and Height Assignment of Operational Atmospheric Motion Vectors. In Proceedings of the Ninth International Winds Workshop, Annapolis, Maryland, USA.

Ester, M., H.-P. Kriegel, J. Sander and X. Xu (1996): A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96), Portland, Oregon, USA, pp 226-231.

Forsythe, M. and R. Saunders (2008): Third Analysis of the Data Displayed on the NWP SAF AMV Monitoring Website. NWP SAF Technical Report 22, pp 1-50.

Genkova, I., R. Borde, J. Schmetz, K. Holmlund, J. Daniels and C. Velden (2008): Global Atmospheric Motion Vector Inter-comparison Study. In Proceedings of the Ninth International Winds Workshop, Annapolis, Maryland, USA.

Sohn, E. and R. Borde (2008): The Impact of Window Size on AMV. In Proceedings of the Ninth International Winds Workshop, Annapolis, Maryland, USA.

Trucco, E., and A. Verri (1998): *Introductory Techniques for 3-D Computer Vision*. Prentice Hall, pp 1-343.

Velden, C. S., and K. Holmlund (1998): Report from the working group on verification and quality indices (WG III). In Proceedings of the Fourth International Winds Workshop, Saanenmoser, Switzerland, EUMETSAT, pp 19-20.