Computing Morphing Vectors for Version 06 IMERG

Jackson Tan, George J. Huffman

18 March 2019, rev. 6 October 2020

The refereed publication that summarizes changes in V06 morphing is Tan et al. (2019).

1. Vector Source

In IMERG up through V05, the cloud motion vector computation approach used is that pioneered in CMORPH (Joyce et al. 2011), in which motion vectors are computed from 4-km geosynchronous infrared (GEO-IR) brightness temperatures. Hence, the motion vectors reflect cloud top motions. However, there are two main limitations in using GEO-IR. The first limitation is that cloud top motions may not match precipitation motions due to both wind shear and the growth and decay of precipitation systems. The resulting motion mismatch necessitated the use of climatological scaling factors based on US ground-based radars in CMORPH, although it is unclear how applicable these scale factors are to other regions. Moreover, this scaling only corrects for differences in mean speed and not differences in direction. The second limitation is that the GEO-IR dataset used in CMORPH and IMERG (the NOAA/NWS/CPC 4-km Merged IR) are confined to the latitude band 60°N-S. Poleward of this band, the viewing angle is too oblique to produce useful data outside 60°N-S. In addition, as we approached the start of V06 processing we found that issues with access to the necessary GEO-IR data would inhibit a complete retrospective processing.

Due to these drawbacks in using GEO-IR data, in IMERG V06 we modified the vector computation subcomponent of the morphing algorithm to derive motion vectors from numerical model variables. For the Final Run, the Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) reanalysis product is used (Gelaro et al. 2017); for the Early and Late Runs, the Goddard Earth Observing System Forward Processing (GEOS FP) forecast product is used (Lucchesi 2017). Both products are computed by the Global Modeling and Assimilation Office at NASA Goddard Space Flight Center using the GEOS model, which helps ensure consistency in the vectors between the Final Run and the Early and Late Runs.

MERRA-2 is generated based on version 5.12.4 of the GEOS model. MERRA-2 variables are available on a 0.5° latitude × 0.625° longitude grid with a temporal resolution of 1 h for most 2D

fields (3 h for most 3D fields). Its latency of about 1 to 2 months is compatible with the processing timeline of the Final Run.

GEOS FP is produced using the latest version of the GEOS model. GEOS FP variables are available on a 0.25° longitude × 0.3275° grid with a temporal resolution of 1 h for most 2D fields (3 h for most 3D fields). Being a forecast product, GEOS FP is run four times a day, each with a different forecast length: the 00Z forecast is run for 240 h, the 06Z forecast for 30 h, the 12 Z forecast for 120 h, and the 18Z forecast for 30 h. Vectors are always computed using variables from two consecutive hours from the latest available forecast, but the extended forecast lengths mean that vectors can still be produced—albeit with a lower reliability—should there be any interruption to forecast data availability. The latency of GEOS FP is currently about 7 h.

In this study, we evaluated the morphing scheme using vectors computed from six variables: total (surface) precipitation from atmospheric model physics (PRECTOT), total precipitable water vapor (TQV; also known as total column water vapor), total precipitable liquid water (TQL), total precipitable ice water (TQI), specific humidity at 500 hPa (Q500), and specific humidity at 850 hPa (Q850). TQV, TQL, and TQI are the vertically integrated water content in the three respective phases. Due to the lognormal distribution of precipitation, PRECTOT is first transformed by $\log(P + 1)$, where P is the precipitation rate in mm / h, to reduce the otherwise disproportionate weights given to high values in the correlation computation. Furthermore, to avoid anomalous vectors due to small-scale noise, minimum-value thresholds are applied to the model variable to include only grid boxes considered part of the precipitation system for the correlation computation. Based on preliminary trial-and-error studies, the following thresholds were selected: 0.03 mm / h (PRECTOT), 0.002 kg / m² (TQL), and 0 for all others. [These thresholds will be revisited in future work.] We chose to compute a vector only if there are at least 1,000 pixels (out of a maximum of 2,500 pixels in the 5°x5° template; see next paragraph) that are above the variable threshold for computing the correlation; otherwise, it is assigned as "missing". [In the previous GEO-IR-based morphing scheme, a vector is computed so long as there is at least one pixel above the threshold.] Of the six variables, only TQV and Q500 consistently give vectors at all locations; Q850 is globally complete except in areas of high altitude. In the following, we only show results for MERRA-2, since results for GEOS FP are similar.

The computation is similar for all variables, working on a large-scale template of 5°x5° on a grid of 2.5°x2.5°. Different template sizes and grid spacing have been explored, but differences

in performance were not sufficiently compelling to warrant a change for V06. However, as we will be extending the vector computation from the GEO-IR domain of 60°N-S to fully global, the distortion of the longitudinal size of the template at high latitudes is an issue. In polar regions the area of a 5° × 5° template is much smaller than the same template at the Equator, leading to inconsistency in the scale the vectors represent. [This was less of a problem in the GEO-IR-based scheme, since the area differs by no more than 50% within the 60°N-S latitude belt.] To mitigate this distortion, the template is given a latitudinal size of 5° and a longitudinal size of 5° / cos(latitude). Furthermore, at the poles, the template "wraps" around to the other longitudinal side of the hemisphere. Note that these fixes are driven by the fact that IMERG's native grid is cylindrical equidistance (CED; or equal-angle). We expect to investigate alternative "equal area" native grids for data processing to avoid this issue in future versions of IMERG, even while maintaining the 0.1° CED for end users.

One important criterion for the choice of vector source is its temporal resolution. Our experience suggests that a higher temporal resolution is more important than higher spatial resolution in computing the vectors. This limits the range of the reanalysis products that are viable for this purpose. We have explored using ERA5 as the data source, but the vectors produced are highly similar to MERRA-2, even though the hourly data have a native 31 km grid size and the computational approach is independently developed. This demonstrates that the vector computation method—which is actually based on the large-scale motion of the variable and not its exact values—is not sensitive to the precise dataset used.

In the GEO-IR scheme, data are provided on a 4-km-equivalent CED grid, which gives fine-scale features and large sample sizes within the 5° × 5° template. With the lower spatial resolution of the numerical analysis products, we choose to bilinearly interpolate the variables to 0.1° × 0.1° to increase sample size. Development work indicated that the absence of fine-scale features does not appear to be an important factor in the motion vectors using the 5° template. However, this could become a factor if we choose to compute vectors on a finer scale in the future.

2. Propagation

Since the MERRA-2 and GEOS FP products are at the hourly resolution, the vectors computed from successive fields represent changes from one hour to the next. Referring to the example in Fig. 1 (top), vectors computed from the 10:00–11:00 field and the 11:00–12:00 field are used to propagate *hourly* precipitation from 10:00–11:00 to 11:00–12:00. However, since IMERG has a

resolution of half-hour, we need to temporally interpolate the vectors to the same resolution. This is done in two parts, as shown in Fig. 1 (bottom). First, to propagate the *half-hourly* precipitation from 10:30–11:00 to 11:00–11:30, we simply divide the magnitudes of the hourly vectors from 10:00–11:00 and 11:00–12:00 by two. Second, to propagate the *half-hourly* precipitation from 11:00–11:30 to 11:30–12:00, we first average the *hourly* vectors from 11:00–12:00 and 12:00–13:00. This produces vectors at every half-hour.

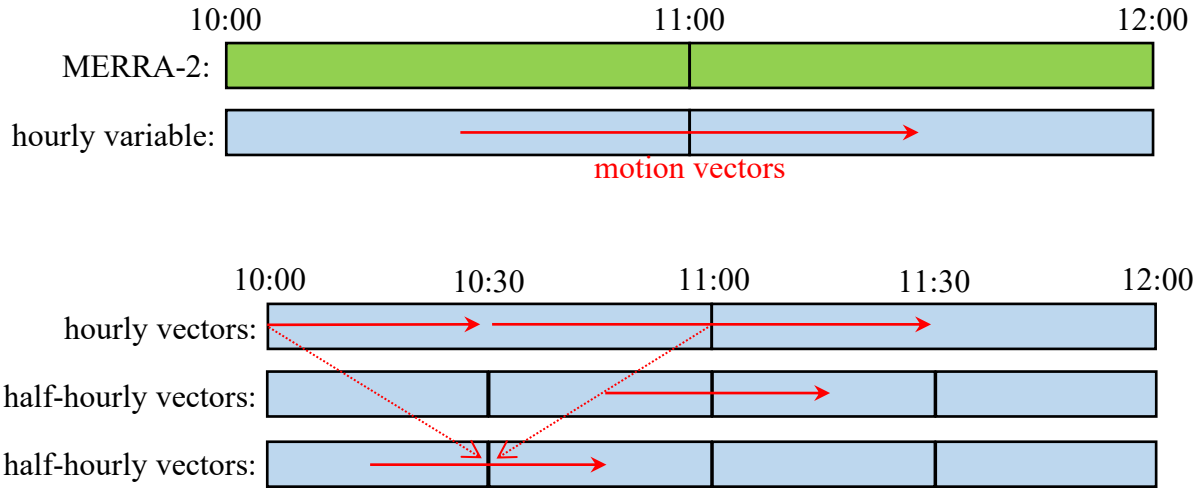

Fig. 1 Illustration of computing (top) hourly motion vectors from MERRA-2 hourly data and (bottom) half-hour motion vectors for use in IMERG.

Spatially, the vectors are computed at every 2.5°. Note that if they had been computed at every 0.1° grid box, the computational demand would be 25 × 25 = 625 times higher. In the original CMORPH GEO-IR-based vector computation, the precipitation pixels are propagated with the same vector across the entire 2.5° block, i.e., the motion of a 0.1° precipitation pixel is based on the closest 2.5° vector. In the new scheme, we use bilinear interpolation from the surrounding 2.5° blocks to obtain the vector at the 0.1° grid box (Fig. 2), allowing for smoother motions.

The climatological scaling factors derived for the IR-based vectors applied in the original CMORPH scheme are no longer necessary, as the motion vectors are now based on numerical model estimates. As in the original CMORPH treatment of converging/diverging vectors, if more than one pixel is propagated into the same destination, the average of all pixels is taken. If no pixel(s) is(are) propagated into a grid box, the box is filled by spatial interpolation. This is done

by searching for the neighboring grid boxes in increasing radius (i.e., the 8 grid boxes surrounding that grid box, then the 16 grid boxes surrounding those 8, and so on), stopping when at least three non-missing values are found. If at least half of these non-missing values are non-zero, their average will be used to fill in the missing value in the grid box; if less than half of these non-missing values are nonzero, the missing value is set to zero. This spatial interpolation scheme is a slight change from the practice in CMORPH, where all boxes are always averaged, and it is an attempt to balance between slightly increasing the area covered by precipitation and preserving the global-mean precipitation rate. As with the vector computation, pixels propagated beyond 90° latitude will be wrapped around to the other longitudinal hemisphere. This consideration is not necessary in the original CMORPH, since it only covers 60°N-S.

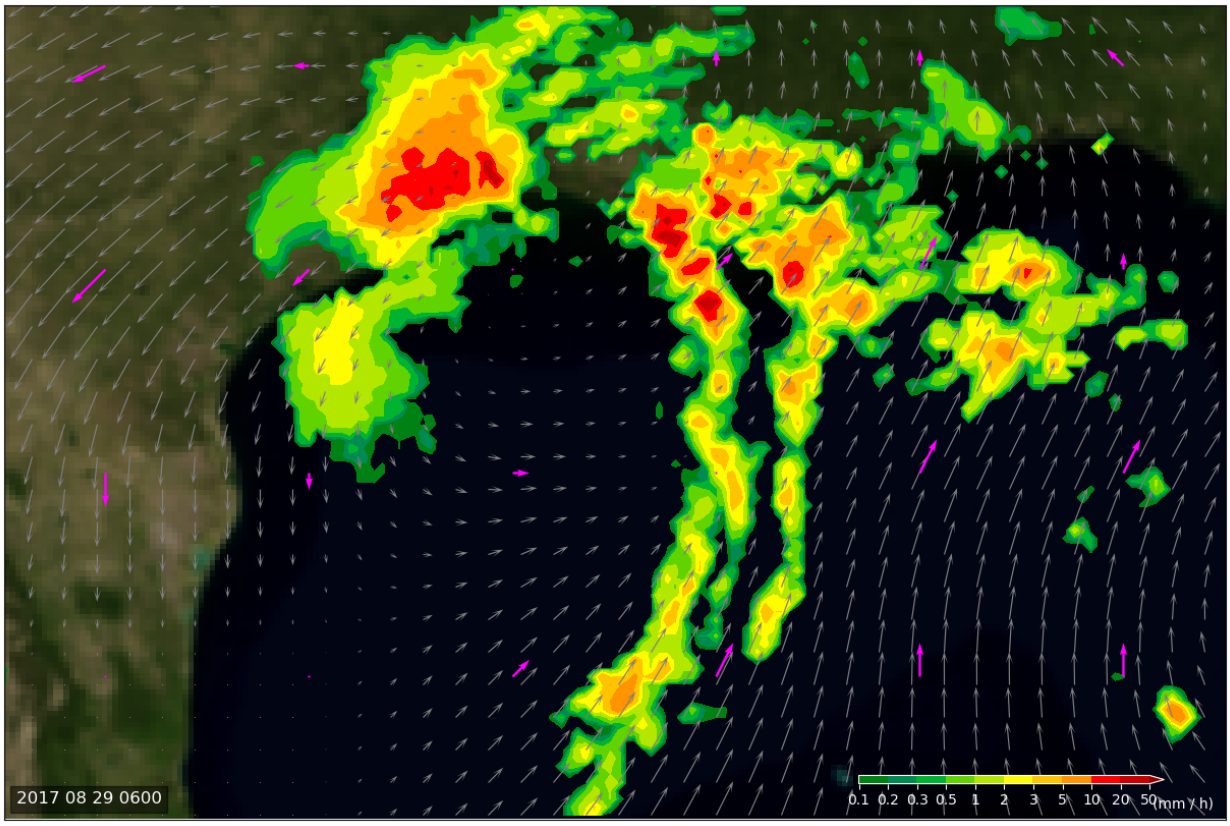

Fig. 2. Example of original (2.5° grid) and interpolated (0.1° grid) propagation vectors for 0000 UTC on 29 August 2017 in the western Gulf of Mexico, together with the rainfall associated with Hurricane Harvey.

3. Evaluation of the New Morphing Scheme

To evaluate the performance of the morphing scheme using vectors derived from different model variables, we propagate actual merged fields of passive microwave (PMW) precipitation estimates (the IMERG HQ field) from one half-hour to the next and compare it with the available observed PMW merged field in that next half-hour. This allows us to compare the accuracy of the propagation globally. However, inter-sensor differences and evolution of the precipitation systems means that the propagated precipitation will never be identical to the observed PMW even if the propagation is perfect; that is, the evaluation will not give perfect scores. Instead, a meaningful evaluation of a variable's performance requires comparison against benchmarks. We have chosen two: the GEO-IR-based morphing scheme as implemented in V05 ("IR") and no motion ("None").

We use three metrics to assess performance: Heidke Skill Score (HSS) for a 0.2 mm/h threshold, Pearson correlation coefficient ("correlation"), and normalized root mean square error (NRMSE). The evaluation presented here uses three months of data from August 2017 to October 2017.

3.1 Global Average

Fig. 3 shows the HSS, correlation, and NRMSE of all propagated precipitation across 60°N-S (limited to the latitudinal extent of the GEO-IR field). Precipitation propagated using vectors derived from MERRA-2 variables outperform both the GEO-IR-based propagation in V05 and no motion ("None"). In fact, total precipitable water vapor (TQV) has the best performance, higher than model precipitation (PRECTOT), total precipitable ice water (TQI), and 850 hPa specific humidity (Q850). In addition to giving the best results, vectors derived from TQV are globally complete; i.e., there is a vector defined at all locations and spatial or temporal interpolation is not required to fill in gaps, other than as a result of divergent vectors.

3.2 Zonal Average

Fig. 4 shows the evaluation as a function of latitude. Again, TQV consistently demonstrates superior performance of all the variables tested, often scoring higher than the IR-based vectors of V05. The improvement over "None" is most salient over the mid-latitudes. Interestingly, in the tropics and subtropics, TQI—and occasionally IR—perform more poorly than "None"; in other words, the introduction of morphing actually worsens the performance. This likely stems from (i)

the different motions between the cloud tops or vapor concentrations and the underlying precipitation, and (ii) the smaller motions in lower latitudes than in mid-latitudes.

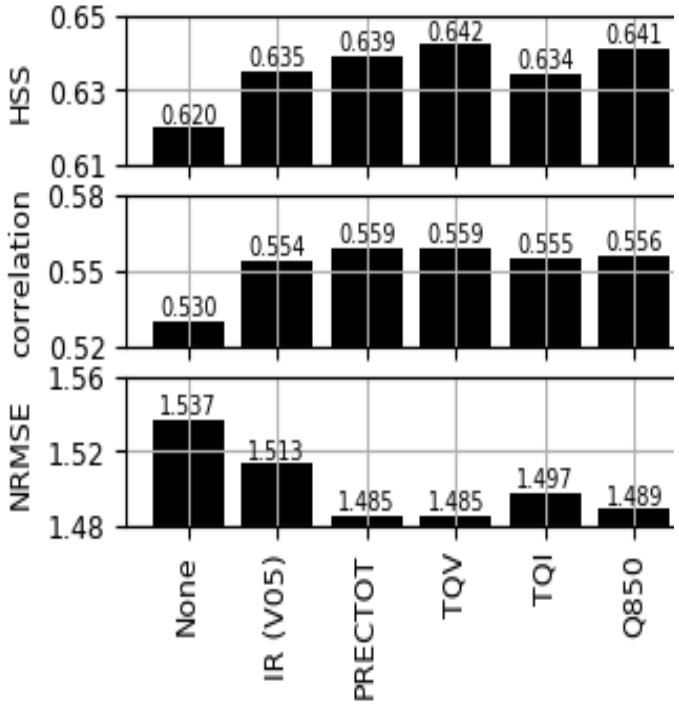

Fig. 3. Evaluation of precipitation propagated for one half hour using vectors from different sources; statistics are averaged over the band 60°N-S.

4. Summary

IMERG V06 introduces a new model-based morphing scheme in which motion vectors are computed from model variables instead of GEO-IR Tb observations. Testing suggests that motion vectors derived from TQV (total precipitable water vapor) outperform both the GEO-IR-based vectors employed in V05 and those computed from other variables. In addition, TQV gives globally complete vector fields, largely eliminating the need for spatial or temporal interpolation to fill holes.

Evaluation will continue after the release of V06 to assess if other variable(s) can give an additional edge in performance over TQV. The start of V07 will provide the next opportunity to implement changes in the vector source.

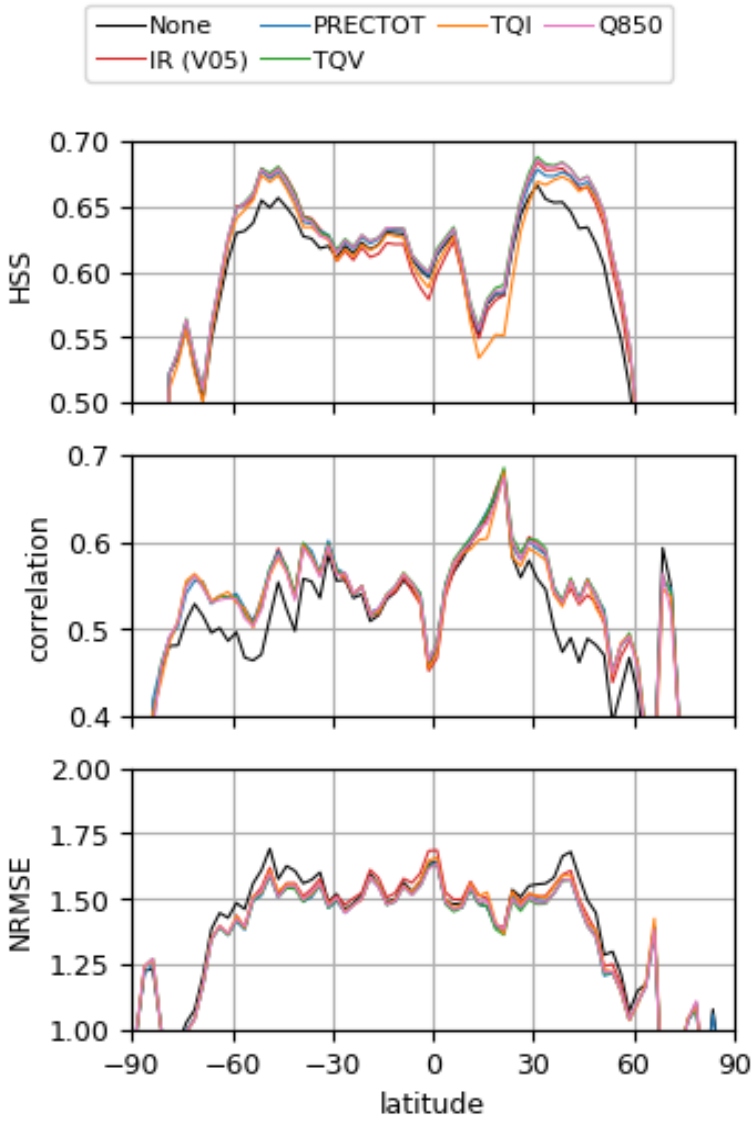

Fig. 4. Evaluation of precipitation propagated for one half hour using vectors from different sources; statistics are shown as latitude averages.

5. References

Gelaro, R., W. McCarty, M.J. Suárez, R. Todling, A. Molod, L. Takacs, C.A. Randles, A. Darmenov, M.G. Bosilovich, R. Reichle, K. Wargan, L. Coy, R. Cullather, C. Draper, S. Akella, V. Buchard, A. Conaty, A.M. da Silva, W. Gu, G.-K. Kim, R. Koster, R. Lucchesi, D.

Merkova, J.E. Nielsen, G. Partyka, S. Pawson, W. Putman, M. Rienecker, S.D. Schubert, M. Sienkiewicz, B Zhao, 2017: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J. Climate.*, **30**, 5419–5454. *doi:10.1175/JCLI-D-16-0758.1*.

Joyce, R.J., P. Xie, J.E. Janowiak, 2011: Kalman Filter Based CMORPH. *J. Hydrometeor.*, **12**, 1547-1563. *doi:10.1175/JHM-D-11-022.1*

Lucchesi, R., 2017: File Specification for GEOS-5 FP. GMAO Office Note No. 4 (Version 1.1), 61 pp. *https://gmao.gsfc.nasa.gov/products/documents/GEOS_5_FP_File_Specification_ON4v1_1.pdf* .

Tan, J., G.J. Huffman, D.T. Belvin, E.J. Nelkin, 2019b: IMERG V06: Changes to the Morphing Algorithm. *J. Atmos. Ocean. Technol.*, **36**(12), 2471-2482. *doi:10.1175/JTECH-D-19-0114.1*