Conical-scanning Microwave Imager/Sounder Algorithm Theoretical Basis Document

Volume 1: Overview
Part 3: Imagery EDR and Gridding

Version 1.6 – 19 April 2005

Submitted by:
Atmospheric and Environmental Research, Inc.
131 Hartwell Avenue
Lexington, MA 02421-3126

With contributions by:
Janusz Eluszkiewicz, John Galantowicz, Thomas Nehrkorn,
Ross Hoffman, Alan Lipton, and Jean-Luc Moncet

Prepared for:
Boeing Satellite Systems
919 Imperial Avenue
El Segundo, CA 90245

AER Document P1130-TR-I-ATBD-OVERVIEW-IMG-V1.6-20050419

This page intentionally left blank.

REVISION HISTORY

Version	Release Date	POC	Comments
1.0	11 Jan., 2001	Galantowicz	Initial partial draft release.
1.1	6 Feb., 2001 (PDR)	Galantowicz	Minor edits.
1.2	22 Feb., 2001	Galantowicz	Filled sections for imagery and gridding, added appendices C and D.
1.3	29 Oct., 2004	Galantowicz	Created separate documents for Footprint Matching (vol.1 part 2) and Imagery and Gridding (vol. 1 part 3).
1.4	10 Dec., 2004	Eluszkiewicz	Added implementation and performance for recursive filter.
1.5	18 Feb., 2005	Galantowicz	Added implementation and performance for earth gridding
1.6	19 Apr., 2005	Galantowicz	Modified Imagery EDR descriptions and overall CMIS processing flow figure.

RELATED CMIS DOCUMENTATION

Title	Version	Authorship	Date
CMIS SRD for NPOESS Spacecraft and Sensors	3.0	Associate Directorate for Acquisition NPOESS IPO	2 March 2001

Title	Description
CMIS ATBD	Multi-volume document covering TDR, SDR, and EDR algorithms and code documentation. See “CMIS ATBD Volume 1: Overview, Part 1: Integration” for the complete list of associated volumes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. Introduction

Three interrelated CMIS algorithm tasks—footprint matching, gridding, and imagery generation—deal with the spatial properties of sensor-sampled data, EDR inputs, and EDR products. *CMIS ATBD Vol. 1 Part 2 Footprint Matching* covers footprint matching and this document covers gridding and image generation.

- Footprint matching is the process by which multiple sensor samples distributed horizontally on the earth's surface (or, more generally, on a surface defined at any height relative to a reference geoid) are used to produce a single composite sample at or near a particular location (interpolation) and with a particular spatial weighting pattern (pattern matching). For our purposes, the composite footprint location is always defined relative to the along-scan and along-track coordinate system of the sensor samples on the geoid. Typically, the objective of footprint matching for the retrieval of a given EDR is to create composite footprints with horizontal spatial resolution (HSR) equal to the EDR horizontal cell size (HCS) for each channel used in the EDR's derivation.

- Gridding is the process by which one geophysical field is transformed from its sampled coordinates to another coordinate set. Examples include 1) regridding of slant-path atmospheric profile retrievals with altitude-dependent coordinates to the coordinates of an EDR reporting grid and 2) regridding of satellite scan/track-coordinate retrievals to a fixed earth-grid (map) coordinate system. Our approach to regridding includes analysis of the expected spatial properties of the geophysical field sampled and the measurement's estimation errors and spatial properties.

- Imagery generation is a CMIS requirement through the imagery EDR. The content of the imagery EDR is brightness temperature data with properties consistent with the retrieval of the other EDRs and suitable for display. Imagery data at the CMIS sampled resolution is available through the SDR products (see *CMIS ATBD Vol. 17 TDR and SDR Algorithms*). Since all EDRs are retrieved based on composite footprint brightness temperatures, composite TBs will also be provided for each channel and composite footprint size combination used in the generation of the EDR. For example, the Land Surface Temperature EDR's HCS is 50 km and channels at 10 GHz and up may be used in its derivation. The imagery product will therefore include ~50 km HSR composite footprint brightness temperatures at each of these channels sampled in the scan/track coordinate system—that is, the brightness temperatures will be located along the main conical arcs scanned by the sensor.

The purpose of this document is to provide all the information necessary to understand, operate, further develop, and use the products of the CMIS gridding and image generation algorithms. Section 2 deals with gridding and section 3 with the imagery EDR. We described the relevant instrument characteristics and the algorithms' requirements, historical background, mathematical formulation, and predicted performance measures.

2. Gridding

The regridding of the core module output occurs at the post-processing stage and includes three main processes:

1. Vertical remapping (VR) to produce atmospheric vertical temperature profile (AVTP) and atmospheric vertical moisture profile (AVMP) EDRs satisfying prescribed reporting requirements.

2. Composite field of view (CFOV) interpolation to produce a background (and first-guess) state vector at a fine spatial sampling scale from more coarsely sampled Core Module outputs. This is needed as part of the cascade algorithm integration design described in *CMIS ATBD Vol. 1 Part 1*. A swath of Core Module outputs at larger CFOV locations are interpolated to the locations of a finer-resolution swath of CFOVs, such as interpolating 40-km CFOV retrieved products to provide background data inputs to 25-km CFOV retrievals.

3. CFOV interpolation to a fixed Earth grid, to provide input to land and ice EDR algorithms.

In all these cases, the regridding is performed on irregularly spaced input data, with the irregular sampling arising because of the satellite viewing geometry and the occurrence of missing data, such as in precipitation areas where the Core Module does not produce retrievals.

2.1. Background Perspective on Interpolation Algorithms

The problem of interpolation of data present on either a grid, or irregularly spaced locations, to a set of grid points has been the subject of study for a long time in meteorology. Typically, fields are undersampled in meteorology, and diverse methods have been devised to spread information to all grid points in the neighborhood of observations. A review of a number of different methods is given in Daley (1991) and Thiebaux and Pedder (1987). We only briefly sketch the different approaches here.

2.1.1. Surface Fitting

Early methods were based on surface fitting, in which functional forms (e.g., polynomial expansions) were fitted to the available observations. Functional fitting can be done either locally (different functions are used for different analysis grid points) or globally (a single functional form is fitted to all observations in the domain of interest). With either approach, there can be serious problems of underfitting (fewer degrees of freedom in the functional form than in the data, leading to a poor fit to the data) or overfitting (more degrees of freedom in the functional form than in the data, leading to a poor analysis in data-sparse areas).

2.1.2. Distance Weighting Schemes

This encompasses a large number of techniques that compute analyzed values from a weighted average of surrounding observations. The weights are specified *a priori* based on the distance between the analysis grid point and the observation. Typically, these techniques employ successive passes, with decreasing length scales of the weighting functions. The two most commonly used techniques are the Cressman (1959; also called the successive correction

method) and Barnes (1964) schemes, which differ in the functional form of the weighting functions (the Barnes scheme uses a Gaussian weighting function).

In their original formulations, the Cressman scheme makes use of a background field (an *a priori* estimate from another source such as climatology or a short-term forecast), whereas the Barnes scheme does not.

The choice of the adjustable parameters of these techniques has been the subject of extensive study, and methods have been devised for the Barnes scheme to select weighting functions based on either the average data spacing (Koch et al. 1983) or the correlation length scale of the field to be analyzed (Seaman 1989).

There are well-known limitations to these distance-weighting schemes: anisotropic distributions of observation can lead to unrealistic analyzed values at the edge of data swaths, since the weight given to the observations does not take into account the observation locations relative to each other.

2.1.3. Statistical Interpolation and Variational Techniques

More recently developed methods explicitly take into account the error statistics of the observations (and background field). In statistical interpolation, the weights given to the observations are determined by minimizing the estimated analysis error. In variational techniques, an analyzed field is derived by minimizing a cost function that measures the appropriately weighted misfit to the observations and the background. These methods are usually implemented in the context of an NWP data assimilation system, and require a much larger computational and scientific infrastructure than simpler regridding techniques.

Lorenc (1992) compared the theoretical basis and performance of statistical interpolation with those of iterative schemes (like the successive correction method). He showed that, under certain conditions, iterative schemes can approximate the results of the statistical interpolation method. He further showed the near-equivalence between employing a recursive filter on the analysis increments and explicitly taking into account the covariance of the background error. We describe the recursive filter in more detail below.

2.2. Algorithm Descriptions

2.2.1. Recursive Filter

The recursive filter is described in (Hayden and Purser, 1988) and (Hayden and Purser, 1995), hereafter referred to as HayP88 and HayP95. It was especially designed to provide a computationally efficient interpolation method capable of producing realistic results for datasets with spatial inhomogeneities of coverage.

Its basic computational steps for a single pass of the analysis may be summarized as follows:

- Background values are interpolated to observation locations using bilinear interpolation.
- The observation increments (observed value minus interpolated background) are then spread to the surrounding four grid points using the adjoint of the bilinear interpolation operator.
- The resulting field of increments is then smoothed through repeated application of a digital filter.

It can be shown that, in the implementation used in HayP95, the spectral response of this filter asymptotically approaches that of the Barnes analysis, but with a spatially varying length scale that depends on data density.

HayP95 provide examples and guidance for choosing the adjustable parameters of the filter. The filter has been widely used in the processing of satellite data; it has also been used in variational assimilation systems in the estimation of the background error covariance matrix.

2.2.1.1 Formulation

The analysis scheme is similar to a successive correction method, where at iteration $n+1$ the analysis values $A(n+1)$ are computed as

$$A(n+1) = A(n) + \frac{G * \left[W\left(O - A(n)\right)\right] + W_b\left[A_b - A(n)\right]}{G * W + W_b} \tag{1}$$

where O indicates observed values, W is the product of quality and observation weights, W_b is the weight given to the background field, and $A_b = A(0)$ indicates the background (or first guess) field. This formula is a generalization of Equation (13) in HayP95, accounting for the deviations of the analysis field from the background field with the weights W_b. The operator $G*$ indicates the distribution and smoothing of values from the observation points to the analysis grid points.

The operator $G*$ is implemented in two stages: an interpolation step and a smoothing step. In the interpolation step, values at observation points are distributed to the neighboring four analysis grid points using the adjoint of a bilinear interpolation formula. Contributions from all observations are added to each applicable grid point.

The formula for the sum in one dimension at a single grid point is given by Equation (12) of HayP95

$$\begin{aligned} X_i &= \sum_{k:\left|dx_k\right|<\delta}^{K} \left(1 - \frac{\left|dx_k\right|}{\Delta}\right) \hat{X}_k \\ \hat{X}_k &= \hat{W}_k \left(\hat{O}_k - \hat{A}_k\right) \end{aligned} \tag{2}$$

where dx_k is the distance between observation k and the grid point, and Δ is the grid spacing (only observations within Δ are considered in the sum). The values of A at the observation points k are obtained by bilinear interpolation from the surrounding analysis points.

An entirely analogous operation is performed for the quality weights W (more on the definition of W below). The smoothing step, in a horizontal implementation, is a two-dimensional smoother that is applied one or more times, both to the weights W and weighted residuals $W(O-A)$. The filter consists of a forward and reverse filter applied in both horizontal dimensions.

The fundamental filter equation for one dimension is given by Equation (1) of HayP95

$$A_i^{'} = \alpha A_{i-1}^{'} + (1-\alpha) A_i, \quad 0 < \alpha < 1 \tag{3}$$

where A is the input field, A' the output, and $\beta=(1-\alpha)$ is the smoothing parameter that controls the spatial scale of the filter. The corresponding equation for the reverse filter is given by HayP95 Equation (4)

$$A_i^{''} = \alpha A_{i+1}^{''} + (1-\alpha) A_i^{'} \tag{4}$$

The result of L iterations of the filter (3) and (4) asymptotically approaches that of a single application of a Gaussian filter given by HayP95

$$G_j \approx G_o \exp\left[-\frac{|j|^2}{2L(\lambda\Delta)^2}\right]$$
$$G_o \approx \frac{1}{\left[2\pi\left(L\lambda^2\right)\right]^{1/2}} \tag{5}$$

with length scale R given by HayP95 equations (10) and (11):

$$R^2 = 2L(\lambda\Delta)^2 \tag{6}$$

$$R^2 = \frac{2L\alpha\Delta^2}{(1-\alpha)^2} \tag{7}$$

The L and Δ are constants, R is prescribed for the analysis pass, and β=(1-α) is obtained by inverting (7). As can be seen from Equation (2) in HayP95, small values of β, i.e., large values of α, cause the analyzed field at gridpoint i to be affected by data at gridpoints removed many grid spaces from i. In other words, small β's imply substantial mixing across several grid spaces (e.g., in data-poor areas) and thus can be used as a diagnostic of input data density (see Section 2.5.3). The variable scaling is obtained by allowing the R to be defined separately for each pass and for each grid point. Complete details are given in HayP95.

The observation quality weights W are defined as the product of an *a priori* reliability estimate (between 0 and 1, initialized to a nominal value of 1 in the absence of ancillary information about data quality), and a quality estimate based on a scaled difference between the observation and the analysis from a previous pass. See HayP95 Equations (18) and (19) for details.

2.2.1.2 Implementation

The top-level flowcharts for the vertical remapping and FOV interpolation modules are shown in Figure 2-1 and Figure 2-2 respectively. The two basic steps in each module are segmenting of the input data and the application of the recursive filter on each segment. The segmenting procedure is described in Appendix A.

For each processing segment, the recursive filter (RF) is applied quasi-horizontally at each output pressure level (for the VR) and for each element of the state vector (for FOV interpolation). As described in Appendix A, the size (length and width) and spacing Δ of the analysis grid are driven by reporting and accuracy requirements. For FOV interpolation, the horizontal coordinates of the input data used in the analysis are the surface latitudes and longitudes, transformed to a quasi-rectangular coordinate system centered along the orbit track. For VR, the horizontal coordinates are the surface coordinates, shifted at each pressure level by a "parallax" amount accounting for the satellite slant-path viewing geometry, and transformed to the ground-track-centered coordinate system.

The computation of the parallax shift involves integration of the hypsometric equation to obtain the altitude at each pressure level:

$$z(p) - z_s = R_g \int_p^{p_s} \frac{1}{g} T_v d \ln p \tag{8}$$

where R_g is the gas constant, g is the gravity (computed as a function of latitude, longitude, and height), T_v is the virtual temperature, z_s is the surface altitude, and p_s is the surface pressure. For pressure levels below p_s, the hypsometric equation is integrated starting from $z_s = 0$ and p_s = lowest pressure level.

Once $z(p)$ is calculated, the latitude and longitude at each pressure level are computed from the surface latitude and longitude and the slant-path Earth incidence angle. These pressure-dependent coordinates are then transformed to the new coordinate system centered on the satellite track (the details of the coordinate transformation are described in Appendix A) and used as the horizontal coordinates of the input data used by the RF.

Once the analysis grid and the coordinates of input data ("observations") are defined, the RF is applied with tuning parameters adopted largely from HayP95 (Table C1 in their Appendix C), except for the grid length Δ (see Appendix A), smoothing parameter f, and tolerance Tol used in the definition of the quality estimate.

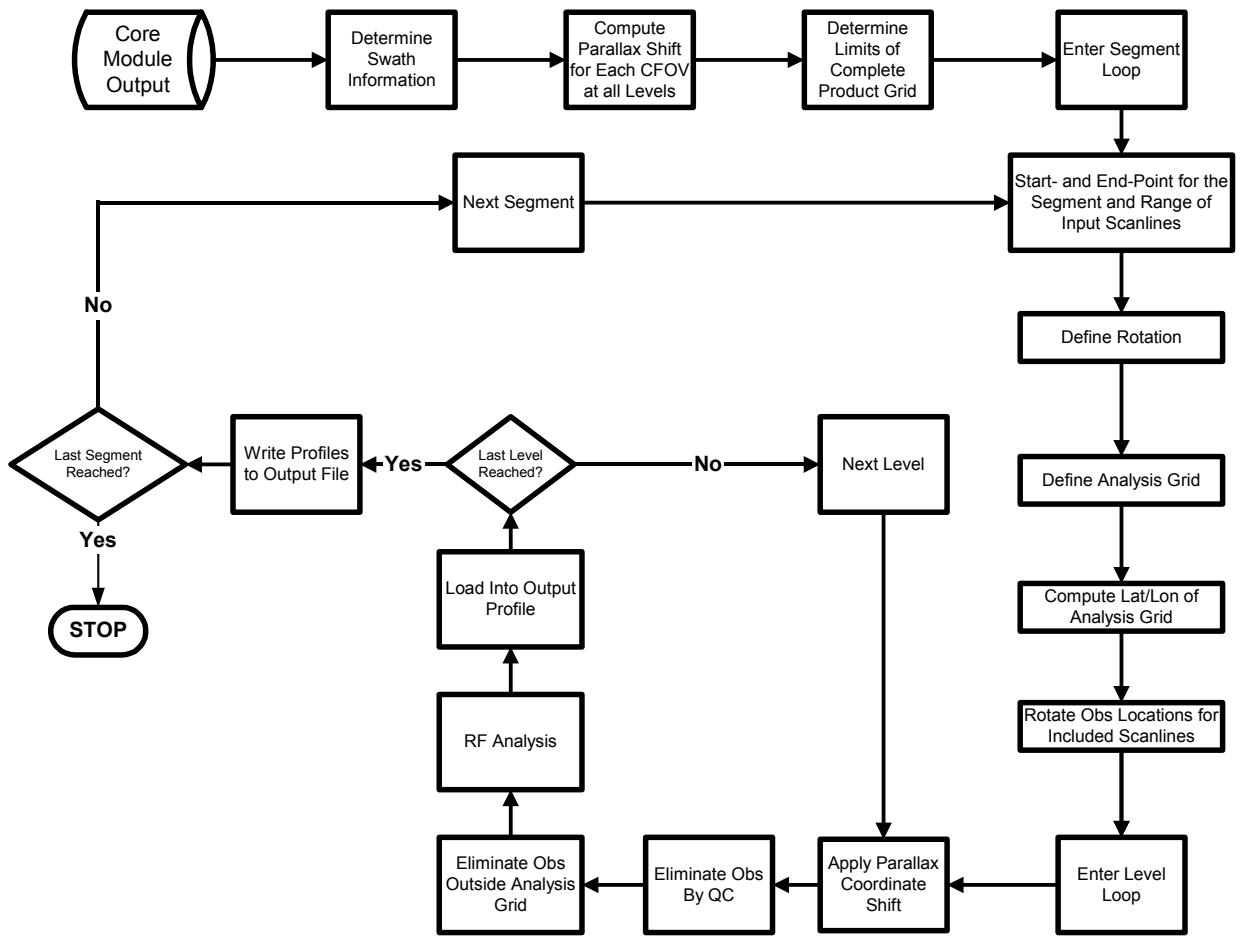

Figure 2-1: Flowchart for vertical remapping.

The basic steps of the RF analysis are as follow:

1. *Select filter parameters*: Select Δ based on the required EDR reporting grid spacing (for VR) and accuracy and timing requirements (for VR and FOV interpolation), and use the default values for f (=1) and Tol (=1).

2. *Initial background analysis*: HayP95 recommend a preliminary background analysis in cases where a first-guess field is available (e.g., from an NWP analysis). In the implementation for CMIS gridding, it is preferable that this process be independent of any external data, so we chose to set the background to the mean value of the field computed from all observations available in the input scene.

3. *Perform RF analysis*: Using the mean value for each field as the prior analysis at each grid point, perform the RF analysis using five analysis passes (except for the 20- to 15-km cascade which uses 7 passes), with $L = 3$ iterations of the smoother at each pass. The characteristic scale of the filter is decreased from 6Δ to 0.9Δ over the five analysis passes (see Table B-1 below; the extra two passes for runs with seven passes employ characteristic scales of 18Δ and 9Δ, respectively), with all other parameters given in Table C1 of HayP95.

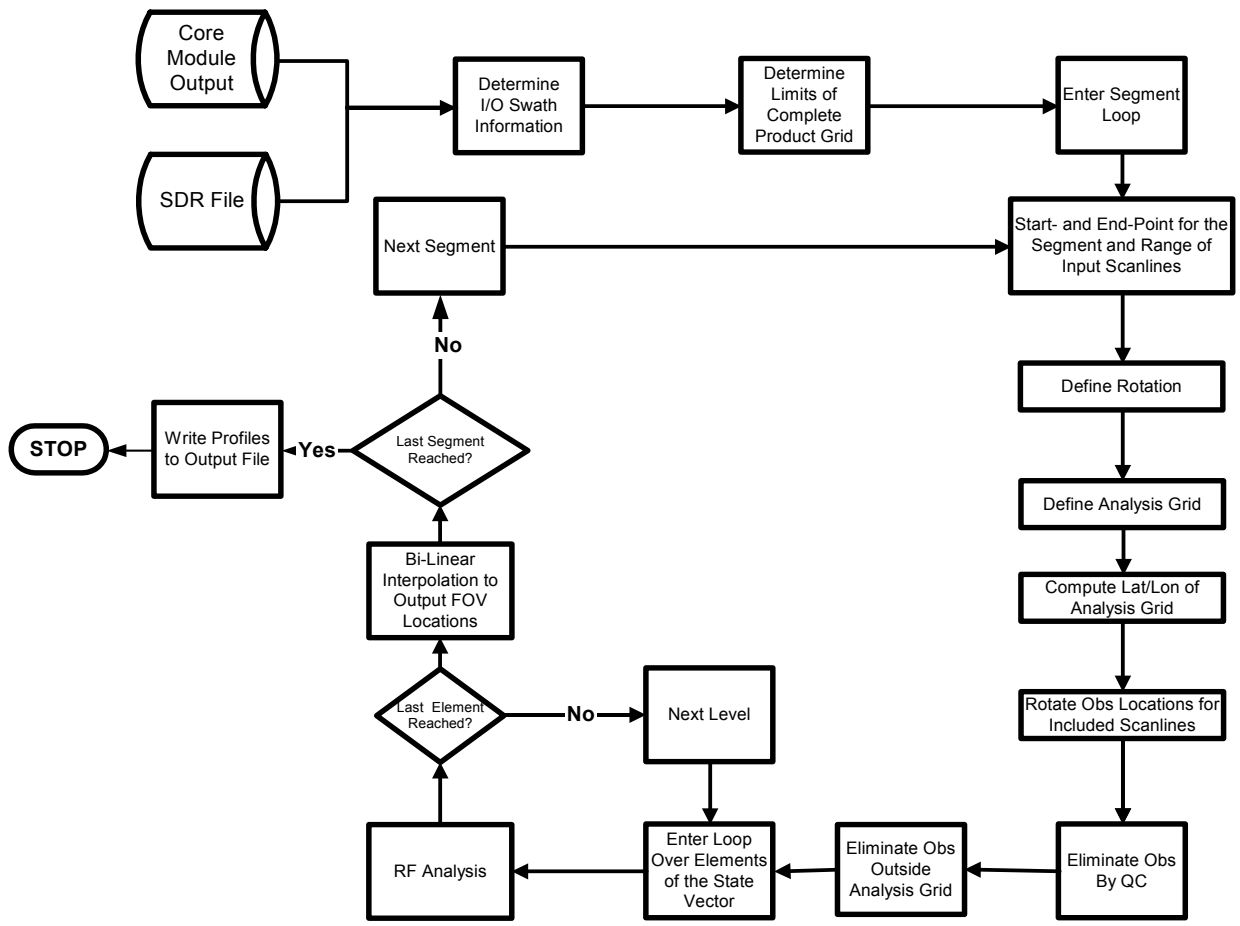

Figure 2-2: Flowchart for FOV interpolation.

Each analysis pass encompasses the following individual steps:

- Interpolate prior analysis to observation locations.

- Interpolate weight to observation locations.
- Evaluate $G*W$ and $G*(W(O-A))$. This involves both the interpolation (2) and smoothing steps [L iterations of (3) and (4)].
- Update the analysis using (1).

In the case of FOV interpolation, the execution of the RF places the data on a fine-resolution quasi-rectangular grid, and is followed by a bi-linear interpolation step to go from that grid to the locations of the output data, which have a conical scan pattern.

The rectangular grid should be fine enough that the error introduced by the bi-linear interpolation step is negligible, after the more demanding (longer distance) interpolation task has been performed by the RF step.

2.2.2. CFOV to Earth Grid Remapping

The CFOV to earth grid remapping algorithm encompasses a family of non-iterative distance-weighting formulations. Each method produces an estimate at a given earth grid point from a weighted-average of CFOV observations or retrievals from the surrounding neighborhood. None of the methods need either the grid or CFOV swath to regularly spaced or orthogonally oriented. Earth grid remapping will be applied to surface parameters such as emissivity and land surface temperature (LST) that may have significant spatial variability at the scales of CFOV averaging and interpolation beyond the edges of a sampled area cannot be expected to be accurate. Therefore, all the methods are implemented such that only grid points in regions covered by a minimum density of qualified CFOV swath data are filled. Section 2.4 gives preliminary tuning results for each method using spatially complex scenes. Method selection and final tuning will be needed for each EDR and Core Module output parameter to be remapped.

2.2.2.1 Nearest-Neighbor Resampling

The simplest of the methods, nearest-neighbor resampling populates each grid point with the nearest qualifying point on the CFOV swath. Like the other methods, the algorithm loops over each CFOV and finds the grid points for which that CFOV is a candidate data source. After the entire swath section has been processed, the algorithm loops over all grid points that were close to at least one CFOV and selects the closest CFOV to each. Quality flags attached to the CFOV data can be used to reject CFOVs at the discretion of the end-product application. The candidate CFOVs must also be within a maximum distance threshold D_{nn}, set to the lowest value guaranteed to fill all grid points within the CFOV data swath when there is no missing data, i.e.:

$$D_{nn} \geq \frac{1}{2}\left(\Delta y_{scan}^2 + \Delta x_{sample}^2\right)^{1/2} \tag{9}$$

where Δy_{scan} is the along-track scan spacing and Δx_{sample} is the along-scan sample spacing. See *CMIS ATBD Vol. 1 Part 2 Footprint Matching* for a complete description of the CFOV scan patterns. Only grid points within D_{nn} of qualifying CFOVs are filled, including those at the swath edges; grid points around missing or unqualified CFOVs are only filled if they are within D_{nn} of another, good CFOV. Algorithm output includes the actual nearest-neighbor distance as a quality indicator as well as the actual geolocation of the nearest-neighbor CFOV. Applications may then use the original data geolocation instead of the grid locations if more precise geolocation fidelity is needed.

2.2.2.2 Distance-Based Interpolation

The remaining methods expand the neighborhood of CFOVs contributing to each grid point estimate and combine them using distance-based weights. They refine upon the nearest-neighbor approach by attempting to infer the inter-CFOV values of the underlying data field, and, as shown in section 2.4, they are generally capable of providing better estimates at the grid points than the nearest neighbors. Because they require a larger sampling of CFOV data than the nearest neighbor, the methods 1) are more susceptible to contamination by poor data or missing data drop-outs and 2) cannot be traced back to a single accurately geolocated CFOV in the same way that the nearest neighbor can be. In tuning each method, the number of grid points potentially affected by missing data is mitigated by keeping only the minimum number of contributing samples needed to produce the most accurate results. Furthermore, grid points just at or beyond the edge of the swath are skipped because the absence of surrounding data makes estimation of the underlying gradients unreliable.

The objective of the interpolation methods is to provide an estimate of the sampled parameter at grid point g as if a CFOV observation or retrieval had been made there. The estimate is given by the weighted sum of the N nearest CFOV observations:

$$\hat{O}_g = \sum_{i=1}^{N} W_i O_i \,. \tag{10}$$

For ***Inverse-Distance Interpolation***, the weights are:

$$W_i = \frac{1}{D_i} \tag{11}$$

where D_i is the distance from the ith CFOV to the grid point.

For ***Inverse Distance-Squared Interpolation***, the weights are:

$$W_i = \frac{1}{D_i^2} \,. \tag{12}$$

For ***Linear Interpolation***, the weights are

$$W_i = D_{\max} - D_i \tag{13}$$

where D_{max} is a tunable maximum distance parameter and W_i goes to 0 for $D_i > D_{max}$.

And for ***Gaussian-Weighted Interpolation***, the weights are

$$W_i = e^{-\ln(16) D_i^2 / D_{hw}^2} \tag{14}$$

where D_{hw} is a tunable parameter equal to the size of the circle at which W_{ifov} = 0.5, that is, $W_i(D_i = D_{hw}/2) = 0.5$.

2.2.2.3 Implementation

Implementation of the earth gridding algorithm presents fewer challenges than the recursive filter discussed in section 2.2.1.2. Algorithm performance will degrade near any swath edges so data segments must overlap but only by 3 scans with the baseline tuning (discussed below). Gridded quality control parameters include the distance to and geolocation of the nearest CFOV as well as the worst-case quality control parameters from among the contributing CFOVs.

2.3. Recursive Filter Algorithm Performance

HayP95 give several examples of the successful application of the filter:

- A surface temperature analysis over the continental US and adjoining oceans, which is a difficult field because of the rapid changes in observed values due to topography, and rapidly varying data density at the coast line
- An analysis of 500-hPa height over North America and the North Pacific using gradient wind estimates from geostationary sounder retrievals of temperature and moisture. Data density in this case was high in isolated patches, surrounded by large data-void areas.

We have explored two methods of evaluating the performance of the recursive filter. A theoretical approach applied to gridded meteorological scenes is described in Appendix B. The rest of this subsection is devoted to a more realistic approach tested on simulated orbital scenes.

2.3.1. Orbital Simulation Scenes

In order to the test the performance of the remapping program on realistic orbital scenes, we have generated simulated retrieval scenes (and corresponding validation truth) using the CMIS testbed. The atmospheric data that drove the simulations were taken from the innermost grid of very high-resolution mesoscale model forecasts created using the University of Wisconsin version of the Colorado State University Regional Atmospheric Modeling System (RAMS). These data were the same as had been used in the initial theoretical analysis reported in Appendix B, where the data are described. By using the same data for both analyses, we provide traceability between the results of the two methods.

The 3-D RAMS model output of surface and atmospheric profile data from the Hurricane Bonnie simulation, which covers a small cell domain (200×200 at 2.5-km resolution), were tiled by "matchbook" flipping to fill a 1122×1122 domain with a 3-km nominal resolution. The testbed simulated a CMIS sensor orbital path crossing this domain and vectors were drawn from each domain cell surface point to the sensor as an estimate of the path followed by CMIS-sensed radiation during sensor scans. Slant-path profiles were constructed for each domain point for calculating top-of-atmosphere radiances along CMIS lines-of-sight. Composite footprint-weighted spatial averages of the following parameters were then computed at each CMIS sample location to represent a simulated CMIS retrieval vector: RAMS model surface temperature, pressure, and elevation, land-water fraction, cloud properties, and air temperature and water vapor mixing ratio at each of 40 pressure levels. For verification data, composite footprint-weighted spatial averages were also computed for the vertical profile data. Composite footprints are approximately circular-Gaussian functions with half-peak widths between 15 and 50 km. For this work, 15-, 20-, 25-, 40- and 50-km footprints were used.

2.3.2. Seaming

As a practical matter, orbital data are divided into segments for execution of the RF, as discussed in Appendix A. The appendix describes the method to provide continuity along the “seams” between segments. One parameter used in that method is the boundary width B, which controls overlap of data during processing. The value of B adopted for the recursive filter should be $B \equiv \max(G, R)$, where G is the maximum expected size of the precipitation areas. To test the algorithm performance, we required an estimate of G.

Based on the NOAA15/AMSU retrievals of rain rate provided by NOAA's Office of Satellite Data processing and Distribution, an example of which is shown in Figure 2-3, we estimated *G* to be on the order of 300 km.

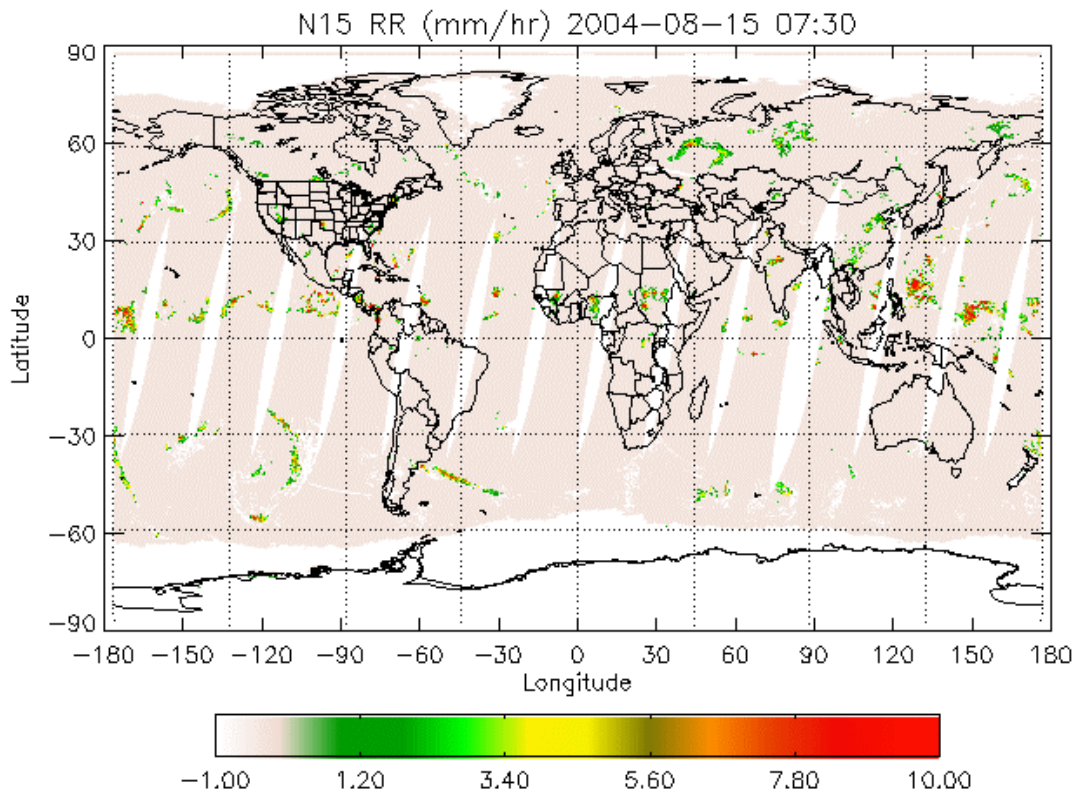

Figure 2-3: Rain rate retrieved from the NOAA-15/AMSU radiances for August 15, 2004 (figure obtained from http://www.osdpd.noaa.gov/PSB/IMAGES/MSPPS_day2.html).

The performance of the method's treatment of seam continuity is illustrated in a semi-quantitative way in Figure 2-4, which shows scatterplots of interpolated 1-mb temperatures at the boundary between the first and second segments in a vertical interpolation run with the 15-km scene. In this run, the lateral offset between the two segments has been set to zero so that the RF products along the last row of the first segment should be nearly identical to the RF products along the first row of the second segment. Figure 2-4 shows the differences in temperatures between those two rows.

The two panels in Figure 2-4 correspond to B = 300 and 50 km, i.e., $B > R$ and $B < R$, respectively (in this case R = 90 km). The seaming properties are clearly better in the first case (which corresponds to our default tuning), with the differences being very small in relation to the dynamic range of temperature along the seam. These results are consistent with the theoretical properties of the RF, which predict that the boundary effects extend a distance R from the boundaries.

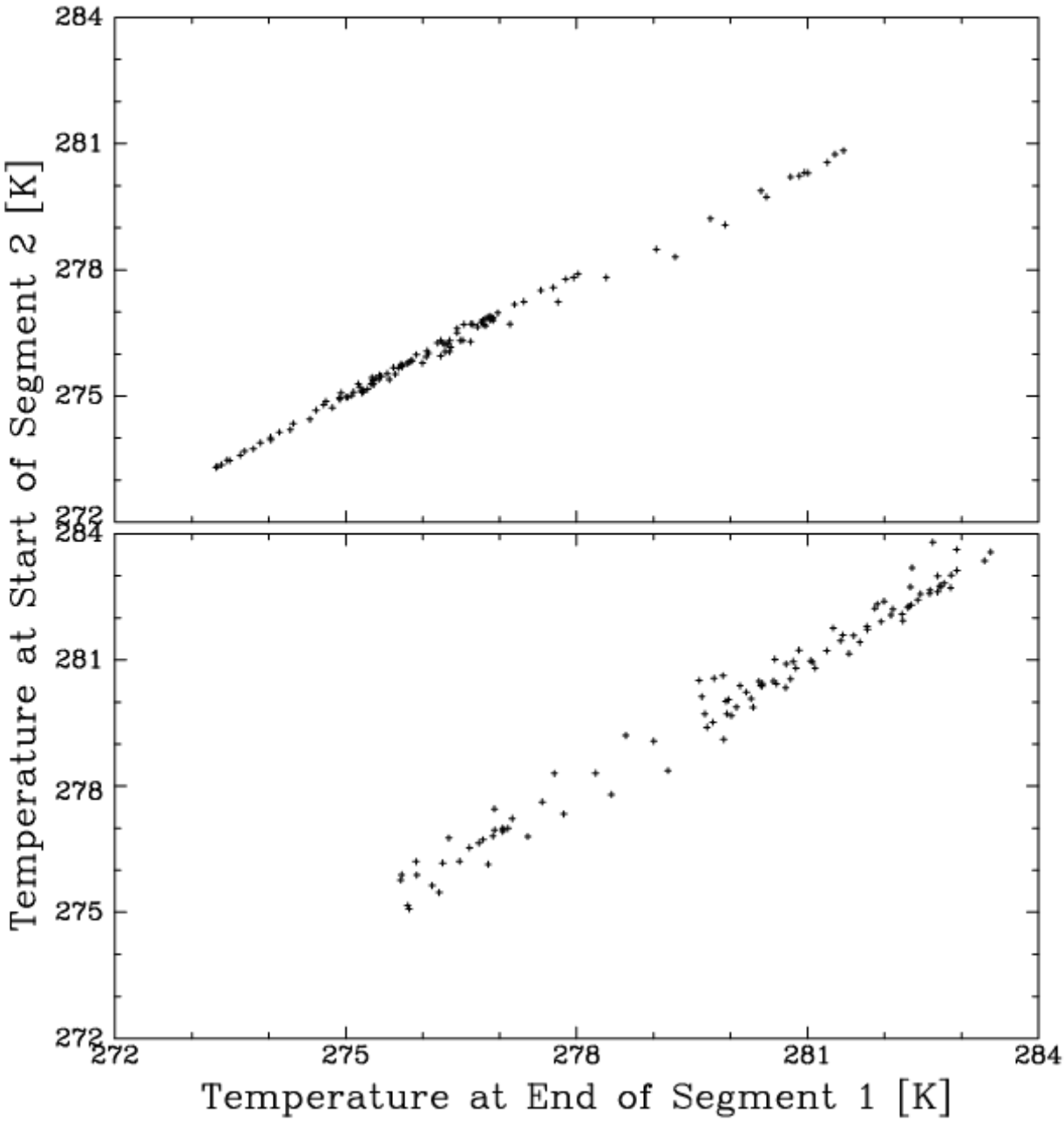

Figure 2-4: Scatter plots of interpolated 1-mb temperatures in the last column of the product grid for segment 1 versus temperatures in the first column for segment 2. The upper and bottom panels correspond to *B* = 300 and 50 km, respectively.

2.3.3. Vertical Remapping

The performance of the vertical remapping program has been assessed separately for the two primary vertical profile EDRs, AVTP and AVMP. The assessments were done under circumstances with a high density of Core Module retrievals that passed quality control: that is, without gaps in Core Module products as would occur where precipitation is present.

The VR algorithm will produce interpolated products in such gaps, but the EDR performance requirements exclude areas where precipitation significantly degrades the Core Module radiometric inputs and retrieval outputs. The assessment of the VR algorithm, therefore, is based on its performance in the areas where the EDRs themselves will be assessed against their performance requirements. The VR algorithm provides quality metrics, derived from the RF algorithm, to flag the gap areas where EDR quality is degraded.

2.3.3.1 Atmospheric Vertical Temperature Profile

The AVTP EDR is produced at 40-km resolution from the 40-km slant-path data. The test slant-path data from Bonnie are shown in Figure 2-5 for the 200-mb level.

The regridded product generated from this input is shown in Figure 2-6 while Figure 2-7 shows the "truth" field generated by a direct averaging of the vertical profiles contained in the high-resolution RAMS scene. It should be noted that because of the buffer zones applied to each analysis segment (see Appendix A), the domain of the regridded product is significantly smaller than the domain covered by the swath data. It is clear that the regridded field captures the variability of the input field (allowing for the parallax shift) and matches the true field very closely. As shown in Figure 2-8, the difference between the regridded and true fields are indeed very small (generally much less than 0.1 K), except near the swath edges where the boundary effects cause the differences to be larger.

The regridded field shown in Figure 2-6 has been generated using a spatial resolution of $\Delta = 5$ km for the regular grid employed by the recursive filter and subsampling this grid every 8 points in both horizontal directions to generate a 40-km resolution for the output product. This subsampling approach has been adopted when it was discovered that a straightforward application of $\Delta = 40$ km, together with the default values for the radius of influence R_m (expressed relative to the grid spacing, see Table B-1) produced substantial RMS errors. These errors can be significantly reduced when the values of R_m/Δ are scaled down from their defaults, but even in this case the performance is still somewhat degraded compared to runs employing small Δ and subsampling.

Figure 2-9 shows RMS errors for temperatures that have been layer-averaged as per the AVTP vertical cell size requirements (1-km layers from surface to 300 mb and 3-km layers above). Results are given for four cases: Case 1: $\Delta = 40$ km, R_m/Δ parameters reduced by a factor of 16 from their default values (shown in Table B-1), Cases 2-4: default R_m/Δ parameters with $\Delta = 10$, 5, and 2.5 km, respectively. In Cases 2 through 4, the 40-km output has been generated by subsampling the finer resolution analysis grid. The smallest errors are generated when the highest resolution ($\Delta = 2.5$ km) is adopted, although the differences between $\Delta = 2.5$ km and $\Delta = 10$ km are very small. Small errors can also be obtained when $\Delta = 40$ km is employed, but the radius of influence is reduced at each analysis pass by a factor of 16 to be consistent (in absolute numbers) with the values employed for the $\Delta = 2.5$ km run. However, the performance in this case (Case 1 Figure 2-9) is still systematically worse than in runs employing smaller Δ (0.06 vs 0.04 K in the mid-troposphere). On the other hand, as discussed in Section 2.5.1, computationally expense increases with decreasing Δ, as the timing of the recursive filter is proportional to the total number of points in the analysis grid and is thus inversely proportional to Δ^2. This speed-versus-accuracy trade needs to be taken into account when choosing the optimal value of Δ.

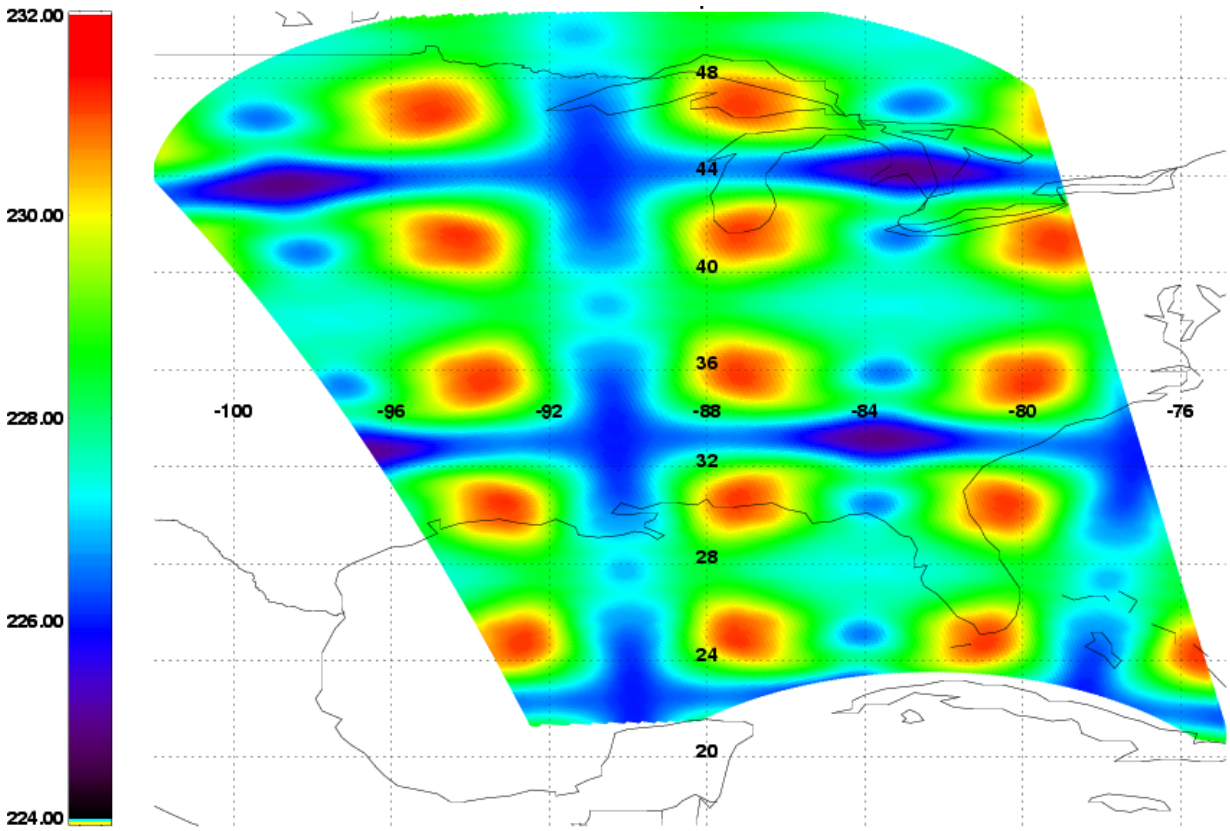

Figure 2-5: 40-km FOV resolution slant-path temperature field at 200 mb serving as input to the vertical remapping of the AVTP field.

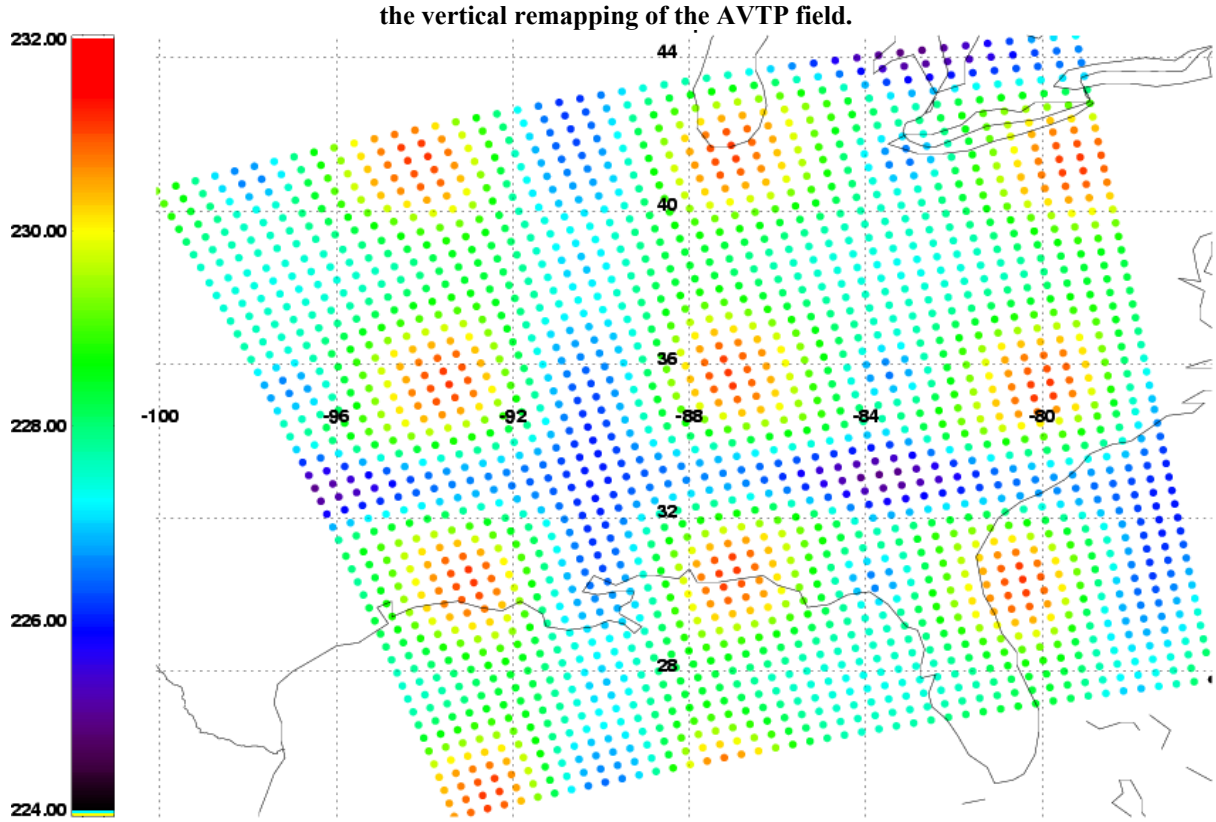

Figure 2-6: Vertically-remapped temperature field at 200 mb.

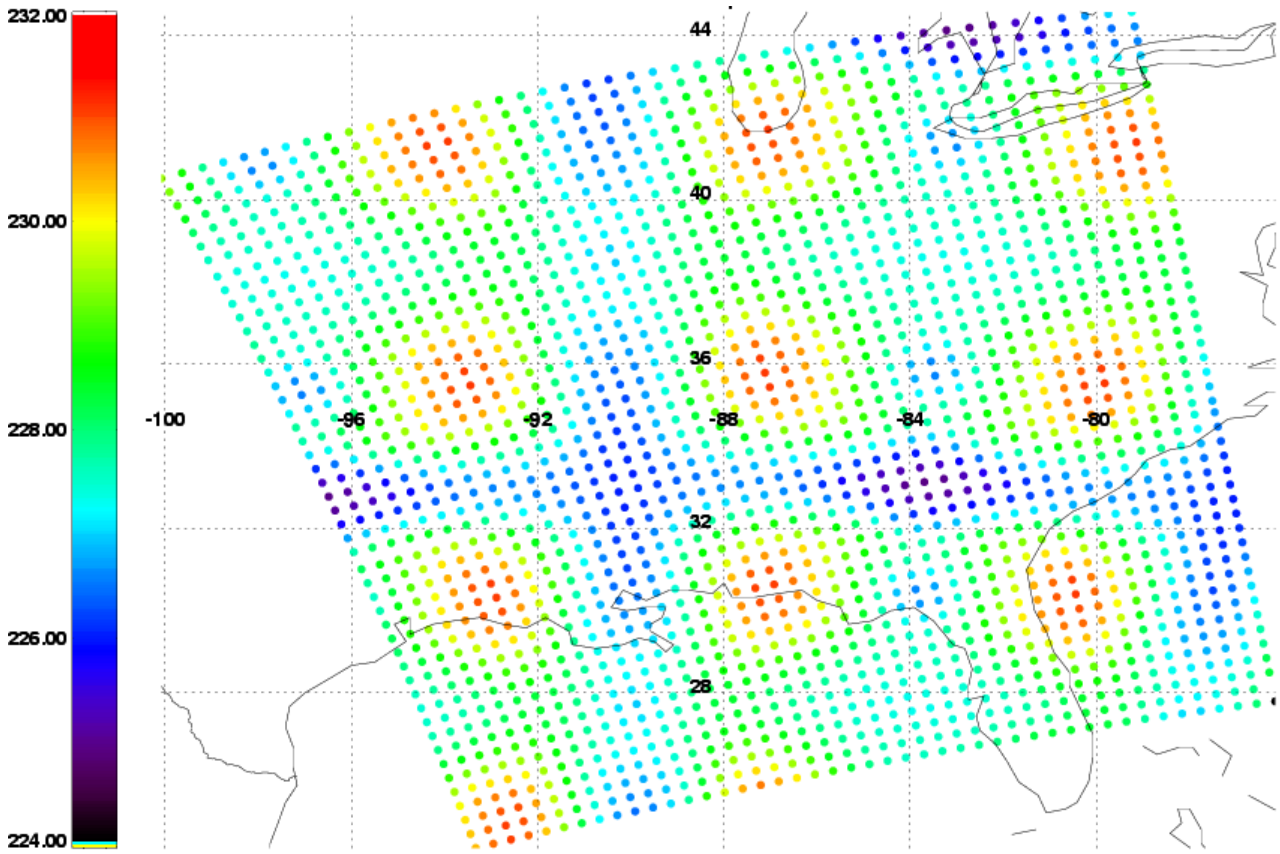

Figure 2-7: "True" vertical-path temperature field at 200 mb.

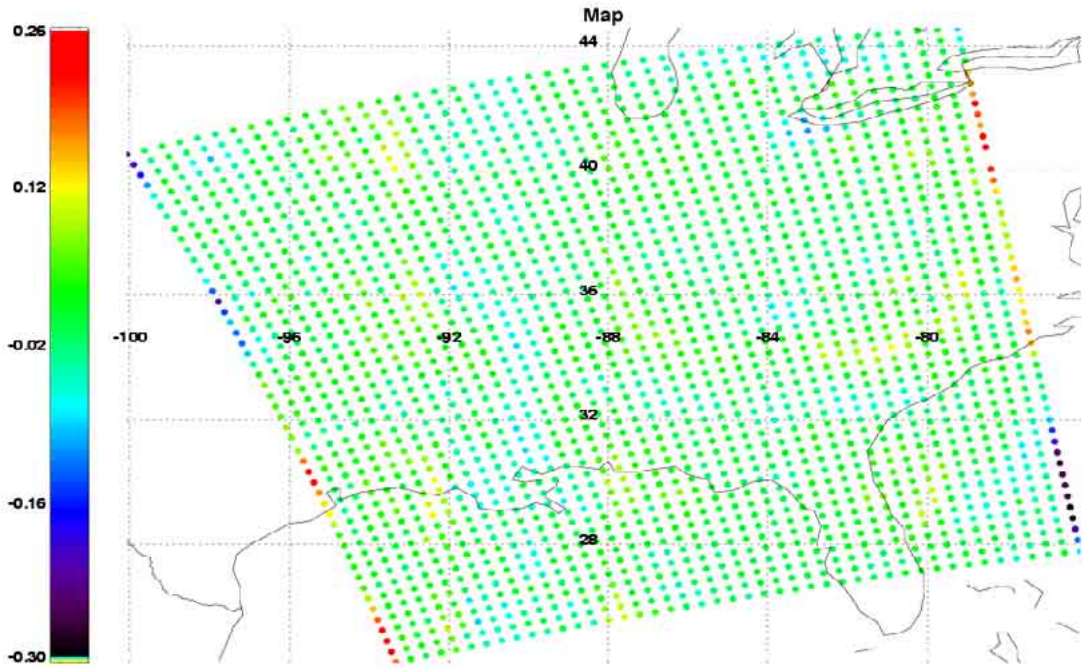

Figure 2-8: Difference between the regridded and true fields shown in Figure 2-6 and Figure 2-7 respectively.

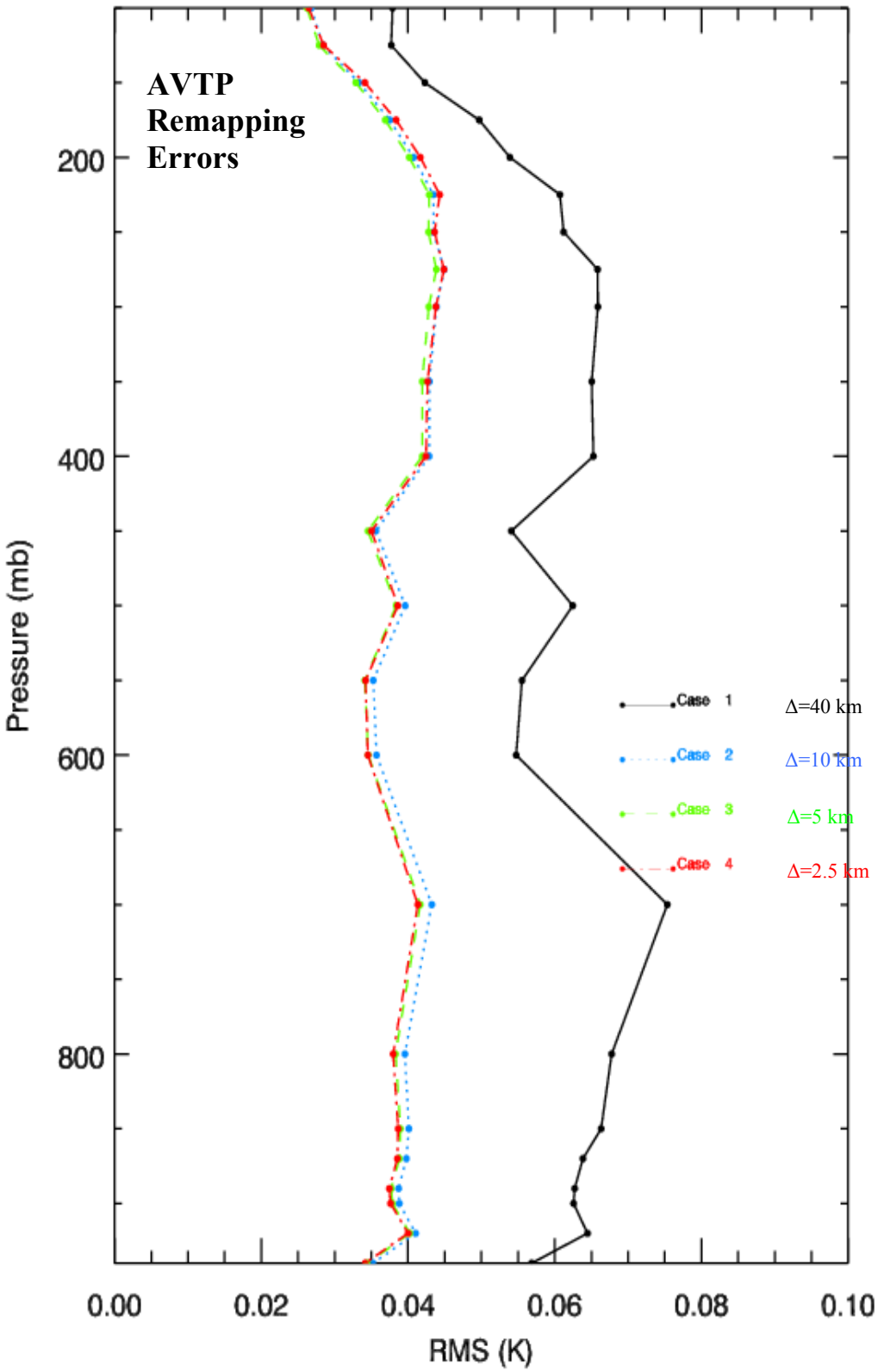

Figure 2-9: Layer-averaged RMS error in the vertically remapped AVTP field.

The four cases are:
Case 1: $\Delta = 40$ km, R_m/Δ parameters reduced by a factor of 16 from their default values
Cases 2,3 and 4: Default R_m/Δ parameters with $\Delta = 10$, 5, and 2.5 km.
In Cases 2 through 4, the 40-km output has been generated by subsampling the finer resolution analysis grid.

Our finding that the VR errors were very small in relation to the dynamic range of the input temperature data is consistent with our expectations, based on theoretical grounds. The input temperature data have a spatial resolution of approximately 40-km and a spacing of 20-km or less, in compliance with the Nyquist criterion for negligible interpolation error. The CFOV spacing was, in fact, chosen on the basis of this criterion and for this reason. As discussed in the discussion of footprint matching in *CMIS ATBD Vol. 1 Part 2*, the 40-km CFOVs are generated at each 12.5-km scan arc and are spaced 12-km apart along each scan line.

Based on the results shown in Figure 2-9, the error budget due to vertical remapping of the AVTP field is on the order of 0.04 K between 1000 and 100 mb. The RAMS scenes have a lid at 50 mb and are thus inadequate to test the performance of the remapping program in the upper atmosphere, but the weak altitude dependence of errors evident in Figure 2-9, and the generally smoother nature of the stratospheric temperature fields, lead us to believe that the remapping errors are unlikely to be significantly larger there. For altitudes above the 20-mb pressure level, the VR algorithm operates on data with a 200-km resolution and a 100-km spacing, so the Nyquist criterion is met and VR errors should be similarly small.

2.3.3.2 Atmospheric Vertical Moisture Profile

The AVMP EDR is produced at 15-km resolution from the 15-km slant-path data, as shown in Figure 2-10 for the 200-mb level of the test scene.

The true and the vertically-remapped fields are shown in Figure 2-11 and Figure 2-12 respectively, while the percentage differences between the two fields are plotted in Figure 2-13. The differences, while generally small (less than 1%), do occasionally become large, especially in the regions of sharp gradients surrounding areas of low moisture content. In these regions, the water vapor mixing ratio increases by almost two orders of magnitude over relatively short distances and this leads to remapping errors as large as 30%. It is questionable whether such dramatic gradients actually occur in nature and, at the least, this test scene can be considered unusually stressing.

Overall, however, the remapping errors are small, as illustrated by the RMS errors plotted in Figure 2-14. For AVMP performance statistics, vertical averaging over 2-km layers is applied, and such averaging can partially mitigate errors that may be confined to narrow layers. As in the case of AVTP, a significant improvement in performance can be achieved by subsampling a fine-resolution analysis grid employed by the recursive filter, but the advantages of subsampling must be weighted against the computational cost, especially when compared to the results from a Δ = 15 km run employing a scaled-down radius of influence at each analysis pass (Case 1 in Figure 2-14).

Based on the results shown in Figure 2-14, the error budget for the vertical remapping of the AVMP field is 0.4%, 0.9%, and 1.2% for the layers surface-600 mb, 600-300 mb, and 300-100 mb, respectively. While the numbers in the mid- and upper troposphere are somewhat larger than the 0.6% nominal budget error, it should be noted that hurricane Bonnie represents an extreme meteorological situation, even in comparison to other stressing situations (see Table B-4) and the global RMS errors are likely to be significantly smaller. Moreover, even these somewhat elevated errors have little effect on the overall error budget for the AVMP EDR, considering that they are uncorrelated with the other error sources and thus combine as a sum of squares.

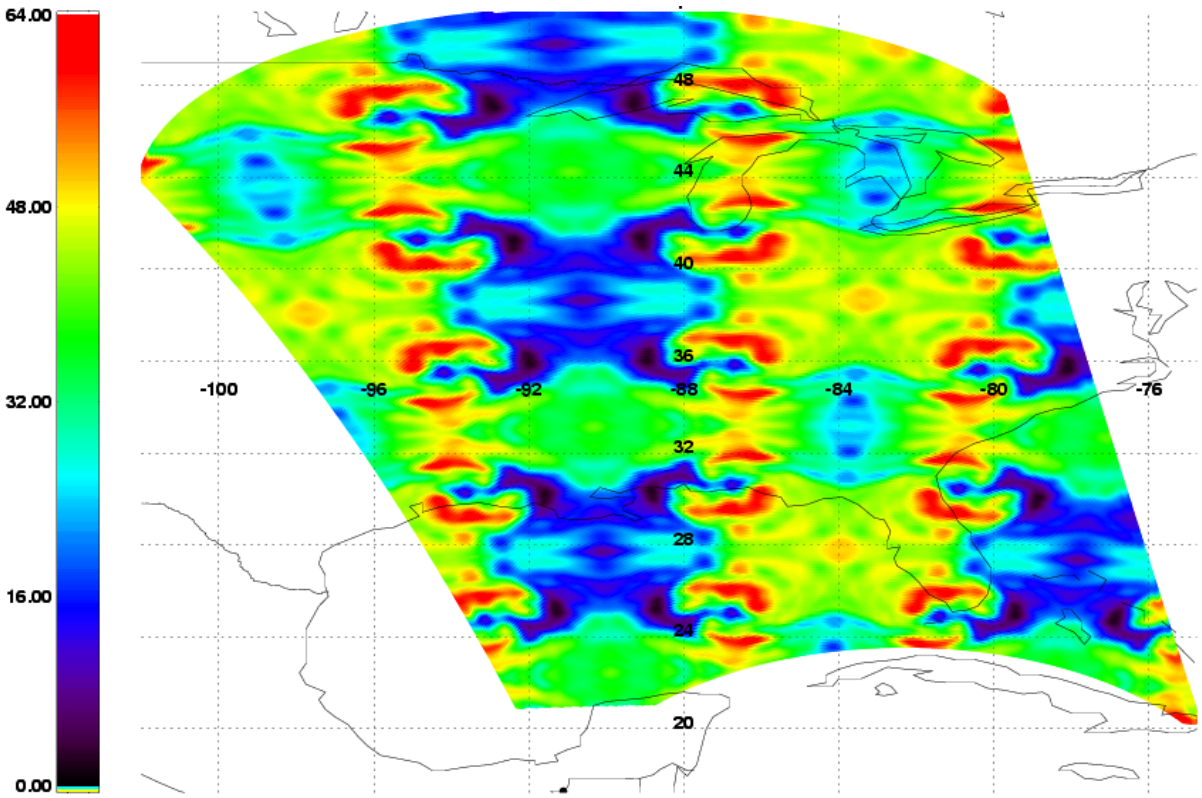

Figure 2-10: 15-km FOV resolution slant-path moisture field at 200 mb serving as input to the vertical remapping of the AVMP field. The units are g/g x 10^6, i.e. parts per million per mass (ppmm).

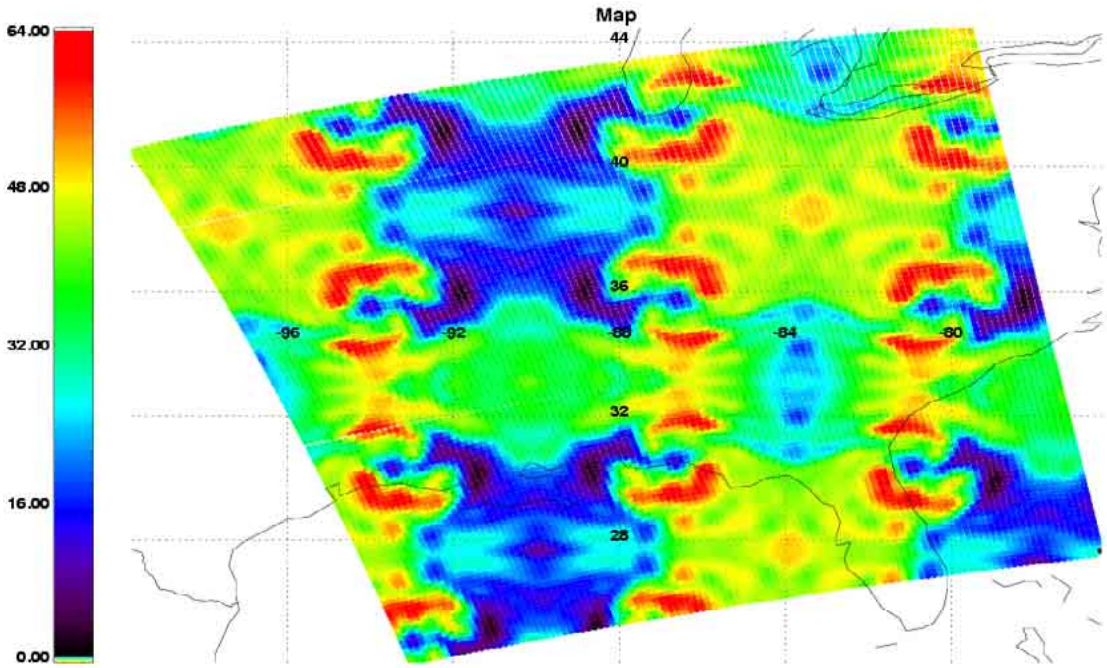

Figure 2-11: Vertically-remapped AVMP field at 200 mb generated from the input data plotted in Figure 2-10.

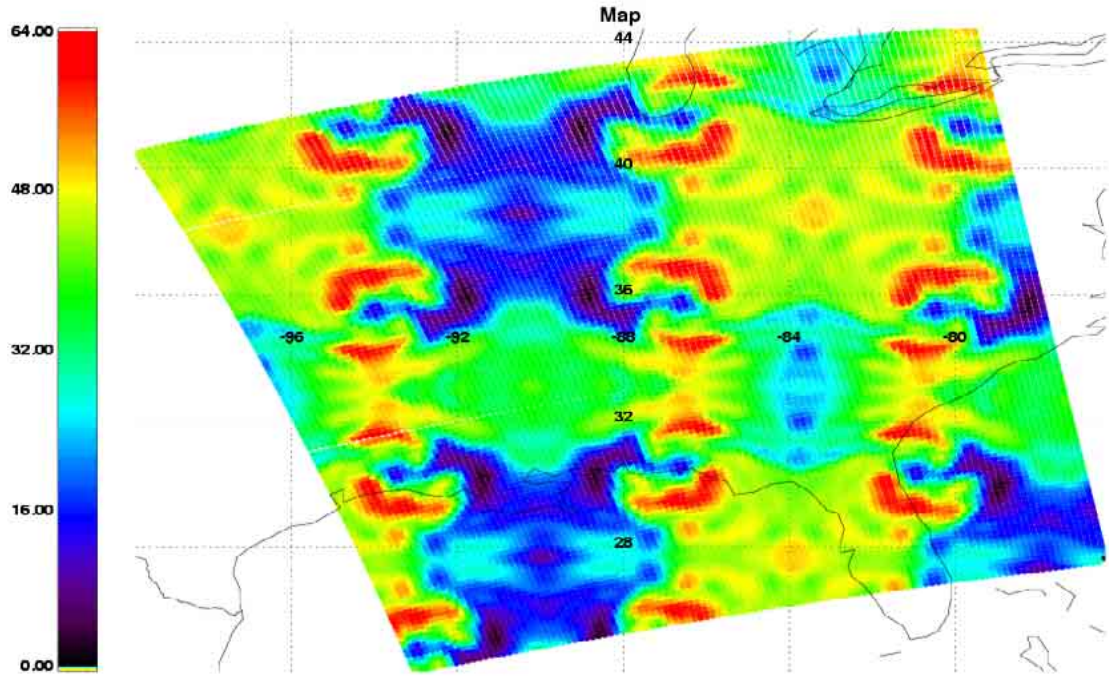

Figure 2-12: True vertical-path moisture field at 200 mb at the grid locations in Figure 2-11 generated by a direct averaging of the high-resolution RAMS data.

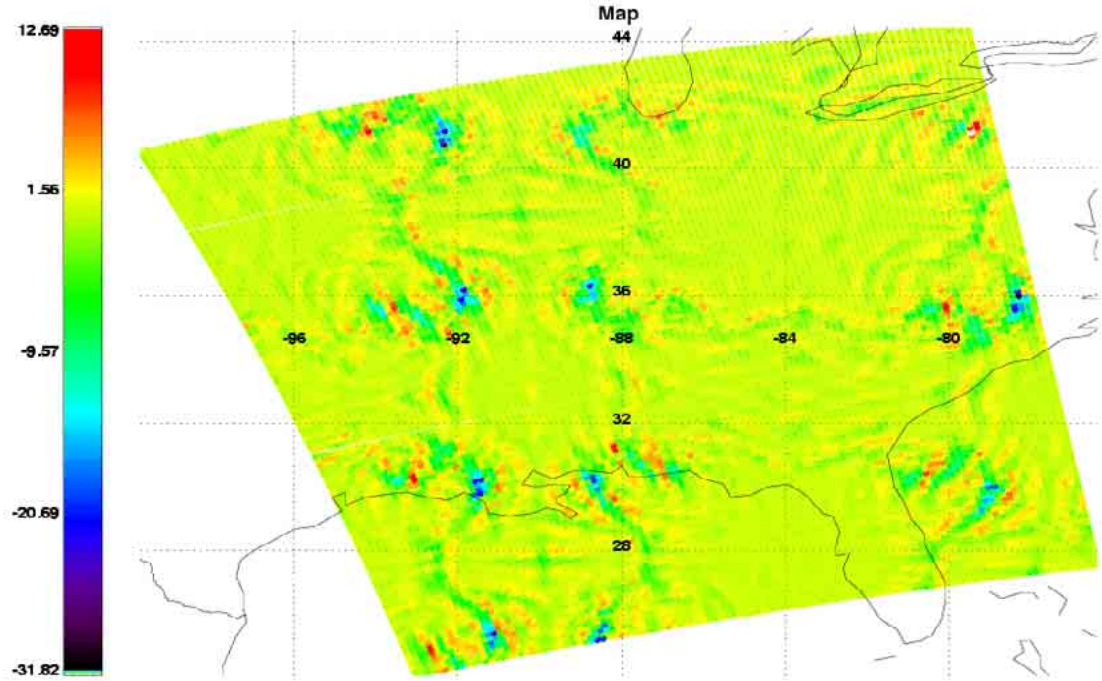

Figure 2-13: Percentage difference between the regridded and true fields shown in Figure 2-11 and Figure 2-12 respectively.

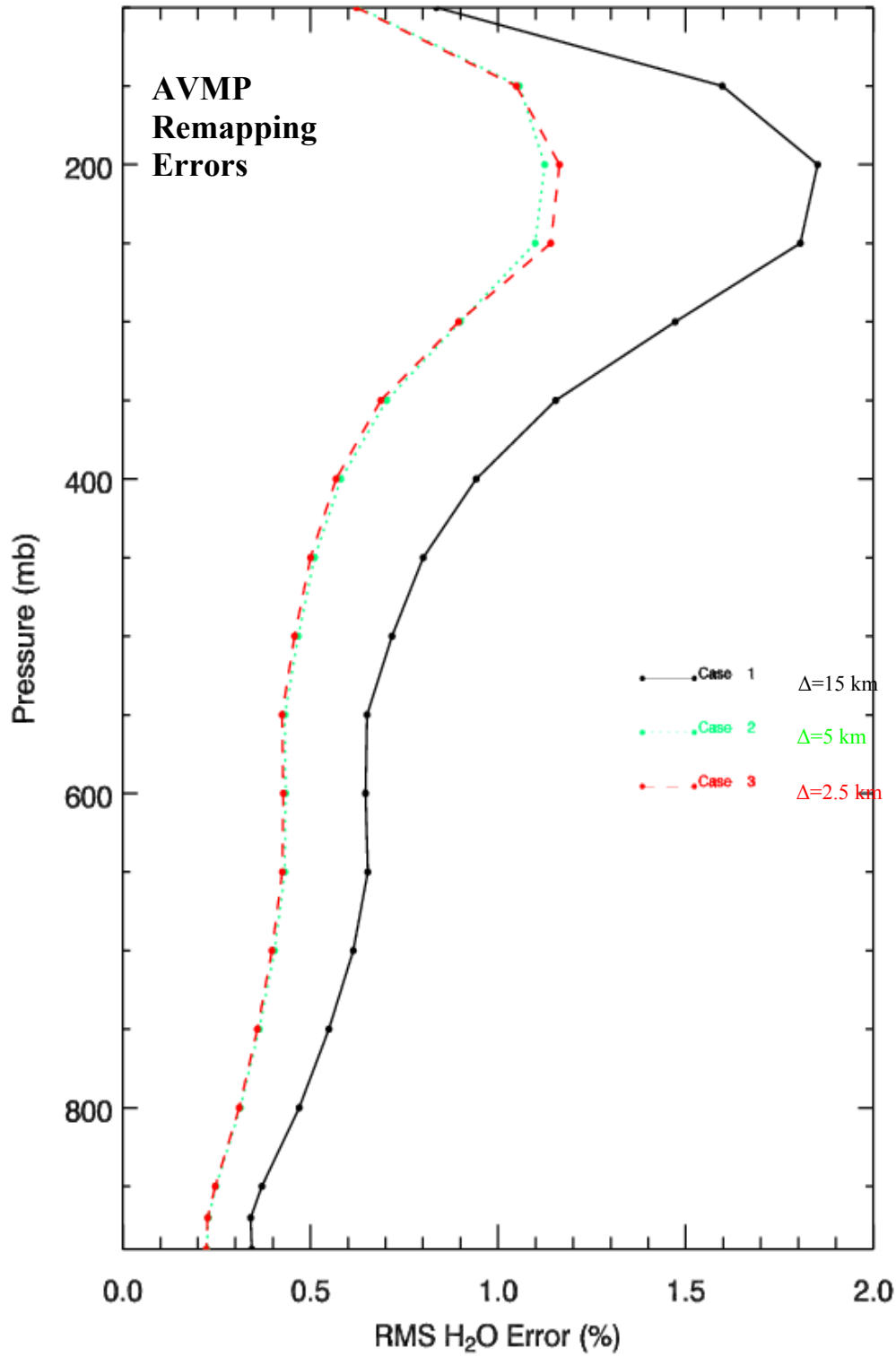

Figure 2-14: RMS errors in the vertically-remapped AVMP field.

The three cases shown are:
Case 1: Δ = 15 km, default R_m/Δ values divided by 6,
Case 2: Δ = 5 km, default R_m/Δ values,
Case 3: Δ = 2.5 km, default R_m/Δ values.
In Cases 2 and 3, the output AVMP field has been generated at the Δ = 15 km resolution by subsampling the small Δ grids (every 3rd and 6th point in both directions, respectively).

2.3.4. FOV Interpolation

2.3.4.1 Data-dense Areas

The FOV interpolation in data-dense areas has been evaluated statistically in three cascading scenarios: coarse-to-coarse (50- to 40-km FOV), coarse-to-fine (50- to 15-km FOV), and fine-to-fine (20- to 15-km FOV). These scenarios span the range of FOV interpolation tasks in the baseline CMIS cascade algorithm design as well as alternative paths through the cascade. The results of the third scenario (20- to 15-km FOV) are presented strictly for evaluating algorithm behavior. The CFOV data at 20 and 15 km are collocated in the baseline implementation of the footprint matching algorithm, so no interpolation is needed for the cascade step from 20 to 15-km CFOVs.

The RMS errors for temperature in the three scenarios are shown in Figure 2-15 through Figure 2-17, respectively, while the corresponding plots for water vapor are shown in Figure 2-18 through Figure 2-20. It should be noted that the errors in these figures are not layer-averaged (i.e., are the values at discrete pressure levels). These test data were derived directly from RAMS model products by spatially averaging over FOVs, and the data were not processed through the radiometric model and the Core Module retrieval algorithm. The Core Module induces some vertical smoothing according to the radiometric averaging that is inherent in the microwave measurements. Atmospheric horizontal gradients vary with altitude, so the vertical smoothing causes some horizontal smoothing. That smoothing was not represented in these test data so, in that respect, the test data were excessively stressing and the interpolation errors we computed are excessively large.

In each scenario, four cases have been considered, one employing an analysis grid with spacing equal to the target FOV resolution and three cases employing progressively smaller analysis grid resolution of 10, 5, and 2.5 km. For temperature and moisture, the best performance is achieved with the finest grid resolution (2.5 km). The performance is only slightly degraded with a 5-km analysis grid, except for upper tropospheric moisture, where the degradation on going from 2.5- to 5-km grid can reach 0.1%. Overall, a performance similar to the 2.5-km case can be achieved by employing the target FOV resolution for the analysis grid, but scaling the default R_m/Δ values by a factor of 6 (which yields radii of influence similar to the 2.5-km case in absolute numbers). The most notable exception to this statement is a degradation in moisture errors for the 50- to 40-km cascade, in which the bi-linear interpolation that follows the RF carried the burden of interpolating over distances up to 20 km. However, even in this case the performance is significantly improved when the default R_m/Δ values are scaled down (see Figure 2-18). The occasional spikes in the performance matrix, mainly in runs employing a rather coarse $\Delta = 10$ km and the default R_m/Δ values (e.g., in Figure 2-17 and Figure 2-20), are associated with localized regions of sharp gradients in the input fields, which apparently are not handled accurately when the radii of influence used by the RF are relatively large.

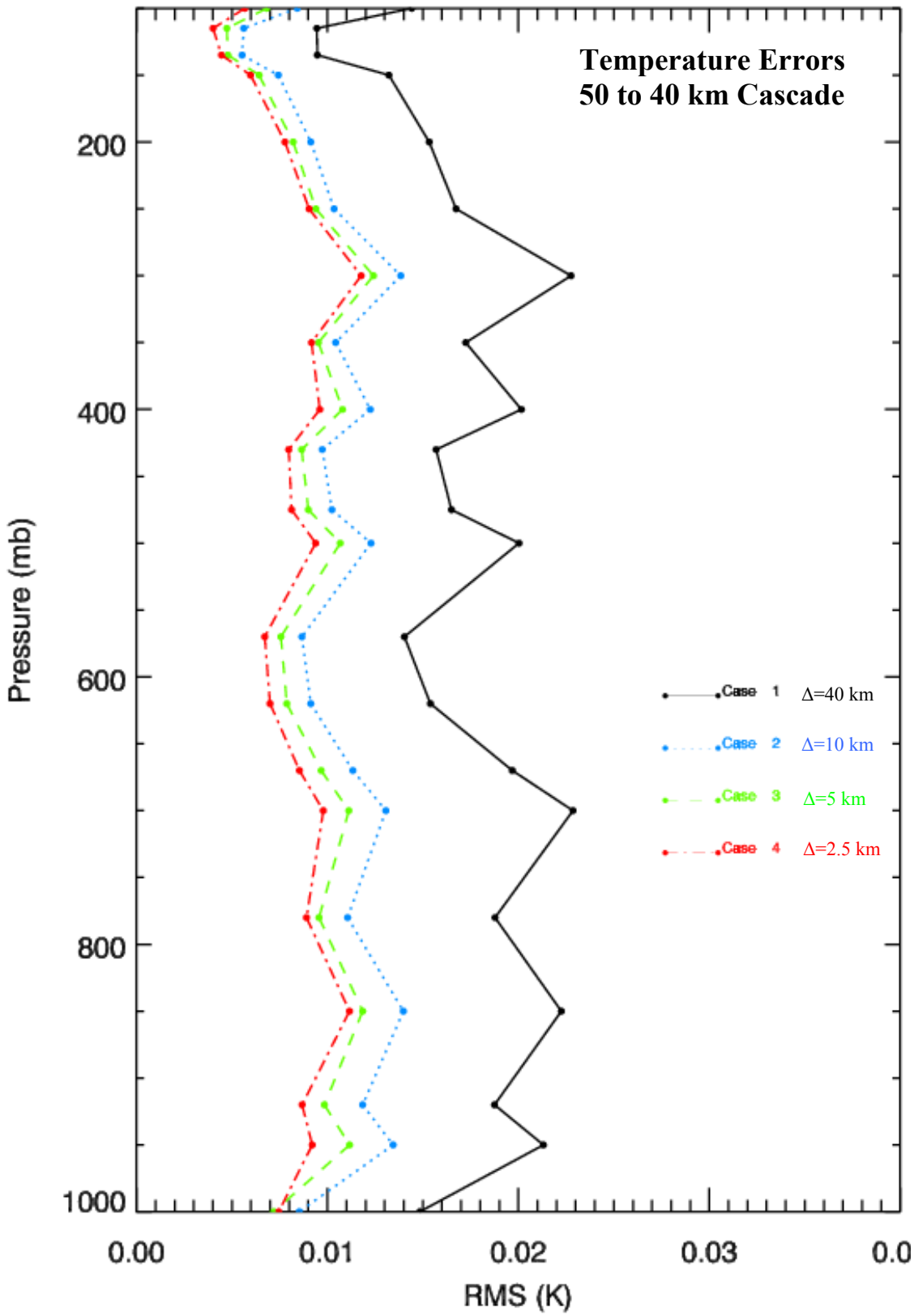

Figure 2-15: RMS errors in temperature for the 50- to 40-km FOV interpolation.

The four cases shown are:
Case 1: Δ = 40 km, R_m/Δ scaled down by 16 from the default values,
Cases 2-4: Δ = 10, 5, and 2.5 km, respectively, default R_m/Δ.

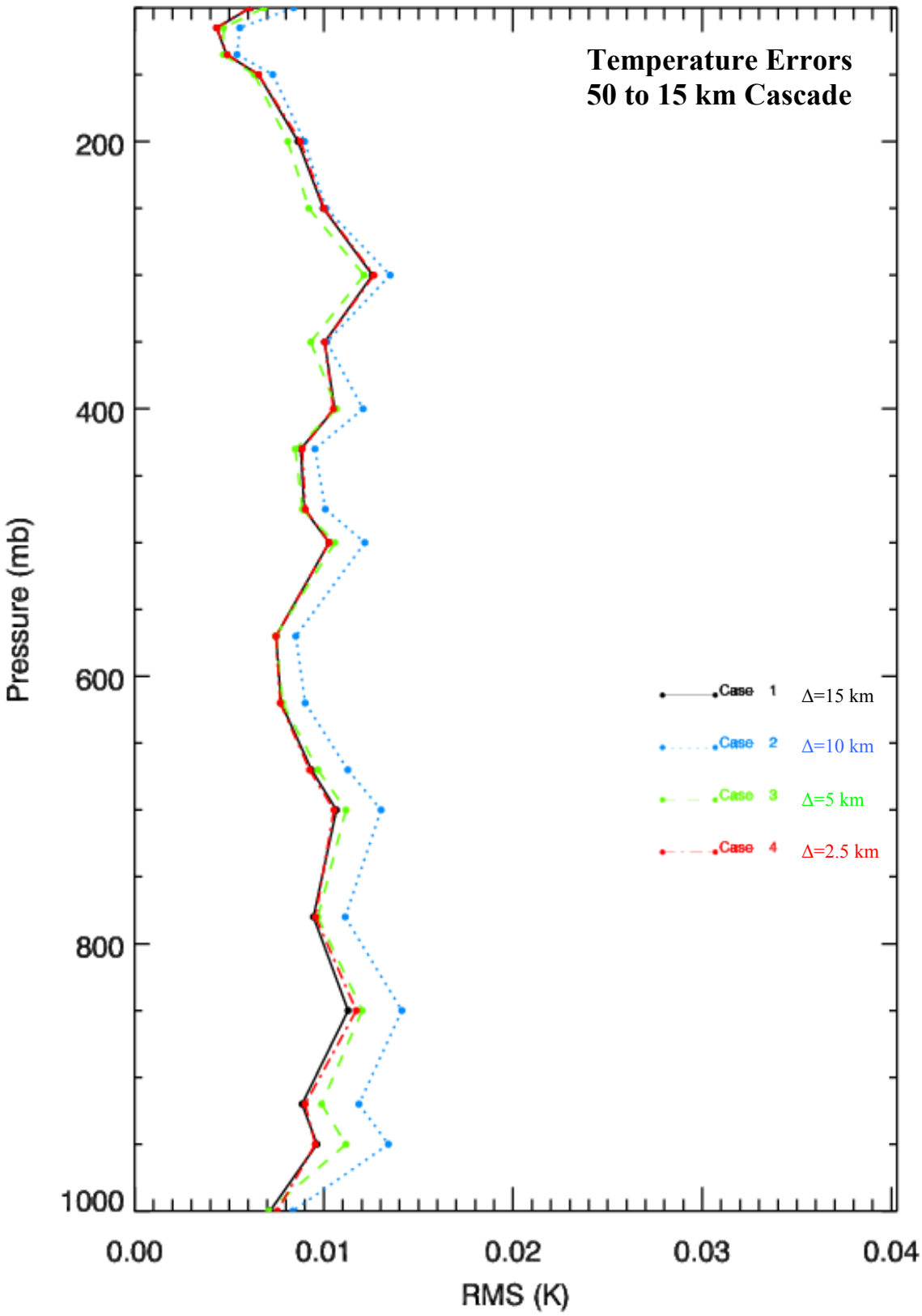

Figure 2-16: RMS errors in temperature for the 50- to 15-km FOV interpolation.

The four cases shown are:
Case 1: Δ = 15 km, R_m/Δscaled down by 6 from the default values,
Cases 2-4: Δ = 10, 5, and 2.5 km, respectively, default R_m/Δ.

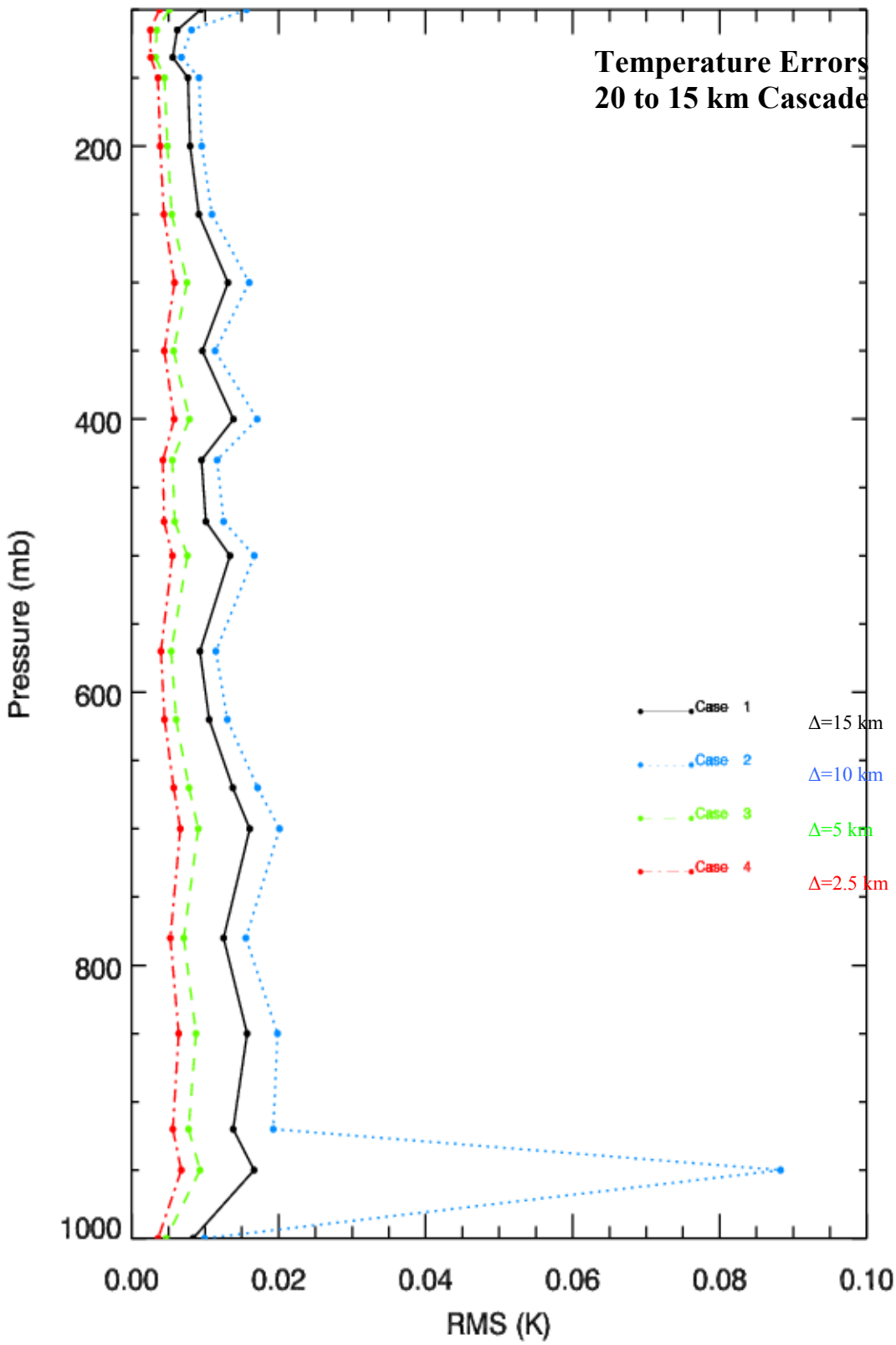

Figure 2-17: RMS errors in temperature for the 20- to 15-km FOV interpolation for the same cases as in Figure 2-16.

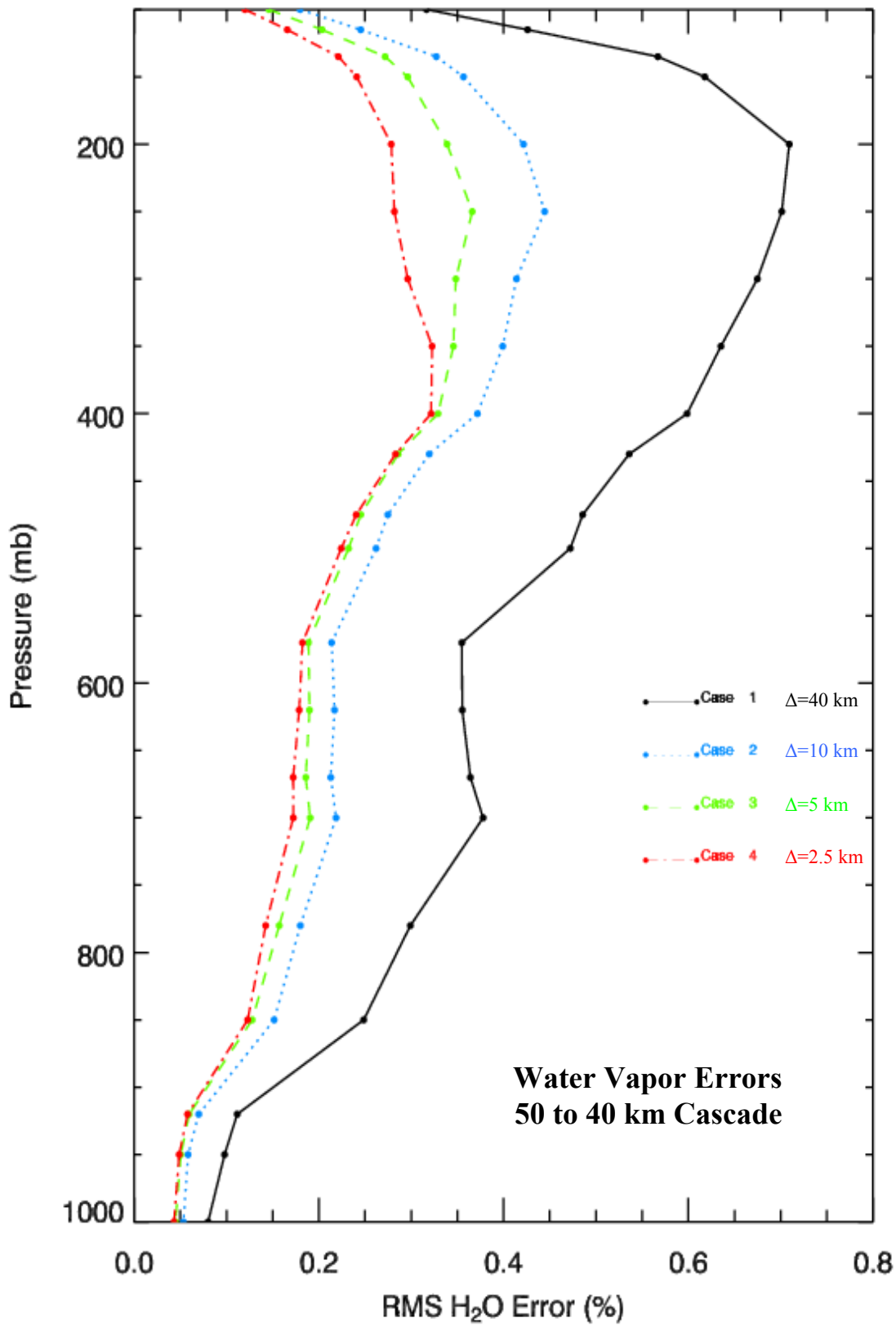

Figure 2-18: RMS errors in water vapor for the 50- to 40-km interpolation in the four cases defined in the caption of Figure 2-15.

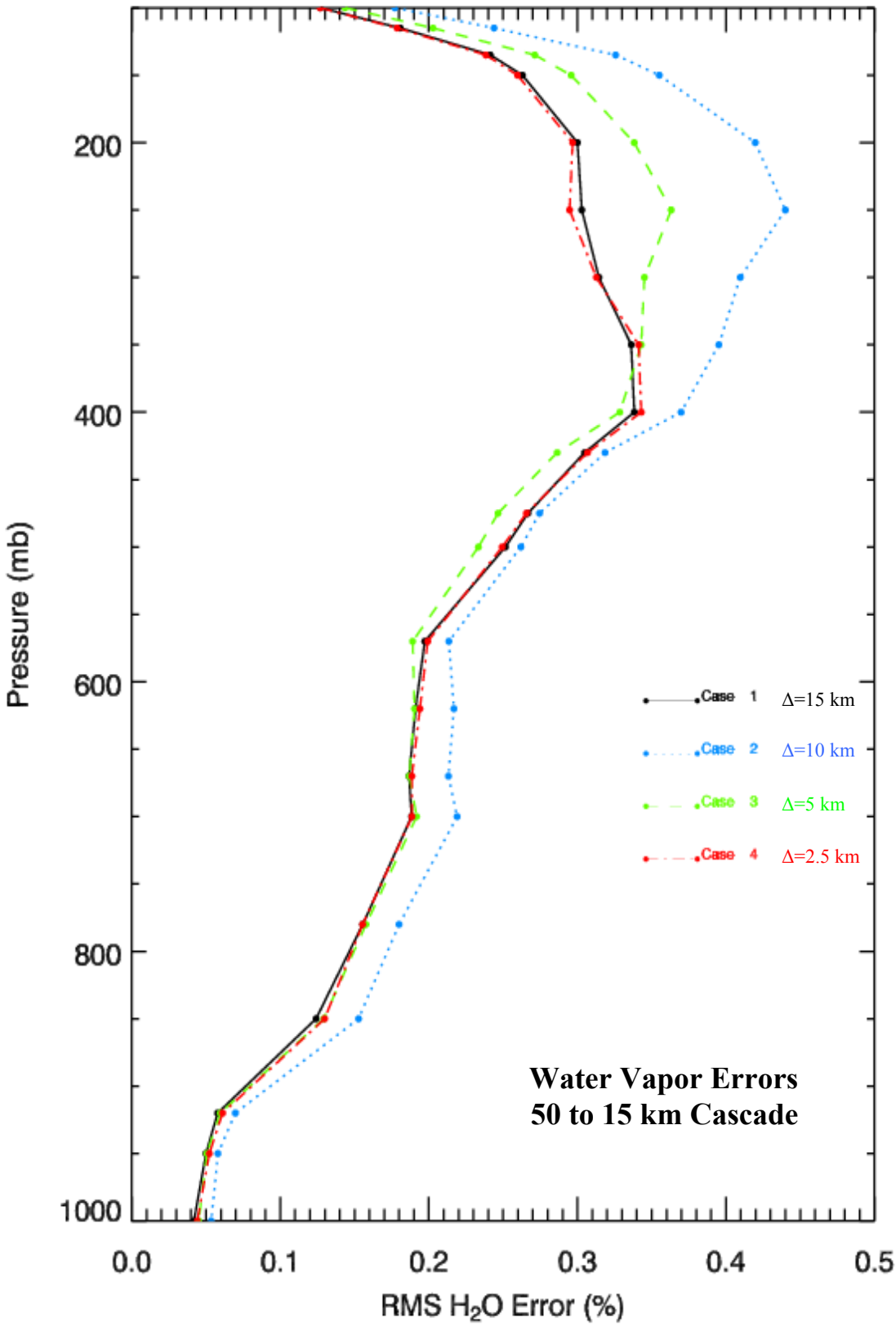

Figure 2-19: RMS errors in water vapor for the 50- to 15-km interpolation in the four cases defined in the caption of Figure 2-16.

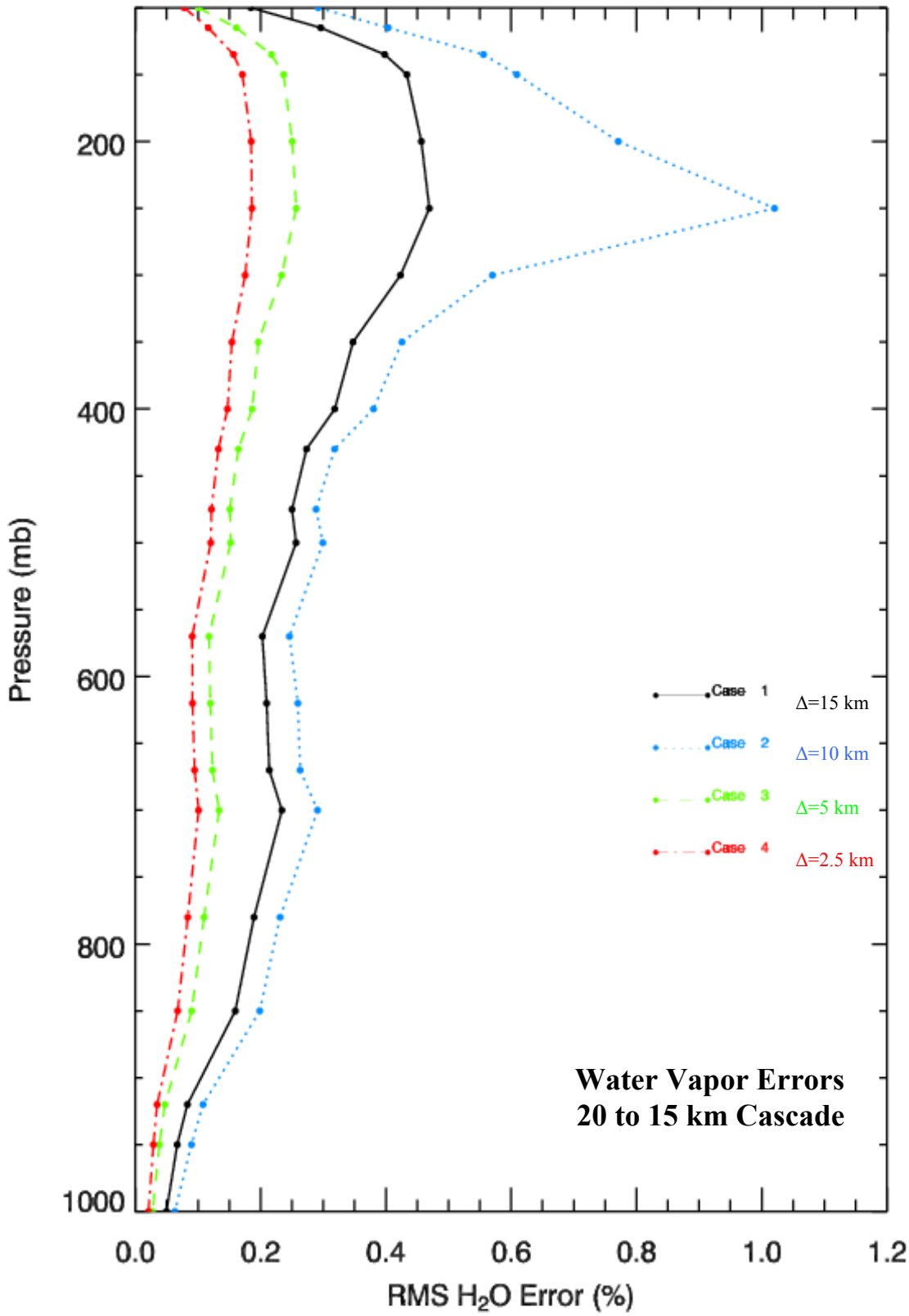

Figure 2-20: RMS errors in water vapor for the 20- to 15-km interpolation in the four cases shown in Figure 2-17.

2.3.4.2 Behavior in Data Gaps

When input data are missing, e.g., in areas of precipitation, the FOV interpolation program cannot, of course, reproduce the "true" fields. The main purpose of the program in this case is to provide a smooth field that interpolates between areas where input data are present. These interpolations provide the background atmospheric characteristics for the precipitation rate EDR algorithm, which is not highly sensitive to errors in the temperature and water vapor profiles. Performance of the FOV interpolation algorithm in such gaps can only be evaluated qualitatively by visual inspection of the interpolated field.

As an example, we show in this section results from the 20- to 15-km interpolation of the moisture and temperature fields at 250 mb.

Figure 2-21 shows the "true" field obtained by direct averaging of the RAMS data performed at 15-km resolution and employing a 20-km spatial average, while Figure 2-22 shows the input 20-km FOV field with artificially created data gaps. The gaps were created by flagging the data as missing on the basis of RAMS cloud liquid water. The flagging threshold was tuned so that the gaps would be about 300-400 km wide (Figure 2-3). The output from a 20- to 15-km FOV interpolation run employing the field shown in Figure 2-22 as input is shown in Figure 2-23. This run corresponds to "Case 3" in Figure 2-20, i.e., $\Delta = 5$ km for the analysis grid and seven passes of the recursive filter.

The output QC field is shown in Figure 2-24. Note that, consistent with the formalism presented in HayP95 (their Appendix B), the numerical values of the QC field (including the limiting value for data gaps) are dependent on R_m and Δ and are thus best used in a relative sense. The field shown in Figure 2-23 captures the main features of the input field, including the precipitation gaps, an extra gap in the SW corner of the swath, and increasing data density towards the edge of swath (with an expected decrease in data density right at the edge of swath).

Unlike in the data-dense areas, the interpolation program appears to behave somewhat better when two extra passes of the recursive filter employing large values of R_m/Δ are applied. This is related to the fact that, in this case, the program is able to reach across data gaps (whose size is comparable to R_m) to provide a more smoothed field within the gaps. In tests without the two extra passes, the algorithm tended to pack the entire cross-gap gradient into the center of the gap.

An interesting behavior occurs in the FOV interpolation of the 250-mb temperature field. In this case, the input data, shown in Figure 2-26, retains isolated patches of warm temperatures corresponding to the relatively precipitation-free warm core of the hurricane. This relationship between gap occurrence and the temperature field, and the relatively smooth nature of the field, allows the interpolation program to reproduce much of the large-scale variability inherent in the input scene. As a result, the interpolated field, shown in Figure 2-27, closely resembles the true field shown in Figure 2-25.

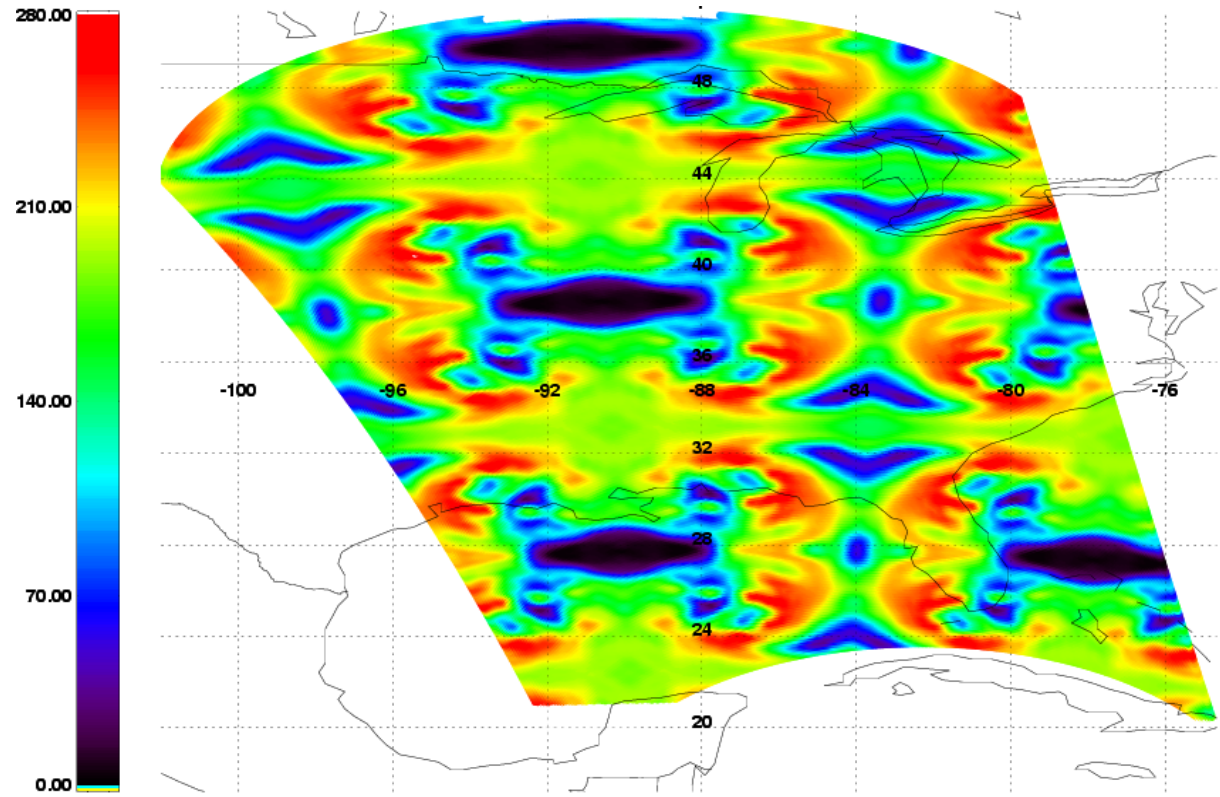

Figure 2-21: True slant-path 20-km spatial average FOV moisture field at the locations of the 15-km FOVs, at 250 mb (in parts per million per mass).

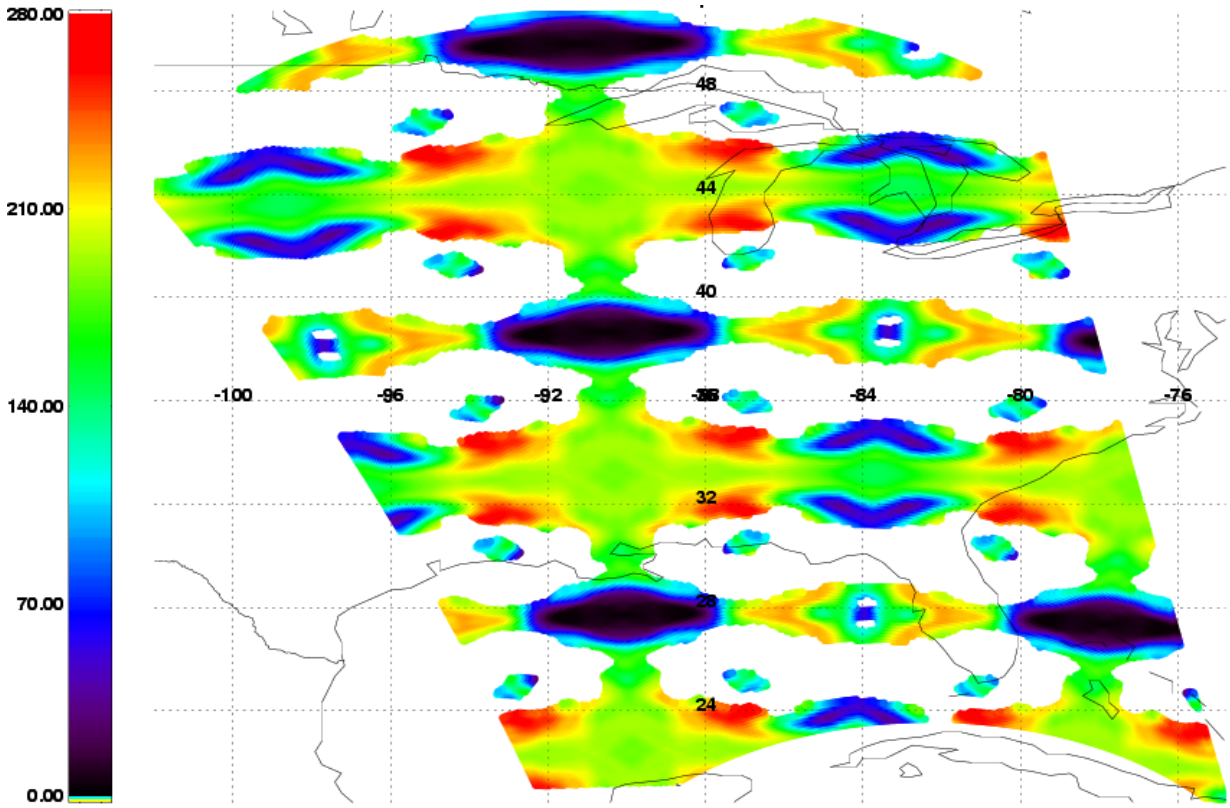

Figure 2-22: Slant-path, 20-km FOV moisture field at 250 mb with data gaps serving as input to the FOV interpolation program.

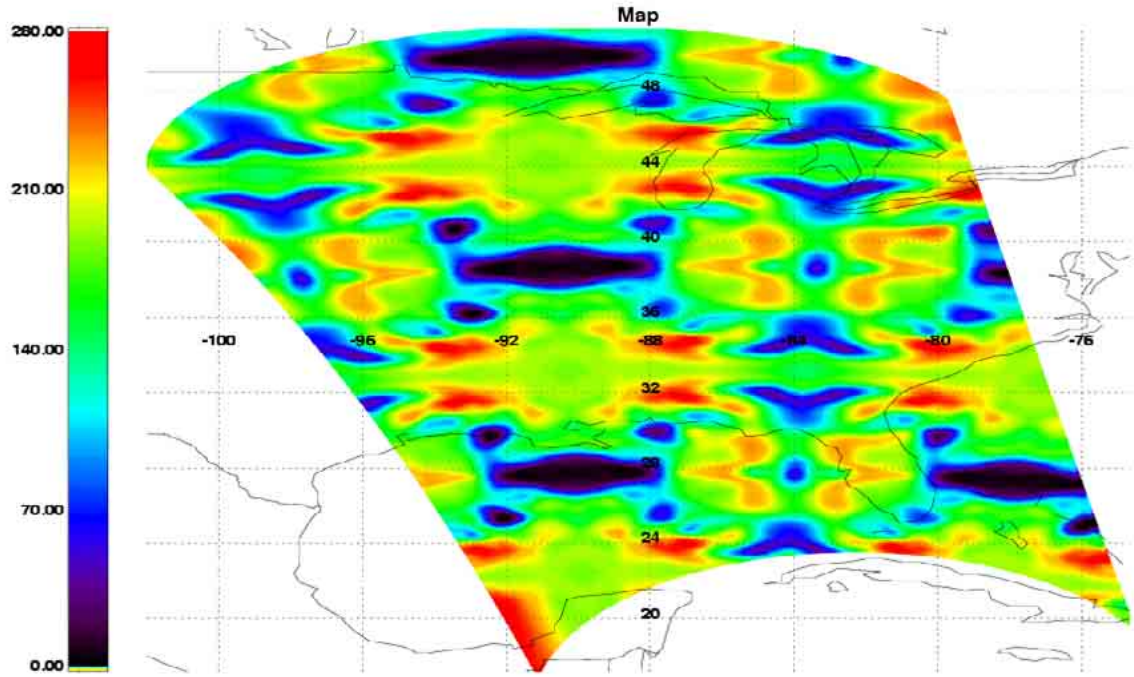

Figure 2-23: 15-km FOV moisture field at 250 mb interpolated from the 20-km FOV field shown in Figure 2-22.

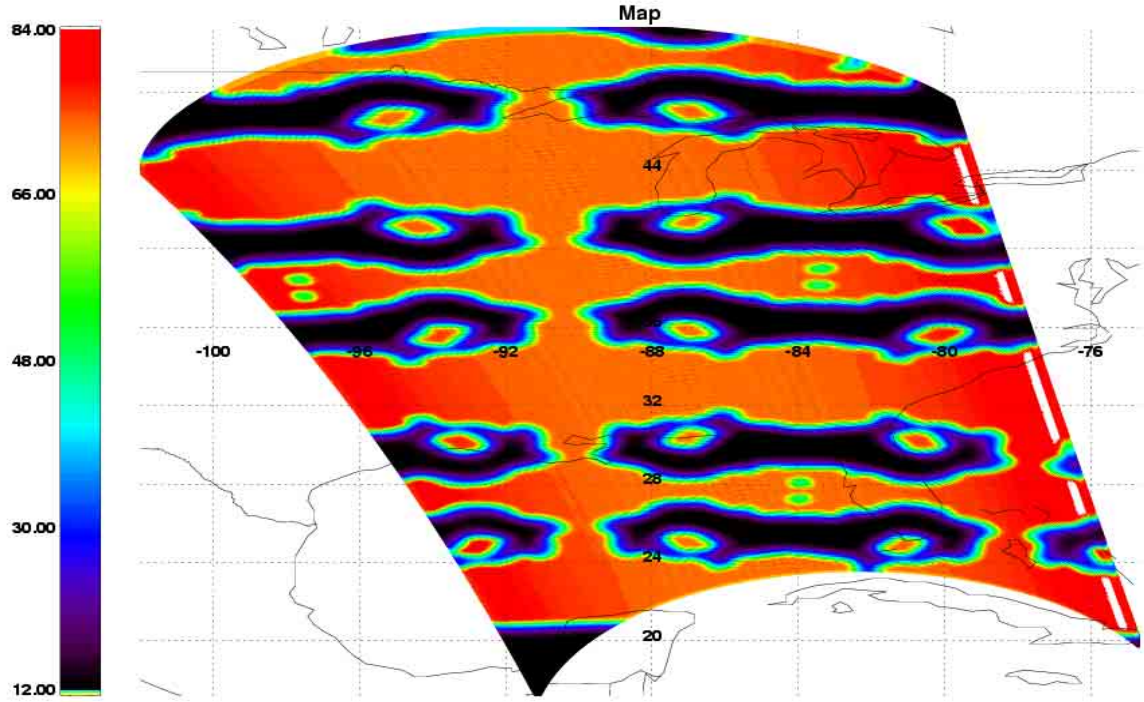

Figure 2-24: Output QC flag for the run used to generate Figure 2-23.

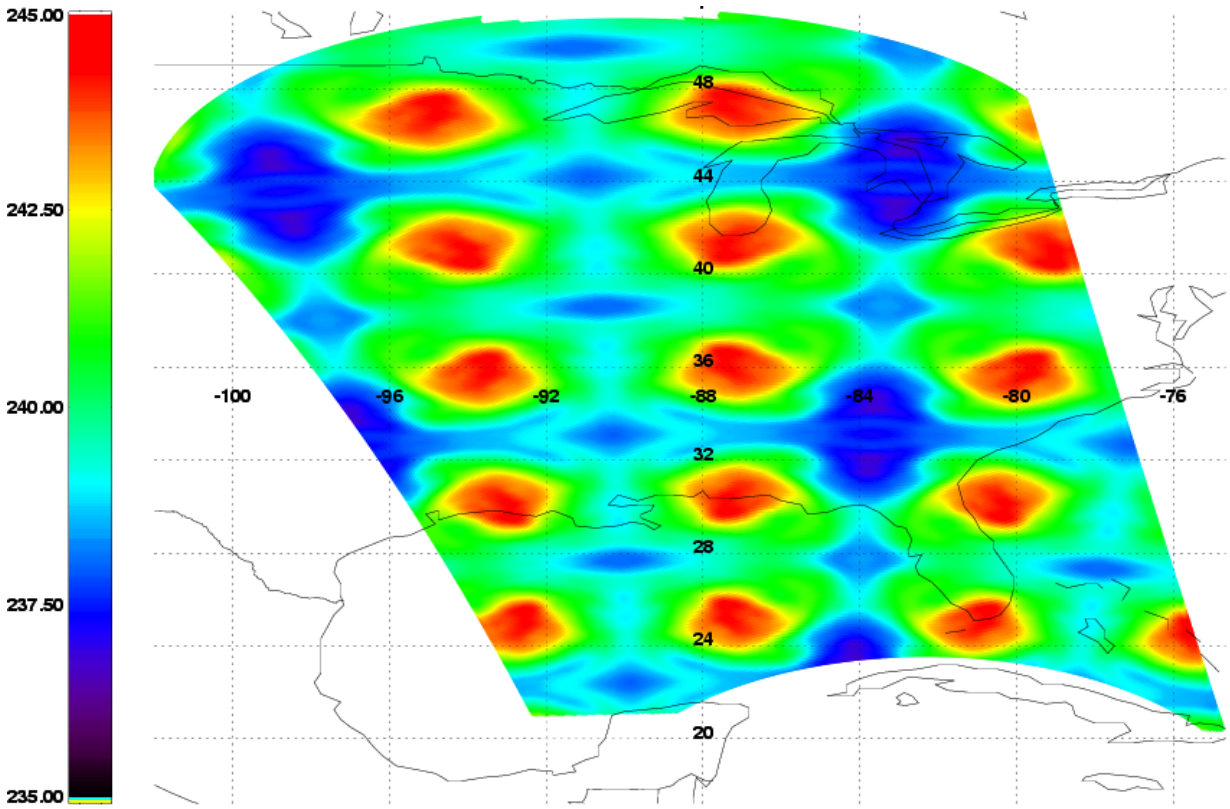

Figure 2-25: True slant-path, 20-km spatial average FOV temperature field at the locations of the 15-km FOVs at 250 mb.

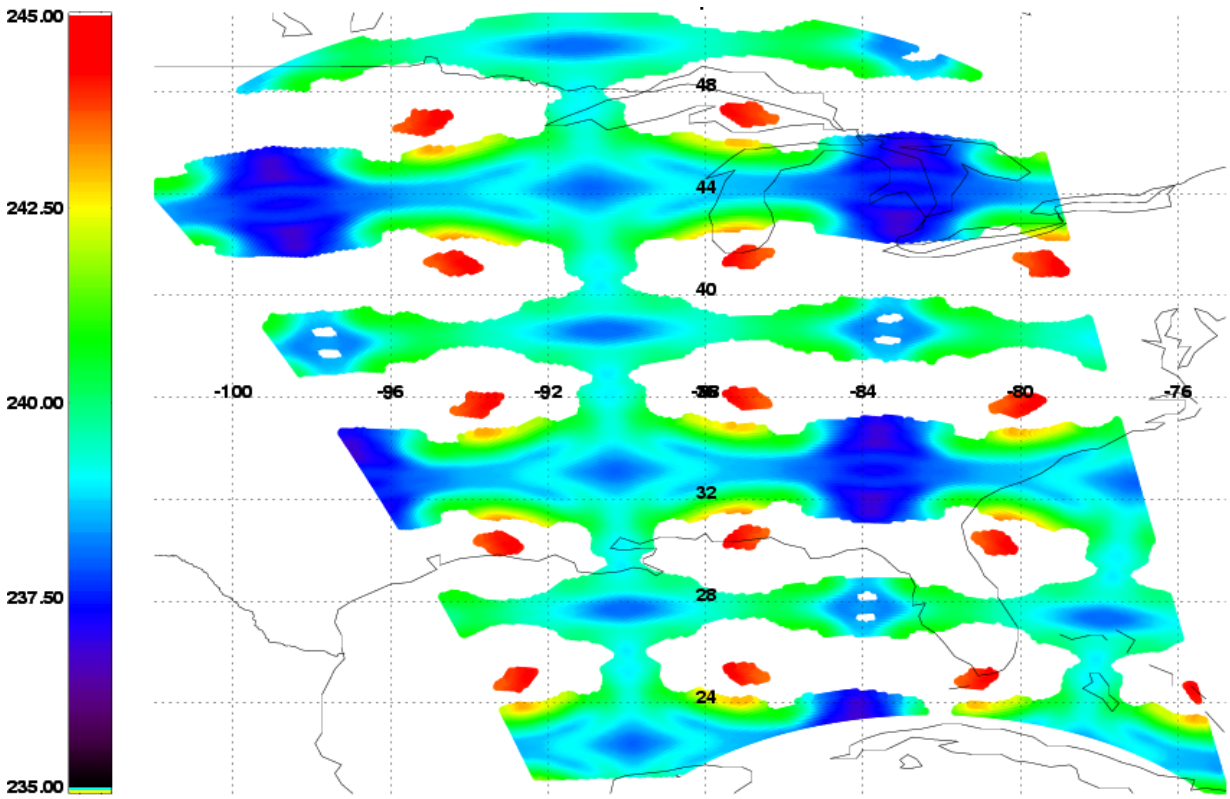

Figure 2-26: Slant-path, 20-km FOV temperature field at 250 mb with data gaps serving as input to the FOV interpolation program.

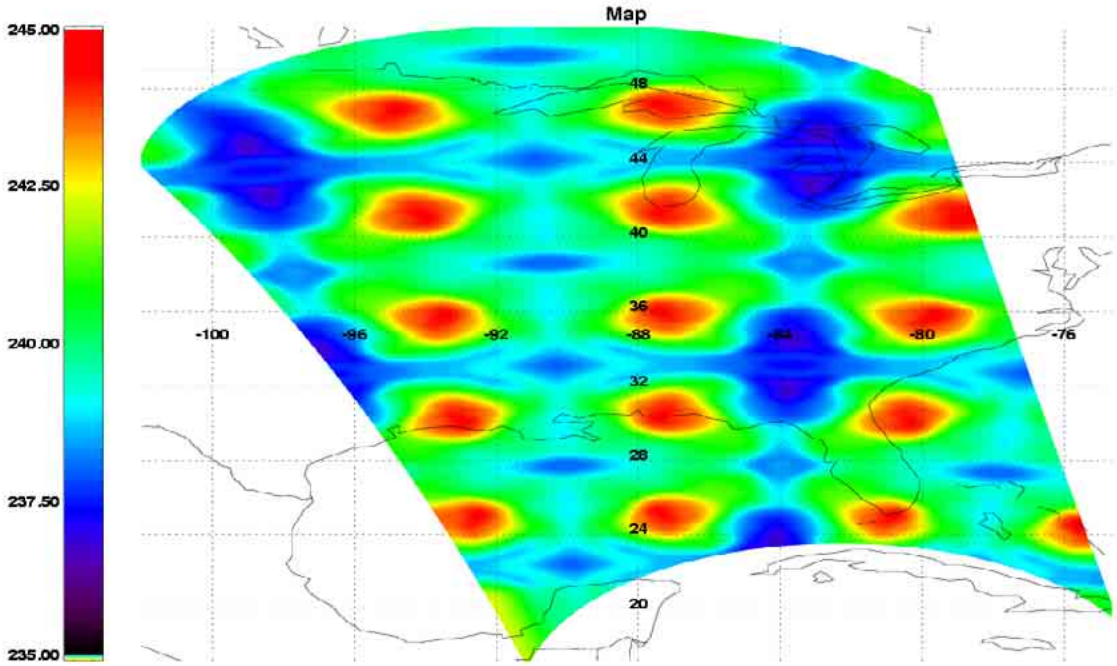

Figure 2-27: 15-km FOV temperature field at 250 mb interpolated from the 20-km FOV field shown in Figure 2-26.

2.4. Earth Gridding Algorithm Performance

We have tested and tuned each of the earth gridding algorithm options using realistic orbital CFOV scenes generated by the CMIS testbed with underlying spatial structure meant to represent maximum variability. The scenes combine surface emissivities build based on land surface type maps and high-resolution atmospheric model forecasts from the University of Pennsylvania Mesoscale Model 5 (MM5). The MM5 surface and atmospheric data fields, which originally covered a 66x66 domain at 3 km nominal resolution, were tilled by "matchbook" flipping to fill a 1122x1122 domain centered on the eastern half of the USA. IGBP (International Geosphere-Biosphere Program) surface types for this domain were drawn from the USGS Global Land Cover Characterization and surface emissivities were added by random selection from a dataset of global emissivity-surface type match-ups. As described in section 2.3.1, emissivity and land surface temperature (LST) fields were averaged over simulated CMIS CFOVs at nominal resolutions of 20 and 50 km to build simulated CMIS swaths. The CFOV-averaged data represent simulations of near-perfect CMIS retrievals without retrieval algorithm errors or spatial inconsistency due to differing composite footprint shapes at each input channel. For verification data, composite footprint spatial averages of emissivities and LST were also generated from the high-resolution domain at the geolocation of each earth-grid point. As addressed below, one minor limitation in these verification data are that they are oriented at a fixed earth-azimuth angle that does not necessary match that of the sensor-simulated CFOVs that surround them.

After initial earth gridding algorithm testing and tuning, a second test scene centered on south Asia was generated by shifting the original 1122x1122 domain 160°. This provided an independent emissivity scene realization with a lower fraction of water and coastlines and dominated by different surface types. LST data were not regenerate so only the emissivity regridding case was tested for this scene.

Figure 2-28 through Figure 2-30 show conditions at the 3 km resolution of the test scenes. The North American scene is primarily a mixture of vegetation types with inland water bodies a long coastline; the Asian scene has less water and more sparsely vegetated or bare areas. The differences are apparent in the 36H emissivity maps. The dominant feature of the North American scene is the land-water contrast with relatively high emissivities across vegetation-dominated land. The greater range of vegetation amounts in the Asian scene result in higher emissivity heterogeneity on land. We chose to use H-pol. emissivities in order to test the algorithms with the most heterogeneous scenes. 6H and 18H emissivities were also tested but performance results were not significantly different from 36H. Figure 2-30 shows the MM5 surface temperature field after being tiled to fill the 1122x1122 domain of the North American scene. The temperature varies by ~10 K across each tile. Temperature test results are only given for the North American scene since both scenes share the same temperature map.

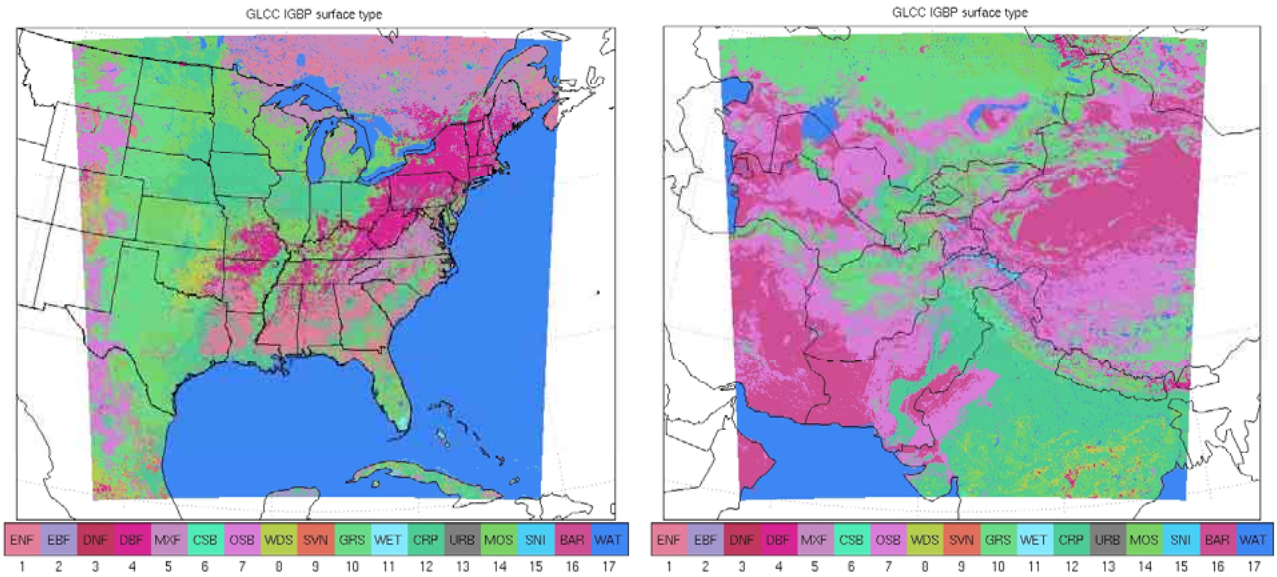

Figure 2-28: IGBP surface types for the North America and Asia test scenes. See *CMIS ATBD Vol. 11, Vegetation/Surface Type EDR* for type descriptions.

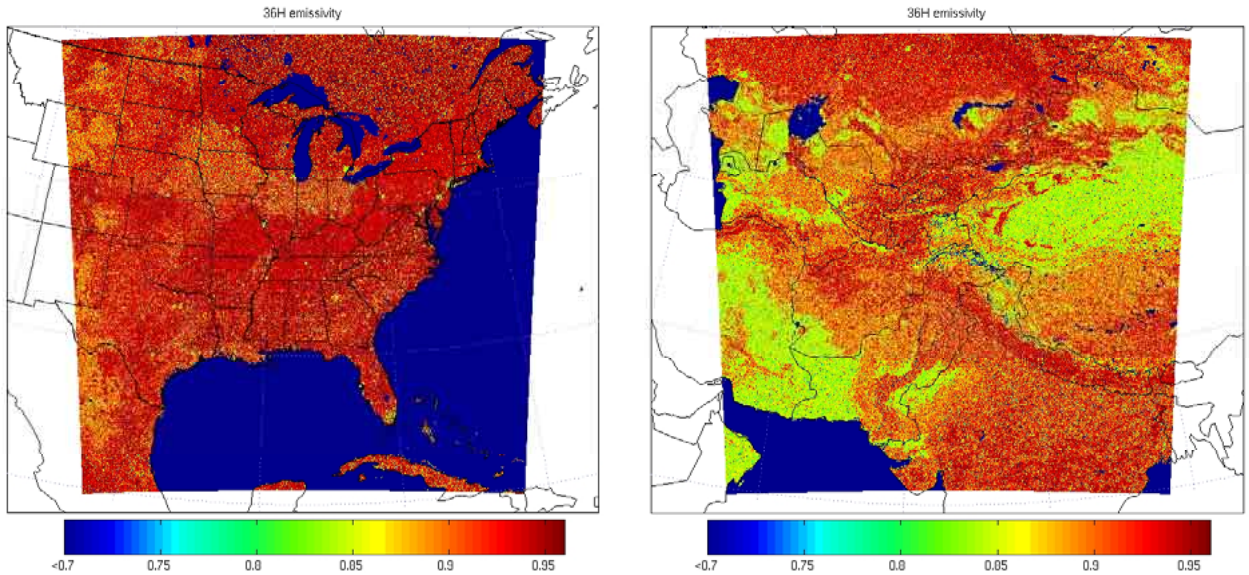

Figure 2-29: 36 GHz H-pol. emissivities for the North America and Asia test scenes.

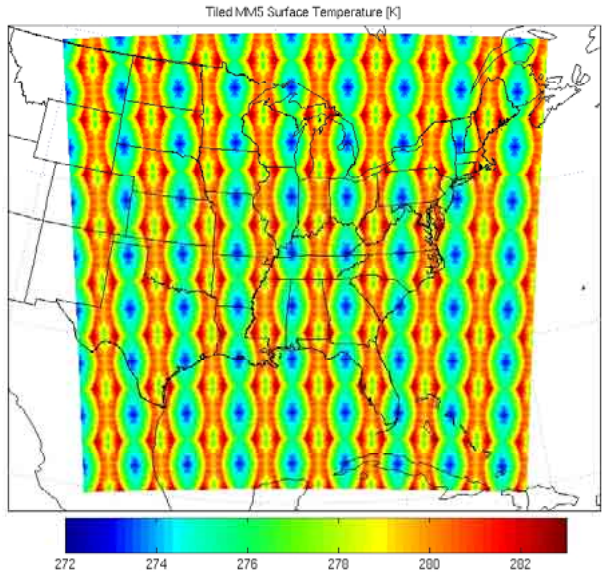

Figure 2-30: 3 km resolution MM5 surface temperature tiled to fill 1122x1122 North America domain.

Figure 2-31 shows 20 km CFOV-averaged surface temperature and 36H emissivity fields. These represent simulated retrieval products from the CMIS Core Module algorithm without retrieval error or sensor noise. Surface EDR algorithms may either conduct further processing on the sensor-CFOV data then map the products to the earth grid or map temperature and emissivity to the earth grid before further processing. In either case, the results of regridding tests with these two data sets can be generalized to apply to the derived products with scaling for dynamic range and heterogeneity as appropriate.

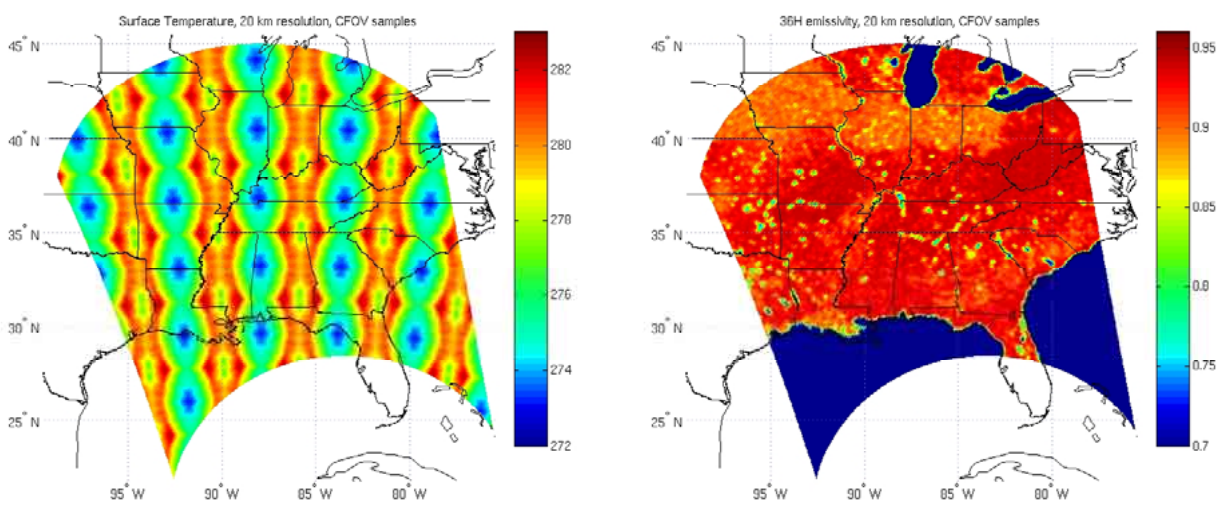

Figure 2-31: Surface temperature and 36 GHz H-pol. emissivity sampled by 20 km CMIS CFOVs.

Figure 2-32 shows examples of gridded surface temperature and 36H emissivity errors for 20 km CFOVs. (Statistical analysis of errors is given below.) The gridding algorithm was applied to the CFOV data in Figure 2-31. For truth, the parameters were also resampled from the high-resolution scene at each earth grid location using the CFOV spatial weighting pattern. This process adds an additional error element due to the fact that the CFOV pattern is not rotationally

symmetric and the truth-sampling process uses a pattern with fixed orientation whereas the sensor-CFOV is rotated in sync with the sensor scan. To test the impact of this effect, a CFOV was rotated through a full 360° near a coastline in the emissivity scene where the emissivity contrast is maximal. The range of sampled emissivities was only 0.004 with a range of CFOV-weighted land fraction of 0.01. For comparison, the land-water emissivity contrast was about 0.6 so the maximum emissivity deviation due to rotation effects was about 0.7% of the local emissivity dynamic range.

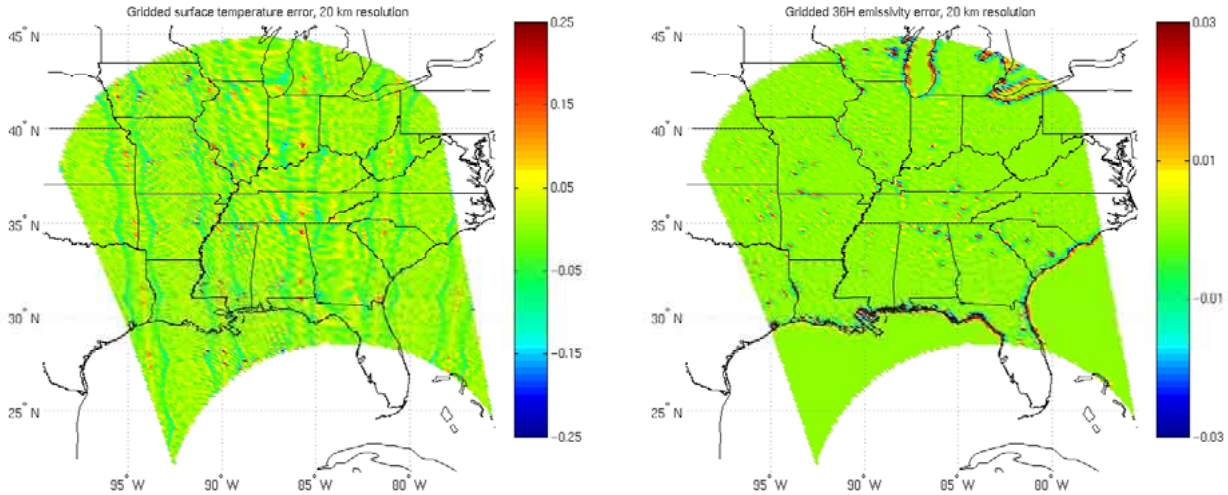

Figure 2-32: 20 km CFOV gridding error maps for surface temperature and 36H emissivity using Gaussian-weighted interpolation.

The surface temperature error map shows residual features from both the surface temperature map and the nearest-neighbor distance map (Figure 2-33). As discussed below, errors are likely to be larger when the grid point is farthest from the nearest CFOV even when interpolation methods are used. The emissivity error map is more dominated by scene features with errors increasing near coastlines and inland water bodies.

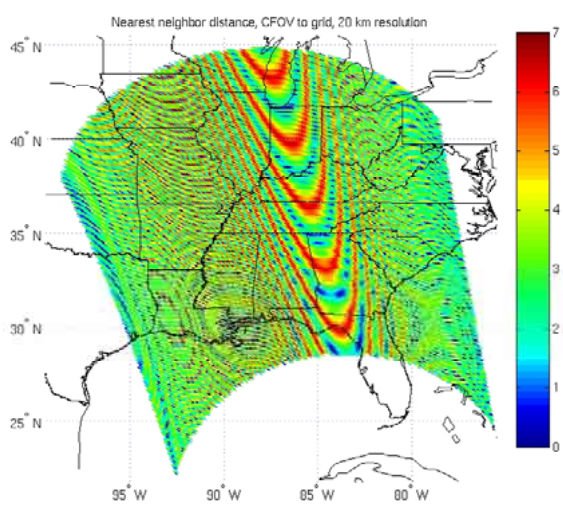

Figure 2-33: Map of distance from grid points to nearest 20 km CFOVs.

Errors for 50 km CFOVs (Figure 2-34) differ from the 20 km CFOVs in a least two respects. First, the surface temperature errors appear to be more a function the spatial heterogeneity of the local scene than the nearest neighbor distance. This is more a function of the spatial scales of the

particular scene chosen than of the regridding algorithm. That is, the surface temperature has a higher fraction of spatially challenging features (e.g., peaks and valleys) at the 50 km scale than at 20 km, where a higher fraction of points fall along easy-to-interpolate temperature gradients. Second, the coastline effects are spread over a broader area at 50 km sampling than 20 km. Coastlines are cause errors at both scales but a larger percentage of 50 km points straddle the coastline.

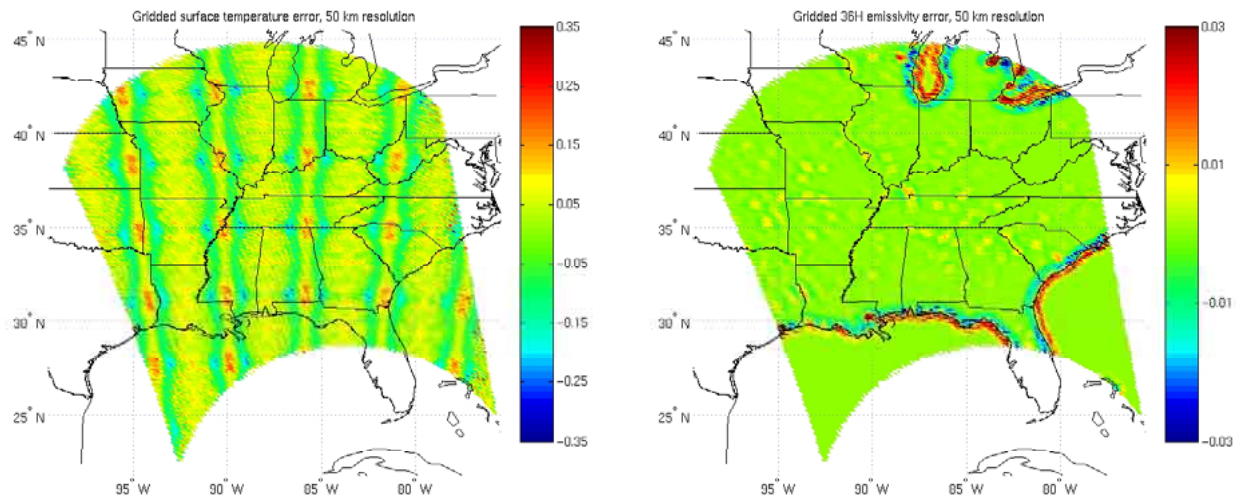

Figure 2-34: 50 km CFOV gridding error maps for surface temperature and 36H emissivity using Gaussian-weighted interpolation.

Example gridding errors for each scene are plotted in Figure 2-35 (20 km CFOV) and Figure 2-36 (50 km). Surface temperature errors are plotted as a function of nearest neighbor distance and emissivity errors as a function of the land fraction. As discussed above, surface temperature errors increase more dramatically with nearest neighbor distance for the 20 km CFOV than for the 50 km CFOV, although overall 50 km errors are larger. Also, 50 km emissivities show a more pronounced coastline-bias effect than 20 km: mostly-land points are negative biased and mostly-water points are positive biased.

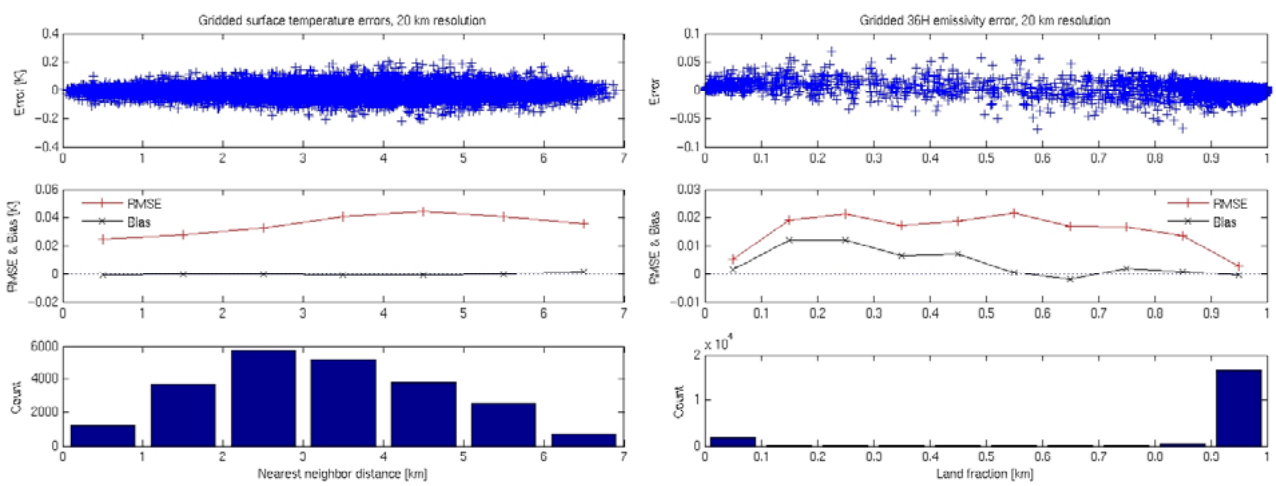

Figure 2-35: 20 km CFOV gridding error plots for surface temperature and 36H emissivity using Gaussian-weighted interpolation.

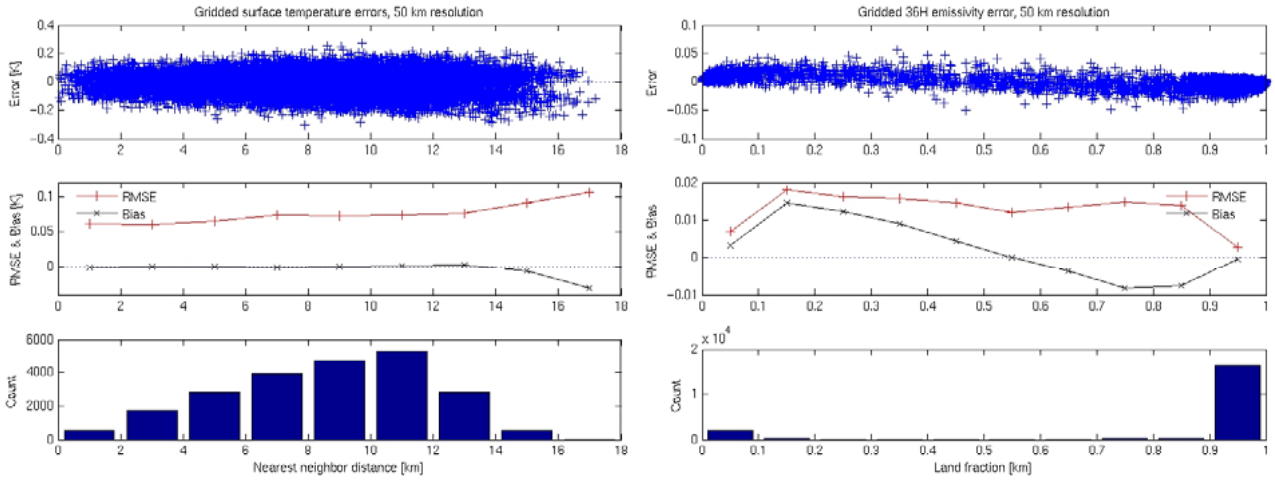

Figure 2-36: 50 km CFOV gridding error plots for surface temperature and 36H emissivity using Gaussian-weighted interpolation.

Table 2-1 through Table 2-4 give the statistical results for each gridding method at 20 and 50 km resolutions. Each method was tuned to determine the optimal number of neighbors to include in the weighted average (given in parentheses). Both 4- and 6-neighbor results are given for Gaussian-weighted interpolation with the best performances highlighted. Of all the methods, Gaussian-weighted interpolation performed best for both the 20 and 50 km CFOV scenes with the 20 km scene using six neighbors and the 50 km scene using four. The 50 km CFOVs are barely Nyquist sampled with 25 km along-track and 24 km along-scan sample spacing. The 20 km CFOVs are more highly sample along-scan (6 km) but undersampled along-track (12.5 km). The higher number of useful neighbors is probably due to the high along-scan sampling. As a first approximation, the number of samples should be set to 6, 6, 5, 5, and 4 for 15, 20, 25, 40, and 50 km CFOV gridding with further tuning as needed for each application. Gaussian-weighted interpolation was also tuned for the scale factor D_{hw} in (14), and the optimal value was found to be half the CFOV size.

Table 2-1: 20 km CFOV RMS surface temperature gridding errors [K] binned by nearest neighbor resampling distance. Error threshold is error exceeded by only 1% of points.

Gridding Method	Nearest Neighbor Distance [km] and Number in Bin							99% Error Threshold
	0-1	1-2	2-3	3-4	4-5	5-6	6-7	
	1206	3704	5779	5209	3790	2531	701	
Near. Neigh.	0.027	0.060	0.095	0.121	0.142	0.159	0.172	0.336
Inverse-D (4)	0.020	0.032	0.041	0.050	0.055	0.054	0.049	0.139
Inverse-D^2 (6)	0.024	0.038	0.035	0.039	0.043	0.046	0.046	0.114
Linear (4)	0.071	0.060	0.058	0.062	0.062	0.061	0.054	0.176
Gaussian (4)	0.026	0.031	0.039	0.050	0.055	0.047	0.040	0.136
Gaussian (6)	0.025	0.028	0.033	0.041	0.045	0.041	0.036	0.116

Table 2-2: 50 km CFOV RMS surface temperature gridding errors [K] binned by nearest neighbor resampling distance. Error threshold is error exceeded by only 1% of points.

Gridding Method	Nearest Neighbor Distance [km] and Number in Bin									99% Error Threshold
	0-2	2-4	4-6	6-8	8-10	10-12	12-14	14-16	16-18	
	553	1741	2833	3928	4715	5318	2818	562	41	
Near. Neigh.	0.045	0.098	0.150	0.211	0.275	0.322	0.366	0.442	0.501	0.747
Inverse-D (4)	0.047	0.073	0.091	0.106	0.114	0.117	0.127	0.109	0.109	0.283
Inverse-D^2 (4)	0.043	0.081	0.097	0.099	0.091	0.087	0.090	0.092	0.106	0.225
Linear (4)	0.182	0.174	0.171	0.170	0.167	0.157	0.167	0.130	0.112	0.440
Gaussian (4)	0.062	0.060	0.066	0.073	0.073	0.074	0.077	0.091	0.106	0.189
Gaussian (6)	0.062	0.064	0.070	0.076	0.078	0.080	0.083	0.095	0.108	0.202

Table 2-3: N. America 20 km CFOV RMS 36H emissivity gridding errors binned by CFOV-weighted land fraction at grid point. Error threshold is error exceeded by only 1% of points.

Gridding Method	Land Fraction and Number in Bin										99% Error Threshold
	0-.1	.1-.2	.2-.3	.3-.4	.4-.5	.5-.6	.6-.7	.7-.8	.8-.9	.9-1.0	
	1845	128	90	88	88	87	97	188	450	16489	
Near. Neigh.	0.007	0.042	0.051	0.054	0.062	0.055	0.052	0.041	0.028	0.005	0.0565
Inverse-D (4)	0.005	0.019	0.022	0.019	0.021	0.022	0.019	0.018	0.014	0.003	0.0251
Inverse-D^2 (6)	0.005	0.019	0.022	0.018	0.020	0.021	0.017	0.017	0.013	0.003	0.0242
Linear (4)	0.007	0.022	0.029	0.024	0.026	0.025	0.027	0.021	0.017	0.004	0.0329
Gaussian (4)	0.005	0.020	0.022	0.020	0.022	0.026	0.020	0.017	0.014	0.003	0.0240
Gaussian (6)	0.005	0.019	0.021	0.017	0.019	0.022	0.017	0.017	0.014	0.003	0.0228

Table 2-4: N. America 50 km CFOV RMS 36H emissivity gridding errors binned by CFOV-weighted land fraction at grid point. Error threshold is error exceeded by only 1% of points.

Gridding Method	Land Fraction and Number in Bin										99% Error Threshold
	0-.1	.1-.2	.2-.3	.3-.4	.4-.5	.5-.6	.6-.7	.7-.8	.8-.9	.9-1.0	
	2042	215	159	160	151	170	171	214	365	16289	
Near. Neigh.	0.012	0.045	0.058	0.063	0.064	0.067	0.064	0.058	0.039	0.006	0.0861
Inverse-D (4)	0.009	0.025	0.020	0.021	0.022	0.019	0.020	0.021	0.019	0.004	0.0333
Inverse-D^2 (4)	0.007	0.020	0.018	0.020	0.021	0.019	0.020	0.019	0.016	0.003	0.0293
Linear (4)	0.012	0.036	0.031	0.035	0.036	0.033	0.033	0.034	0.028	0.005	0.0502
Gaussian (4)	0.007	0.018	0.016	0.016	0.015	0.012	0.013	0.015	0.014	0.003	0.0236
Gaussian (6)	0.008	0.020	0.017	0.015	0.014	0.012	0.014	0.015	0.015	0.003	0.0247

For the Asia scene, both the 20 and 50 km CFOV data had 0.002 RMS emissivity errors in the 0.9-1 land fraction bin with Gaussian interpolation. This is even lower than the 0.003 RMS errors in the same bin for the N. America scenes, with the difference attributable to the higher percentage of near-100% land cells in the Asia scene even within the 0.9-1 land fraction bin. This puts an upper limit of 0.002 on the regridding error when the surface is known to have emissivity contrasts below ~0.2. Any scenes that mix high and low emissivity surfaces (e.g., water, ice, snow, and some deserts) in more than trivial fractions will have higher remapping errors with little dependence on the fractional coverage amount.

2.5. Practical Considerations

2.5.1. Algorithm Tuning

The RF parameter choices given in Section 2.2.1.2 are our baseline values. In the course of on-orbit algorithm validation, additional tuning may be necessary to optimize its performance with regard to grid spacing, data density and quality (including the use of realistic input QC flags), and the sensor characteristics. At the present stage, subjective evaluation of the analyzed fields

and objective methods (comparison against withheld data values) have been used to assess performance of the regridding modules when applied to simulated test scenes.

The baseline earth regridding algorithm parameter choices are given in Section 2.4. Additional tuning and reconfiguration may be needed for specific Core Module parameters, surface EDRs, and other surface parameters. For example, where surface EDR algorithms take gridded data inputs, the algorithm may choose to resample the gridded data to the sensor-CFOV locations or earth-grid the Core Module products before performing the retrieval. Also, optimal tuning may differ for snow and ice EDRs where their spatial variability differs from the scenes tested above.

2.5.2. Recursive Filter Numerical Computation Considerations

The main factors affecting the speed of the recursive filter regridding programs are the spatial extent of input data to be processed (essentially, the number of input scanlines) and the spacing of the analysis grid used by the recursive filter. The number of scanlines affects the timing approximately in a linear manner, while the grid spacing effect is quadratic. For the simulated orbital set analyzed in Section 2.3, which contains 251 scanlines at 12.5-km cross-scan spacing (i.e., roughly 1/12 of an orbit), the timing results for the vertical remapping and FOV interpolation on a RedHat Linux operating system using 2x AMD Athlon 2000 MP are given in Table 2-5 and Table 2-6, respectively.

Note that the grid spacing Δ listed in Table 2-5 refers to the analysis grid employed by the recursive filter, with the remapped EDR product generated at the requirement-specified resolution (40- and 15-km for AVTP and AVMP, respectively) by subsampling of the analysis grid (consequently, Δ = 10 km is not available in the AVMP case as it is not an integer subinterval of 15 km). The default settings are highlighted in red. These timing results serve mainly for comparison purposes and are likely to be updated in the future.

Table 2-5: Timing of Vertical Remapping [CPU + system, seconds].

	AVTP	AVMP
Δ = 40 or 15 km	16 + 0.5	40 + 1
Δ = 10 km	18 + 0.5	N/A
Δ = 5 km	45 + 0.5	67 + 1.9
Δ = 2.5 km	177 + 1.4	189 + 3.4

Note that the timing results for FOV interpolation shown in Table 2-6 have been obtained in runs employing five passes for Δ = 10 and 15 (or 40) km and seven passes for Δ = 5 and 2.5 km, while the RMS errors shown in Figure 2-15 through Figure 2-20 have been obtained in 5-pass runs (since the two extra passes are not necessary in data-dense areas). The timing is roughly linear in the number of passes.

Table 2-6: Timing of FOV Interpolation [CPU + system, seconds].

	50- to 15-km	50- to 40-km	20- to 15-km
Δ = 15 or 40 km	23 + 2	13 + 1	40 + 2
Δ = 10 km	37 + 1	29 + 1	51 + 2
Δ = 5 km	148 + 2	143 + 12	163 + 4
Δ = 2.5 km	699 + 7	703 + 7	723 + 10

2.5.3. Quality Control and Diagnostics

The main quality control index is the smoothing parameter $\beta = 1 - \alpha$ defined in Section 2.2.1.1. As described in HayP95, this parameter controls the spatial scale of the filter and a small value mixes substantially across several grid spaces (this occurs, for example, in data gaps). The actual values for this parameter are dependent on resolution and radius-of-influence (see HayP95, Appendix B) and are therefore most informative when used in a relative sense for a given set of Δ and R_m. For both the vertical remapping and the FOV interpolation programs, this parameter is averaged after the final analysis pass over all variables to which the recursive filter is applied per analysis point. Specifically, it is averaged over all levels in the vertical profile for VR and, for CFOV interpolation, it is averaged overall variables in the Core Module state vector. In the case of FOV interpolation, this averaged parameter is further bi-linearly interpolated to the output FOV locations. The QC parameter stored in the output file is an integer number obtained by multiplying β by a factor of 100. Examples of the output QC fields are shown in Section 2.3.4.2.

In addition to the QC field, the occurrence of non-standard situations during execution of the regridding programs is indicated by means of warning messages.

2.5.4. Exception and Error Handling

The regridding programs contain a series of exit statements for situations when the execution cannot be completed, e.g., for a large number of missing data or inconsistent spatial coverage between the input Core Module data file and the SDR file that specifies the locations at which to interpolate (for FOV interpolation). In each case when the programs are terminated prematurely, an error message explaining the nature of the problem is printed. In situations when the programs are directed to execute in a non-default mode (e.g., varying the prescribed segment length due to missing scanlines), a warning message is printed.

3. Imagery

3.1. Objectives

The imagery EDR complements the environmental EDR products. Because the imagery brightness temperatures are equivalent to the EDR algorithm input values, the imagery EDR facilitates visualization of processes at work in the EDR source data as well as reprocessing using user-defined alternatives to the EDR algorithms. Imagery is also potentially useful for phenomenon location (e.g., storm centers) and tracking (bulk sea ice motion), detection and monitoring of transient signals (e.g., RFI), and data quality verification.

3.2. SRD Requirements

The text below and Table 3-1 are the portions of CMIS Sensor Requirements Document (SRD) section 3.2.1.1.1.1 that apply directly to the Imagery EDR.

Imagery

Brightness temperature data from each microwave channel shall be available for display at the sampled resolution. The threshold horizontal spatial resolution (HSR) is to be consistent with the performance of the related EDRs. The display capability for all imagery should be consistent with the dynamic range of any CMIS channel.

Table 3-1: SRD requirements for the imagery EDR.

Para. No.		Thresholds	Objectives
	a. Horizontal Spatial Resolution		
C40.2.3.1-1	1. Global	Consistent with related EDRs	(TBD)
C40.2.3.1-2	b. Horizontal Reporting Interval	Consistent with related EDRs	(TBD)
C40.2.3.1-3	c. Horizontal Coverage	Global	Global
C40.2.3.1-4	Not Used		
C40.2.3.1-5	d. Measurement Range	Dynamic range of all measurement channels	Dynamic range of all measurement channels
C40.2.3.1-6	e. Measurement Uncertainty (TBR)	Derived	Derived
C40.2.3.1-7	f. Mapping Uncertainty	3 km (TBR)	(TBD)

In addition to these requirements, the SRD specifies:

1. "Science algorithms shall process CMIS data, and other data as required, to provide the [EDRs] assigned to CMIS." (SRD, paragraph SRDC3.1.4.2-1)
2. "Specified EDR performance shall be obtained for any of the orbits described in paragraph 3.1.6.3 …" (SRDC3.1.6.3-2)
3. "As a minimum, the EDR requirements shall be satisfied at the threshold level." (SRDC3.2.1.1.1-3)
4. "… the contractor shall identify the requirements which are not fully satisfied, and specify the conditions when they will not be satisfied." (SRCD3.2.1.1.1-4)
5. "… CMIS shall satisfy the EDR Thresholds associated with cloudy conditions under all measurement conditions …" (SRD SRDC3.2.1.1.1.1-1)

Also note that the CMIS system consists "of all ground and spaceborne hardware and software necessary to perform calibrated, microwave radiometric measurements from space and the software and science algorithms necessary to process … these measurement into a format consistent with the requirements of the assigned [EDRs]." (SRD, section 3.1.1)

3.3. Algorithm Description

Figure 3-1 shows the CMIS processing flow leading to the generation of data for the Imagery EDR. TDR/SDR processing is discussed in *CMIS ATBD Vol. 17 TDR and SDR Algorithms*. The EDR pre-processing module includes footprint matching (described in *CMIS ATBD Vol. 1 Part 2 Footprint Matching*) and radiative transfer model (RTM) calibration of SDR data that removes biases derived post-launch (also described in the *CMIS ATBD Vol. 17 TDR and SDR Algorithm*). The output of the pre-processing module is empirically-corrected brightness temperature (ECBT).

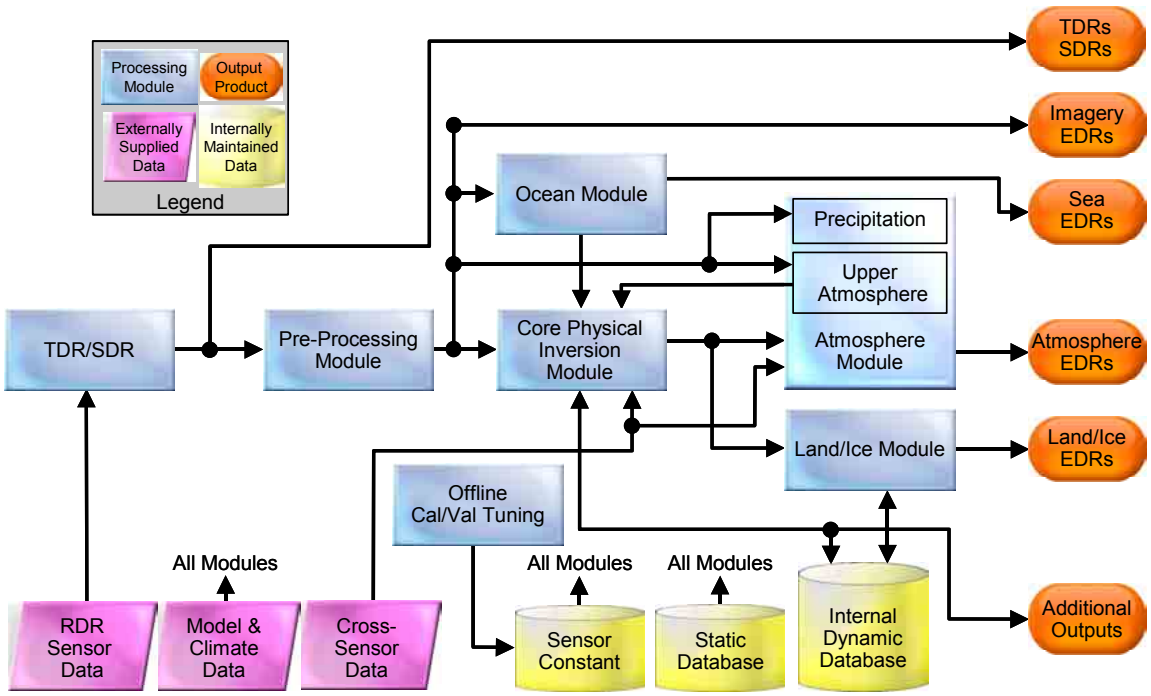

Figure 3-1: Overall CMIS processing flow including imagery EDR.

3.4. Algorithm Performance

Table 3-2 gives the nominal performance characteristics for the imagery EDR. The table gives nominal HSR performance values—typically equal to the EDR horizontal cell sizes—whereas *CMIS ATBD Vol. 1 Part 2 Footprint Matching* provides estimated composite footprint performance per each channel and horizontal cell size combination. Imagery data for each channel is provided at *every* EDR resolution where a high-fidelity footprint match can be made. Table 3-3 lists the few exceptions where the sensor footprints are too large to report useful imagery data for some of the EDR resolutions. Detailed analysis of composite footprint processing and the associated noise and spatial match trade-offs are given in *CMIS ATBD Vol. 1 Part 2 Footprint Matching*.

Table 3-2: Imagery EDR nominal performance.

Para. No.		Thresholds	Objectives	Performance
	a. Horizontal Spatial Resolution			
C40.2.3.1-1	1. Global	Consistent with related EDRs	(TBD)	15, 20, 25, 40, 50, 56×35, 86×52, and 200 km
C40.2.3.1-2	b. Horizontal Reporting Interval	Consistent with related EDRs	(TBD)	15, 20, 25, 40, 50, 56×35, 86×52, and 200 km
C40.2.3.1-3	c. Horizontal Coverage	Global	Global	Global
C40.2.3.1-4	Not Used			
C40.2.3.1-5	d. Measurement Range	Dynamic range of all measurement channels	Dynamic range of all measurement channels	Dynamic range of all measurement channels
C40.2.3.1-6	e. Measurement Uncertainty	Derived	Derived	Derived
C40.2.3.1-7	f. Mapping Uncertainty	3 km (TBR)	(TBD)	3 km

Table 3-3: Imagery EDR excluded conditions.

Exclusion	Rationale
6 GHz at spatial resolution < 50 km	Sensor resolution is flowed from requirements for category 1 and 2 EDRs and does not support creating imagery for this channel at higher resolution without substantial noise amplification
10 GHz at spatial resolution < 40 km	Sensor resolution is flowed from requirements for category 1 and 2 EDRs and does not support creating imagery for this channel at higher resolution without substantial noise amplification
18 and 23 GHz at spatial resolution < 20 km	Sensor resolution is flowed from requirements for category 1 and 2 EDRs and does not support creating imagery for this channel at higher resolution without substantial noise amplification

4. References

Barnes, S. L., 1964: A technique for maximizing details in numerical weather map analysis. *J. Appl. Meteor.*, **3**, 396–409.

Cressman, G. P., 1959: An operational objective analysis system. *Mon. Wea. Rev.*, **87**, 367–374.

Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press.

Hayden, C. M., and R. J. Purser, 1988: Three-dimensional recursive filter objective analysis of meteorological fields. *CIMSS View*, **4**, 1–5.

Hayden, C. M., and R. J. Purser, 1995: Recursive filter objective analysis of meteorological fields: applications to NESDIS operational processing. *J. Appl. Meteor.*, **34**, 3–15.

Koch, S. E., M. DesJardins, and P. J. Kocin, 1983: An interactive Barnes objective map analysis scheme for use with satellite and conventional data. *J. Clim. Appl. Meteor.,* **22**, 1487–1503.

Lorenc, A., 1992: Iterative analysis using covariance functions and filters. *Quart. J. Roy. Meteor. Soc.*, **118**, 569-591.

Panegrossi, G., A. Tassa, S. Di Michele, G. J. Tripoli, A. Mugnai, E. A. Smith, 2001: Using TRMM observations to improve numerical simulations of precipitation within tropical cyclones, Proc. Of Symposium on Precipitation Extremes: Prediction, Impacts, and responses, *81 AMS Meeting*, Albuquerque, NM, pp. 324-326.

Seaman, R. S., 1989: Tuning the Barnes objective analysis parameters by statistical interpolation theory. *J. Atmos. Ocean. Technol.*, **6**, 993–1000.

Thiebaux, H. J., and M. A. Pedder, 1987: *Spatial Objective Analysis: With Applications in Atmospheric Science*. Academic Press, London.

Tripoli, 1992a: A nonhydrostatic model designed to simulate scale interaction, *Mon. Wea. Rev.*, 120, 1342-1359.

Tripoli, 1992b: An explicit three dimensional nonhydrostatic numerical simulation of a tropical cyclone. *Meteor. Atmos. Phys.*, **49**, 229-254.

Tripoli et al., 2000: Orographically induced flash floods on the northern Italian coast, Ninth Conference on Mountain Meteorology, *AMS*, Aspen, CO, 336-339.

APPENDIX

A. Product Grid for Recursive Filter Applications

A.1. Introduction

The recursive filter (RF) is used in two basic applications: 1. Remapping of CMIS core module retrievals obtained at FOV locations determined by the orbital and viewing geometry to a regular grid defined by the EDR reporting requirements (henceforth this application will be referred to as "vertical remapping") and 2. Core module retrieval interpolation from coarse to fine FOV spacing (henceforth referred to as "FOV interpolation"). In principle, the recursive filter should work on the most general format of core module output files, such as full-swath data from orbits completed during a single day. In practice, in order to keep the computational costs within manageable limits, and to meet accuracy and reporting requirements (to be described below), it is necessary to divide the full-swath data into segments, defined as subsets of scanlines contained in the original full-swath file. It is required that the segmenting does not introduce discontinuities in the remapped retrievals (i.e., the recursive filter should lead to a smooth "seaming" at the segment boundaries). In practice, since the recursive filter employs a regular grid, with each segment there is associated a product grid of length D, width W, and spacing Δ (all in kilometers), on which valid retrieval fields are generated. The product grids from all segments combined constitute the complete product grid for the file being processed.

The purpose of this appendix is to describe the computational procedure adopted for defining the RF product grid. The general aspects of the procedure, applicable to both the vertical remapping and FOV interpolation, are presented in Section A.2, followed by a description of the application-specific aspects in Section A.3.

A.2. General Computational Aspects

The complete product grid is a quasi-regular two-dimensional grid, with the x- and y-axes parallel and perpendicular, respectively, to the sub-satellite track. The method of determining the start- and end-points of the grid is application-dependent (input- and output-driven for vertical remapping and FOV interpolation, respectively) and is described in detail in Section A.3. Regardless of the application, the scanline center positions from the input data file are the points of reference for defining product grid segments. The grid definition process thus relies on the geolocation data that accompany the input data, passed down from the CMIS geolocation algorithm, and no redundant geolocation computations need to be made for the gridding process.

A.2.1. Segmenting

For the first data segment, the position $\vec{p}_{1,s}$ (i.e., latitude and longitude) of its starting point is taken as the center-point of a selected input scanline $n_{1,s}$, with the method of selecting $n_{1,s}$ described in Section A.3. The position $\vec{p}_{1,e}$ of the first segment's end-point is then found by locating the point at a distance D downtrack from $\vec{p}_{1,s}$. For this purpose, we interpolate between locations of scanline-center FOVs.

Specifically, $\vec{p}_{1,e}$ is determined as $(1-a)\vec{p}_{1,n_{1,e}} + a\vec{p}_{1,n_{1,e}+1}$ where $n_{1,e}$ denotes the index of the scanline satisfying the conditions:

$$d(\vec{p}_{n_{1,e}},\vec{p}_{1,s}) \le D \;\; and \;\; d(\vec{p}_{n_{1,e}+1},\vec{p}_{1,s}) \ge D \tag{A1}$$

d represents the great-circle distance between two points, and $\vec{p}_n$ are the geographical coordinates of the center-point for the nth scanline.

The weight a is equal to:

$$a \equiv \left[D - d\left(\vec{p}_{n_{1,e}}, \vec{p}_{1,s}\right)\right] \Big/ \left[d\left(\vec{p}_{n_{1,e}+1}, \vec{p}_{1,s}\right) - d\left(\vec{p}_{n_{1,e}}, \vec{p}_{1,s}\right)\right] \tag{A2}$$

In practice, the index $n_{1,e}$ is found using a first guess $n_{1,e} = n_{1,s} + \text{int}(D/\delta)$, where δ is the nominal cross-scan spacing of the FOVs (12.5 km for CMIS, or 25 km when the FOV size is 50 km). If, for the first guess, $d(\vec{p}_{n_{1,e}}, \vec{p}_{1,s}) \ge D$, then $n_{1,e}$ is decremented until $d(\vec{p}_{n_{1,e}}, \vec{p}_{1,s}) \ge D$. If, for the first guess, $d(\vec{p}_{n_{1,e}+1}, \vec{p}_{1,s}) \le D$, then $n_{1,e}$ is incremented until conditions (A1) are satisfied.

For subsequent segments, the coordinates $\vec{p}_{i+1,s}$ are set to the coordinates of $\vec{p}_{i,e} + P$, where P is application-dependent (see Section A.3) and the distance is augmented along the sub-satellite track (defined by the azimuth angle for the line $\vec{p}_{n_{i,e}} \to \vec{p}_{n_{i,e}+1}$). In this manner, a quasi-continuity of the product grid across adjacent segments is maintained.

A.2.2. Coordinate Transformation

As mentioned above, it is desired that the x-axis of the product grid be parallel to the sub-satellite track. To meet this objective and to satisfy reporting requirements with regard to the spatial resolution of the output product, for each segment the RF analysis is performed in a rotated spherical coordinate system (λ', φ'), in which $\vec{p}_{i,s}$ is the point $\lambda' = 0, \varphi' = 0$ and the line $\varphi' = 0$ corresponds to the line $\vec{p}_{i,s} \to \vec{p}_{i,e}$ (λ' and φ' are the angular distances along- and cross-track, respectively). The transformed coordinates in distance units are $x=\lambda' r_e$ and $y=\varphi' r_e$, where r_e is the Earth radius, with the spacing along the x-axis (along track) equal to Δ and the number of points equal to $m_x = D/\Delta + 1$. In the y-direction (perpendicular to the track), the spacing is also equal to Δ and the number of points is $m_y = W/\Delta + 1$.

The geographical coordinates $\vec{p} \equiv (\lambda, \varphi)$ (λ and φ are the longitude and latitude of $\vec{p}$) of any observation points used by the recursive filter to compute field values at the product grid are transformed to this new coordinate system. This transformation is accomplished by computing the vector $\vec{n}' = [\cos\varphi' \cos\lambda', \cos\varphi' \sin\lambda', \sin\varphi']$ normal to the Earth's surface at $\vec{p}$ in the rotated coordinate system. This rotated normal vector is related to the normal vector $\vec{n} = [\cos\varphi \cos\lambda, \cos\varphi \sin\lambda, \sin\varphi]$ in the original geographical coordinates via $\vec{n}' = \Re\vec{n}$, where $\Re$ is the rotation matrix defined as $\Re = \Re_3 \Re_2 \Re_1$, and the three individual rotation matrixes are:

$$\Re_1 = \begin{pmatrix} \cos\lambda_{i,s} & \sin\lambda_{i,s} & 0 \\ -\sin\lambda_{i,s} & \cos\lambda_{i,s} & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \Re_2 = \begin{pmatrix} \cos\varphi_{i,s} & 0 & \sin\varphi_{i,s} \\ 0 & 1 & 0 \\ -\sin\varphi_{i,s} & 0 & \cos\varphi_{i,s} \end{pmatrix}$$

$$\Re_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha_i & \sin\alpha_i \\ 0 & -\sin\alpha_i & \cos\alpha_i \end{pmatrix} \tag{A3}$$

In the above expressions, $\lambda_{i,s}$ and $\varphi_{i,s}$ are the longitude and latitude of $\vec{p}_{i,s}$ and α_i is the azimuth angle for the segment (relative to the north), defined as the azimuth angle of the line $\vec{p}_{i,s} \to \vec{p}_{i,e}$

(note that $\Re$ is computed only once for each segment). Once $\vec{n}' \equiv [n_1', n_2', n_3']$ is computed, the transformed coordinates of $\vec{p}$ are given as $\lambda' = \tan^{-1}(n_2'/n_1')$ and $\varphi' = \tan^{-1}\left(n_3'/\sqrt{n_1'^2 + n_2'^2}\right)$. Note that for the likely segments lengths (see below), the arguments of the arctangents are never greater than 1, so these formulas are computationally stable.

An example of a simulated CMIS scanline set before and after coordinate transformation is shown in Figure A-1 and Figure A-2 respectively. As a result of the Earth's rotation and the conical viewing geometry for CMIS, there is an asymmetry in the rotated FOV locations relative to the orbit track. The phenomenon can be seen conceptually by considering a great-circle line drawn perpendicular to the orbit subtrack ahead of the satellite. As the scan swath arrives at that line, the first part of the line to be measured is the leading edge of the swath at the center of the arcing scan line. Several more rotations of the conical scanner, as the satellite moves forward, are needed before the perpendicular line has been covered by measurements all the way to the swath edge, and the Earth rotates during that time. The magnitude S of the resulting offset is a function of latitude, with a maximum value of about 40 km at the equator and vanishing as the poles are approached. In order to account for this offset and generate a product grid centered on the orbit track, for each segment the coordinate transformation is performed relative to modified pivot points $\vec{p}_{i,s}^{S}$ and $\vec{p}_{i,e}^{S}$ obtained from $\vec{p}_{i,s}$ and $\vec{p}_{i,e}$ by applying a lateral shift $S_{i,s}$ and $S_{i,e}$, respectively, in the direction perpendicular to the track (in other words, $\vec{p}_{i,s}^{S}$ and $\vec{p}_{i,e}^{S}$ are used to define $\Re$).

At any point, this shift is computed as:

$$S = (S_l + S_r)/2 \tag{A4}$$

where S_l and S_r are the cross-track components of Earth's rotation at the left- and right edge of the swath, respectively.

These terms are computed as:

$$S_{l,r} = \chi_e \cos(\varphi_{l,r}) \cos\alpha_H \tag{A5}$$

where α_H is the satellite heading azimuth angle relative to the north (set to the azimuth angle defined by the scanline center-points bracketing $\vec{p}_{i,s}$ or $\vec{p}_{i,e}$), φ_l and φ_r are the latitudes of the left and right edge, respectively, and $\chi_e = 2\pi r_e t_{ce}/(1\text{ day})$ is the amount of Earth's rotation (in km) completed during the time t_{ce} between the center and edge of swath. For CMIS, t_{ce} is computed as $t_{ce} = t_s(m_s - 1)/2$, where t_s and m_s are the sampling interval and the number of samples per scanline, respectively.

Using a planar geometry approximation, the latitudes of the swath edges are computed using the expression:

$$\varphi_{l,r} = \varphi_c \pm \Omega \sin\alpha_H \tag{A6}$$

where φ_c is the latitude of the swath center and Ω is equal to half the angular swath width (the angle between the Earth radius vectors at opposite ends of a scan line). The planar approximation breaks down near the poles, but this introduces small error since the offset vanishes near the poles.

The angular swath width is calculated from:

$$2\Omega = \cos^{-1}\left[\cos(2\gamma)\sin^2\beta + \cos^2\beta\right] \tag{A7}$$

where γ is the azimuth angle for the line joining the center of the scan with the swath edge (half the active scan azimuth range) and β is the angle between the radius of the Earth at the sub-satellite point and the radius at a footprint (constant for conical scanner).

For CMIS, these two angles are computed as:

$$\begin{aligned} \gamma &= t_{ce}\omega \\ \beta &= EIA - \theta \end{aligned} \tag{A8}$$

where ω is the nominal rotation rate, EIA is the Earth incidence (zenith) angle, and $\theta = \sin^{-1}\left(\sin(EIA)\dfrac{r_e}{r_e + z_s}\right)$ is the nadir angle (cone angle of conical scan).

In order to apply the above expressions, the parameters z_s and ω are obtained from the sensor constants file, m_s, t_s, α_H, and EIA are derived from the input data, and the shift S is computed as a function of φ_c.

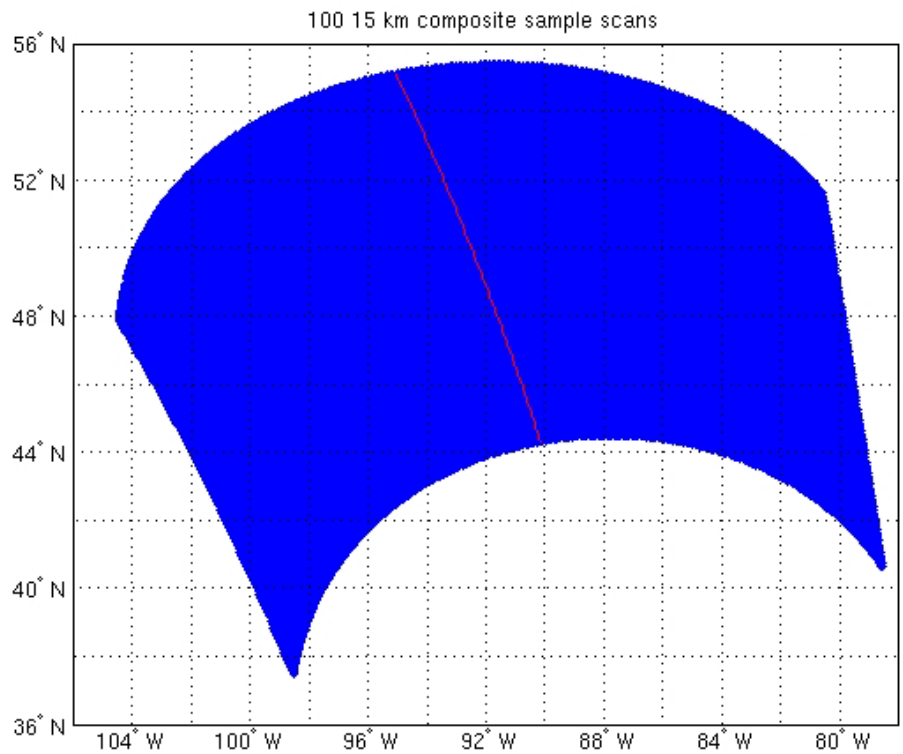

Figure A-1: Simulated CMIS FOVs.
Marked in red is the line $\vec{p}_{i,s} \rightarrow \vec{p}_{i,e}$ used to define the coordinate transformation described in the text.

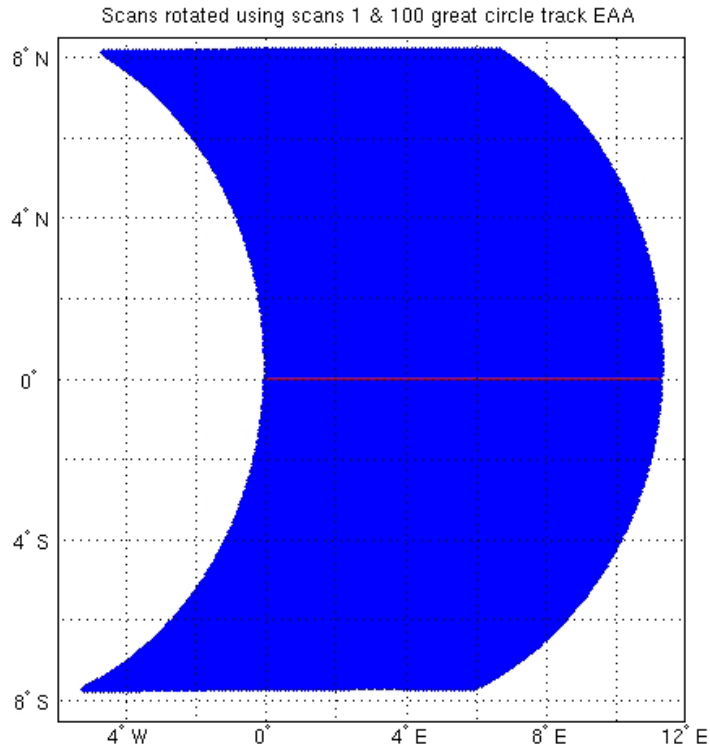

Figure A-2: Positions of the CMIS FOVs shown in Figure A-1 after coordinate transformation (the labels along the axes are for illustrative purposes only, as they do not correspond to the E, W, N, S geographic directions after the transformation). Note the asymmetry of the transformed positions relative to the scan line centers.

A.2.3. Optimal Segment Length

A second relevant effect of Earth rotation is to cause satellite orbit subtrack path to depart from a great circle drawn across the globe. If a great circle is drawn between two points along the subtrack, there will be a deflection Λ of the subtrack relative to the great circle along the path between the end points. Since the recursive filter product grid is defined with respect to a great circle, large deflections cause degradation of the remapped product at the edge of scan. An upper bound on D can be set by requiring that, for each segment, the deflection Λ of the center-points of scanlines $n_{i,s} + 1$, $n_{i,s} + 2, \ldots, n_{i,e} - 1$, $n_{i,e}$ from the great circle defined by $\vec{p}_{i,s}$ and $\vec{p}_{i,e}$ does not exceed a preset value. While Λ is by design very close to 0 at $n_{i,s} + 1$ and $n_{i,e}$, it can assume large values at the intermediate scanlines as the distance between $\vec{p}_{i,s}$ and $\vec{p}_{i,e}$ grows large, especially in high latitudes. This effect is illustrated in Figure A-3, which has been computed from AMSU orbital data. The latitude-dependence of the maximum separation between $\vec{p}_{i,s}$ and $\vec{p}_{i,e}$ (i.e., D) consistent with prescribed deflection thresholds is shown in Figure A-4. The tolerable deflection is application-dependent (see Section A.3), but should always be small in relation to W. For example, if Λ is limited to 6 km, then the global bound on D is about 1500 km.

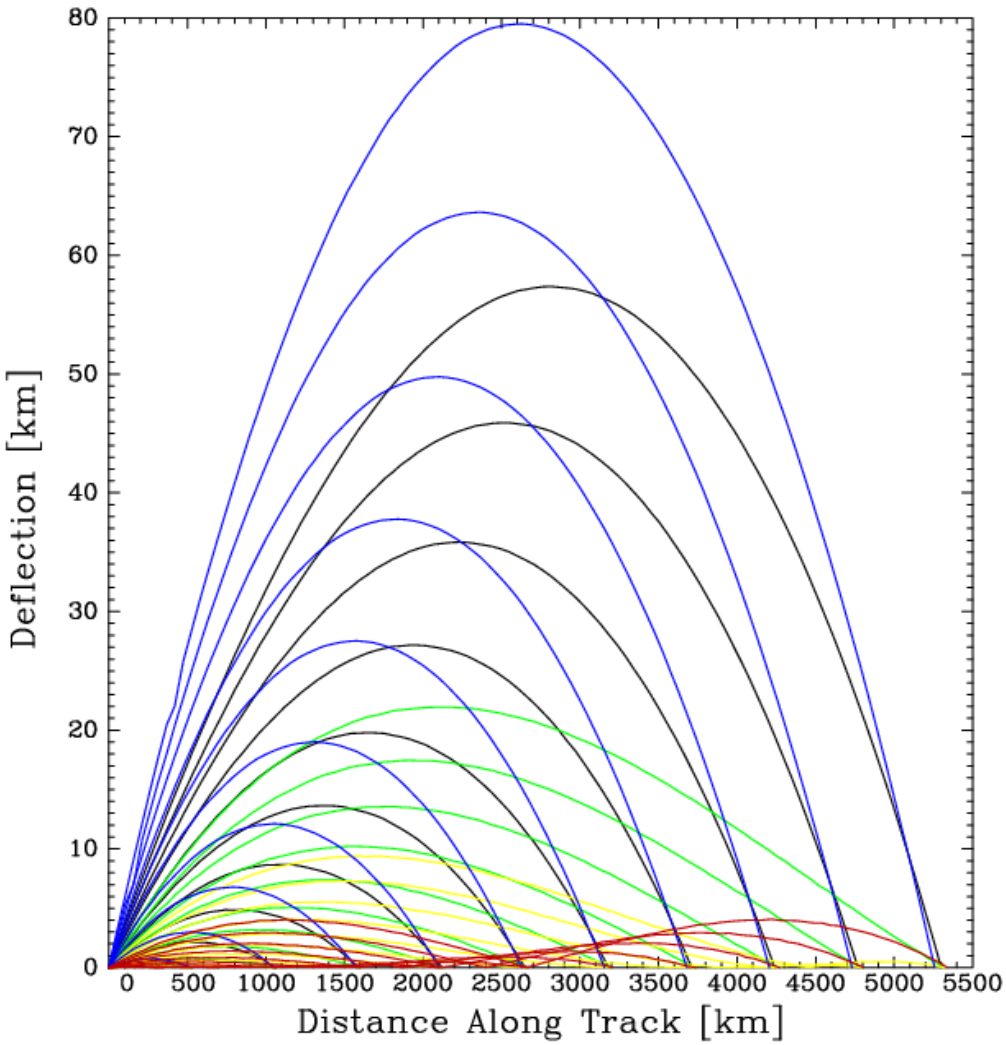

Figure A-3: Deflection of the sub-satellite points from the great circle defined by two reference sub-satellite points.

The different colors represent families of curves computed when the mid-point between the sub-satellite reference points is located near 80°, 45°, 15°, 5°, and the equator (blue, black, green, yellow, and red, respectively). Within each family, the different curves represent deflections computed by varying the distance between the reference points (the distance is equal to the point at which the deflection goes to zero).

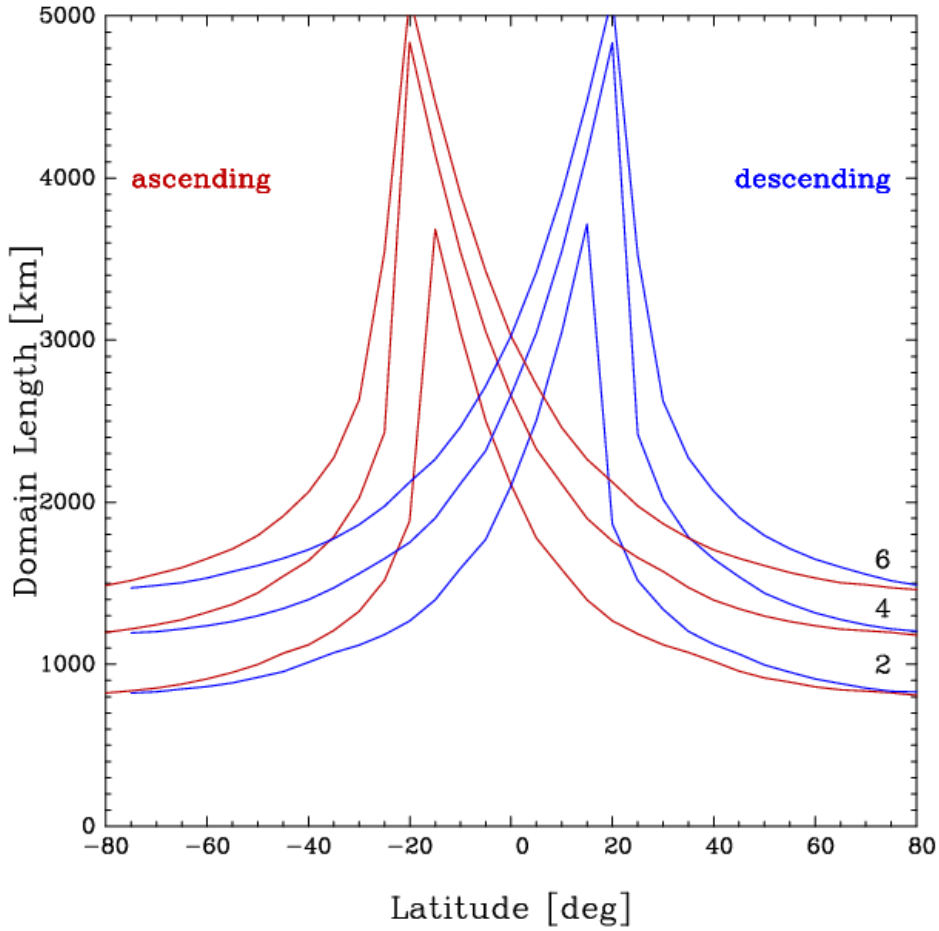

Figure A-4: Maximum distance along the track satisfying a prescribed deflection criterion in kilometers (shown as the black labels to the right) as a function of latitude.

Unlike the curves in Figure A-3: , which have been generated by varying the position of the mid-point between $\vec{p}_{i,s}$ and $\vec{p}_{i,e}$, the curves in this figure have been computed by varying the latitude of $\vec{p}_{i,s}$, thus necessitating a separate consideration of ascending and descending orbits.

A.2.4. Computational Boundary Areas

Consider the possibility that a gap in the input data occurs along the seam between two segments of the product grid. Such a gap may be caused by an area of precipitation causing core model retrievals to fail QC. The RF analysis results will effectively be extrapolations in the gap area, with one segment extrapolating from one side of the gap and the other segment extrapolating from the other side. This behavior could cause discontinuities in the product data when the segments are merged.

To avoid this problem, the length L of the analysis grid used by the recursive filter is longer than the product grid D. The width G of data gaps is likely to be on the order of the size of large-scale precipitation areas, usually 200-300 km. A related concern is boundary effects in the recursive filter related to the finite size of D. With regard to boundary effects, since the recursive filter is approximately Gaussian with a characteristic spatial scale R, it may be expected that the effect of employing a finite domain extends about R from the boundaries (in reality, this effect is limited to distances shorter than R, as the RF boundary conditions, described in Appendix A of HayP95, minimize the finite-size effect). Together, these two effects call for using an analysis grid that is larger than the product grid by a buffer zone of width $B = \max(G, R)$ in both forward and backward directions (note that R is one of the tuning parameters for the RF and is prescribed to decrease at each analysis pass). The length of the RF analysis grid for each segment is thus $L \equiv D + 2B$. While deflection control sets an upper limit on D, the choice of D, and thus L, is determined by a combination of accuracy and timing requirements.

A.2.5. Input Data

For each segment, the input data to the RF consist of valid scanlines whose centers encompass the range covered by L. In addition, the range of input scanlines is extended on the down-track end to account for the forward range loss at the edge of swath *FRLE*. *FRLE* is the difference in forward view range between the center and edge of scan or, equivalently, the distance along track between the center of a scan line and the point that bisects the arc (along a great circle) between the two endpoints of the scan line. *FRLE* is equal to

$$FRLE = [\beta - \tan^{-1}(\cos\gamma \tan\beta)]r_e \tag{A9}$$

To be consistent with the RF formulation, among input scanlines, only scan positions falling within the limits of the analysis grid (in the transformed coordinates) are retained for the analysis.

A.2.6. Computation Sequence

The product grid for a single EDR file is defined as follows:

1. Determine the start-point for the complete product grid (scanline number N_s and position $\vec{P}_s$) and the number of N_e of the scanline that marks the end of the product grid. The application-specific details of this determination are described in Section A.3.
2. For the first segment, set $n_{1,s}$ and $\vec{p}_{1,s}$ to N_s and $\vec{P}_s$.
3. Determine the number of extra input scanlines necessary to cover the analysis grid based on B+*FRLE* and the nominal cross-scan FOV spacing.
4. Check whether the dataset contains enough scanlines to cover the nominal D by comparing the scanline number defined by $\vec{p}_{i,s} + D$ with N_e. If not enough scanlines are available, reduce D and set a flag to stop after this segment is done.
5. Determine $n_{1,e}$ and $\vec{p}_{1,e}$ based on D.
6. Apply cross-track offsets to determine pivot points $\vec{p}^{\,S}_{i,s}$ and $\vec{p}^{\,S}_{i,e}$ for the grid rotation.
7. Define the coordinate transformation $\Re$.
8. Define the analysis grid (product grid plus any required buffer zones on each side)

9. Gather observations from scanlines covering the analysis grid, and transform their locations from lat/lon to analysis grid coordinates.

10. Perform RF analysis over the analysis grid L and output the results over the product domain D (including latitudes and longitudes of the product grid rotated back to the geographical coordinate system). For the VR only, a pre-computed parallax shift is applied to the locations of observations at each vertical level.

11. Compute the starting point of the next segment such that $n_{i,e}$ and $\vec{p}_{i,e} + P$ are $n_{i+1,s}$ and $\vec{p}_{i+1,s}$ for the next segment. The magnitude of the down-track offset P is application-dependent (see Section A.3).

12. Repeat the analysis, going back to step 3.

A.2.7. Missing Scanlines

The segmenting procedure described above assumes that the geolocation data are available for all scanlines contained in the retrieval file being processed. In reality, some scanlines will be completely "missing," i.e., both the state vector and the geolocation data will have unphysical values. While the missing state vector data are accounted for by excluding them from the observation data set used as input to the recursive filter, invalid geolocation data require special attention in the segmenting procedure.

To account for invalid geolocation data, the following procedure has been adopted:

1. If, for the 1st segment, $\vec{p}_{1,s}$ associated with $n_{1,s}$ is invalid, then $n_{1,s}$ is incremented until a valid $\vec{p}_{1,s}$ is found.

2. If, for the 1st segment, $\vec{p}_{n_{1,e}}$ for the first guess $n_{1,e}$ is invalid, then $n_{1,e}$ is decremented until a valid $\vec{p}_{n_{1,e}}$ is found. The line $\vec{p}_{i,s} \rightarrow \vec{p}_{i,e}$ is then used to find the location $\vec{p}_{1,e}$ at a distance $\widetilde{D}$ equal to the nearest multiple integer of Δ downtrack from $\vec{p}_{n_{1,e}}$. The segment is skipped if $\widetilde{D}$ is less than a prescribed minimum segment length D_{min} and the segmenting is redefined using $\vec{p}_{n_{1,e}}$ as the starting point $\vec{p}_{1,s}$ for the first segment.

3. Once $\vec{p}_{i,e}$ is found for the ith segment, $\vec{p}_{i+1,s}$ is set to $\vec{p}_{i,e} + \Delta$ and the process of determining $\vec{p}_{i+1,e}$ described in point 2 is repeated.

A.3. Application-Specific Aspects

This section describes computational aspects of defining the RF grid that are specific to the two main applications, vertical remapping and FOV interpolation.

A.3.1. Vertical Remapping

For this application, the RF product grid is a finer-resolution version of the EDR reporting grid, and is driven by the following considerations:

- *Reporting requirements for the EDRs to be remapped, i.e.,*

1. Horizontal resolution Δ of no more than 15 km for AVMP. For AVTP, it is no more than 40 and 200 km for pressures greater than 20 mb and less than 20 mb, respectively. As described in Section 2.3.3, in order to achieve best accuracy, the resolution of the RF product grid Δ' is related to Δ via $\Delta' = \Delta/m$ (where m is an integer number), with the output EDR product at resolution Δ obtained by sub-sampling the Δ' grid.

2. A swath width of $W = 1700$ km for both AVMP and AVTP, except for upper-atmosphere AVTP, where $W = 1540$ km (the actual value of W adopted for the recursive filter is slightly larger than these nominal values so as to be a multiple integer of Δ).

- *Accuracy.*

  The accuracy is related both to the likely spatial variability of the EDRs to be remapped as well as to the numerical aspects of the RF approach.

- *Computational efficiency.*

  The deflection-related upper limit on D is determined by the need to keep the product grid within the swath of measurements, as much as possible, to avoid degradation along the grid edge due to extrapolation. The swath of measurements narrows with altitude due to the slant viewing geometry, and approaches the width of the required reporting grid for the highest levels of that grid. Extrapolation errors due to the deflection can be held to a small portion of the overall error budget by keeping the deflection small in relation to the reporting grid spacing or the expected scale of spatial variability in the EDR to be remapped. For example, in the case of AVMP, this scale is on the order of 5 km, which limits acceptable deflection to 5/2 = 2.5 km and, based on the results shown in Figure A-5 calls for $D \leq 1500$ km.

The determination of the start- and end-points (expressed as $\vec{P}_s$ and N_e, respectively) for the complete product grid is input-driven and performed as follows:

1. Convert B to scanline units, using nominal spacing of input scanlines. Round up.
2. Start-point $\vec{P}_s$ is the center-point of scanline number $N_s = 1 +$ (result from step 1).
3. Convert $B+FRLE$ to scanline units, using nominal spacing of input scanlines. Round up.
4. End limit is N_e = (total number of input scanlines) – (result from step 3).

A schematic representation of the product grid (in this application equivalent to the reporting grid) and the actual analysis grid for the vertical remapping of CMIS data is shown in Figure A-5. It should be noted that, by design, there is significant overlap between analysis grids (and the observation points) corresponding to neighboring processing segments (i.e., product grids), thus assuring a smooth seaming between segments. Note also that the actual swath width (about 1840 km for CMIS) is larger than W. This calls for the extension of the RF analysis grid by a few rows (i.e., in multiples of Δ) on both sides of the swath so as to encompass input data all the way to the edge of swath. This extension accommodates any residual cross-track asymmetry after the application of the lateral offset and mitigates RF boundary effects in the cross-track direction.

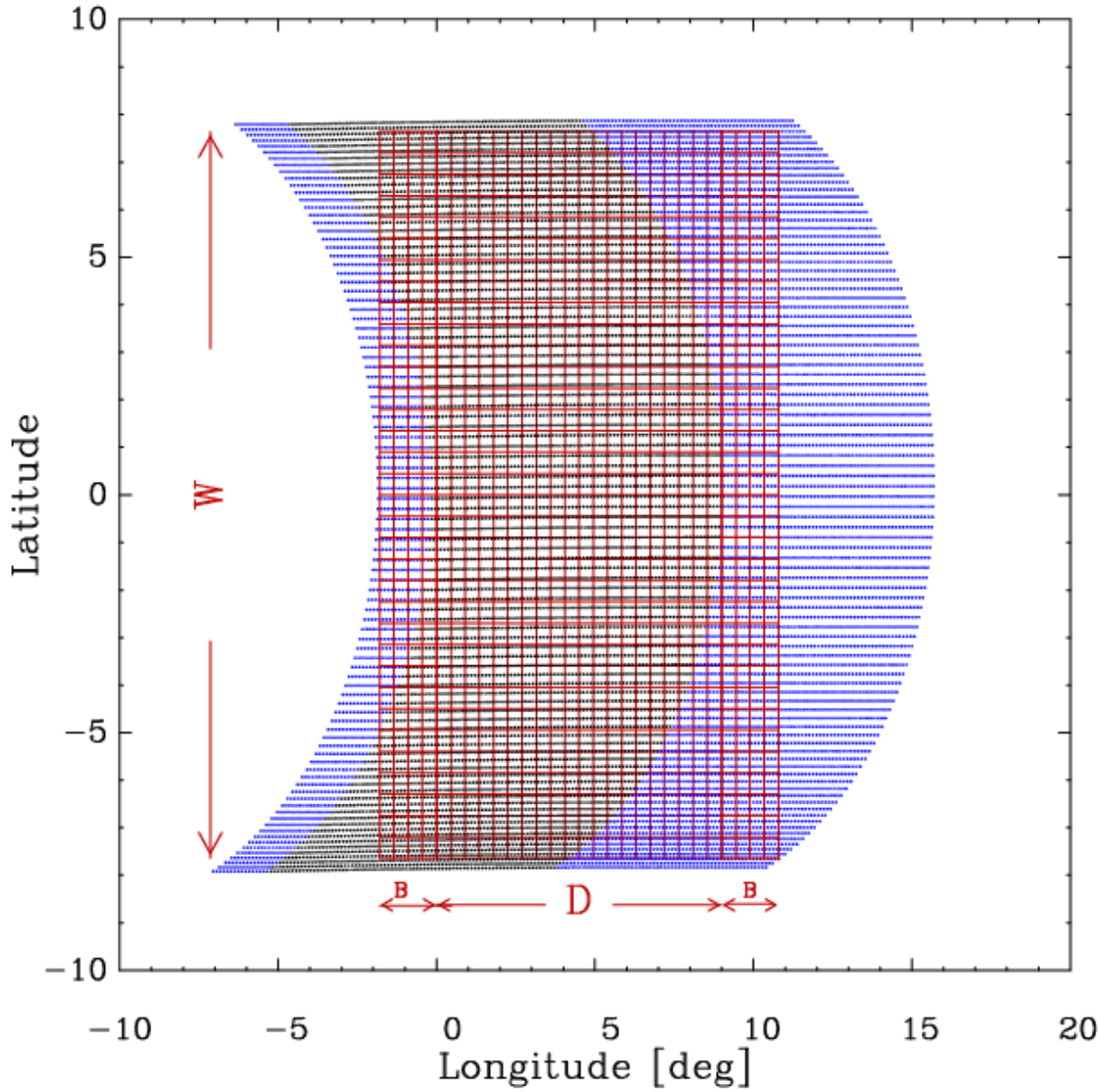

Figure A-5: Example of analysis grid and observational data points for vertical remapping of CMIS retrievals within each processing segment. The black points represent the scanlines used to define the product grid (length *D*), while the blue points represent the additional scanlines necessary to cover the analysis grid (length *L=D+2B*), accounting for *FRLE*.

The along-track offset for the determination of the starting point for the next segment is set to $P = \Delta$ in order to assure a quasi-continuity of the resulting grid and to avoid duplication of the analysis at the segment seams.

A.3.2. FOV Interpolation

In this application, the value of D is determined mainly by timing requirements (larger D allowing fewer segments). The deflection considerations are somewhat less important in this application since the output data covers essentially the same swath as the input data, and so deflections do not cause degradation of the output product quality.

Larger D are thus allowed, although very large D's and their corresponding deflections add to the computational burden by creating a RF product grid wide enough to encompass the deflected swath and thus containing large areas devoid of output FOVs. The width W is determined by the requirement that the product grid encompasses the output swath and is set to the actual swath width at the surface (computed from the sensor constants file), augmented on each side by the maximum (i.e., polar) deflection tabulated for a given D. The spacing of the product grid Δ is driven by accuracy requirements, accounting for the fact that in order to produce retrievals at the FOV output locations, an extra bilinear interpolation step follows the application of the recursive filter on the analysis grid. In order to assure a smooth seaming, for each processing segment the actual observation points used by the recursive filter are augmented by a few input scanlines beyond the scanlines used to define the product grid for the segment.

The determination of $\vec{P}_s$ and N_e defining the complete product grid is driven by the output FOV locations and performed as follows:

1. Among all scanlines, find the center-point $\vec{p}_o$ (and the corresponding scanline number) from the input (i.e., coarse FOV) data that is nearest to the lat/lon of the first scanline center-point of the output locations. The search starts with the first input scanline and discontinues when the distance stops decreasing and starts increasing. If the distance d to the nearest point is greater than f times the nominal spacing of input scan lines (where f is tunable, and we start with a value 1), there is a fatal error (occurs when the input swath starts too far downtrack from the start of the output swath).

2. Round up $FRLE$ and e (see definition below) to the nearest multiple integers of Δ.

3. Compute the start-point $\vec{P}_s$ as $\vec{p}_o - FRLE - d - e$ (measured along the uptrack direction as defined by the azimuth angle at $\vec{p}_o$). The end-point for the first segment is set to $\vec{p}_o$.

4. Among all scanlines, find the center-point (and the corresponding scanline number) from the input data that is nearest to the lat/lon of the last scanline center-point of the output locations. The search starts with the last input scanline and discontinues when the distance stops decreasing and starts increasing. If the distance to the nearest point is greater than f times the nominal spacing of input scan lines, there is a fatal error (occurs when the input swath ends too far uptrack from the end of the output swath).

5. The end-limit is N_e = (result of step 4) + 1 + e. If the result is greater than the number of input scan lines, reset the end-limit equal to the number of scanlines.

In the above, e is an extra amount (in multiples of Δ for $\vec{P}_s$ or a number for N_e) added to account for the fact that there is some approximation employed when we do computations using the nominal scanline spacing and assume a great-circle trajectory rather than using the actual scanline spacing and a trajectory that follows the arc of scanline centers.

The along-track offset P for the determination of the starting point for the next segment is set to a negative number (i.e., the start of the next segment is a small distance uptrack from the end of the preceding segment). This assures that during the bi-linear interpolation step (which is performed in rotated coordinates), no output FOV locations are missed between segments.

B. Errors Due to The Horizontal Interpolation of EDRs

B.1. Recursive Filter Methodology and Formulas

Regridding from the input FOVs to the output locations is performed using the recursive filter (see Section 2.2.1 for a discussion of the recursive filter). The resulting interpolation errors are estimated here by approximating its effects as those of a Gaussian filter. The two-dimensional recursive filter, when applied a number of iterations, approximately corresponds to a Gaussian filter with weights proportional to

$$G(x) = G_o \exp\left(\frac{-x^2}{R^2}\right) \tag{B1}$$

where R depends on the data density and the analysis pass. In the configuration recommended by HayP95, R at a given grid point R_i is obtained from:

$$R_i = \frac{f\Delta}{\sqrt{(W_i)}} \tag{B2}$$

subject to the constraint that:

$$R_m < R_i < R_0 \tag{B3}$$

where f is an adjustable parameter, Δ is the grid spacing, W_i is the local sum of data quality weights, R_0 is the specified maximum value of R, and R_m is a minimum value specified for each analysis pass.

The sum of weights is obtained by the formula:

$$W_i = \sum_k \left(1 - \frac{|d_k|}{\Delta}\right) W_k \tag{B4}$$

where d_k is the distance between observation k and grid point i, W_k is a data quality weight assigned to observation k, and the sum is accumulated over all k for which $|d_k| < \Delta$.

For the case of regridding from one set of approximately regularly spaced points to another, the following simplifications can be made:

$$Wk = 1 \text{ for all k} \tag{B5}$$

If we further choose the output grid no finer than the input grid, the sum of weights $W_i \geq 1$. This analysis neglects the effects of missing data in the input grid, because in that case larger smoothing radii are used locally. Using the value $f = 1$ suggested in HayP95 results in $R_i \leq \Delta$ for all grid points i, which is smaller than the limiting value R_m suggested in HayP95 for all but the last pass of the analysis. Therefore, the expression for R_i simplifies to $R_i = R_m$.

HayP95 suggest 5 analysis passes, with values of R_m as follows:

Table B-1: Default Values for Radius of Influence.

Pass	R_m/Δ
1	6.0
2	3.5
3	2.1
4	1.3
5	0.9

The passes are applied as successive corrections:

$$A_m = A_{m-1} + G(O - A_{m-1}) \tag{B6}$$

where O is the original field, A_0=0, and G() represents one pass of the recursive filter analysis.

B.2. Error Budget Calculations

B.2.1. Methodology

For the purpose of estimating the interpolation error introduced by the recursive filter, we simulate its effect as a series of five successive correction analysis passes given by (B6) with G() represented by the Gaussian filter (B1) with the scale parameter R_m/Δ for analysis pass m given in Table B-1. We note that this filtering is applied to the actual retrieved quantities, which are themselves approximately Gaussian averages, whereas the desired retrieved quantities are averages over square areas. The errors resulting from this mismatch have been separately estimated.

B.2.2. Computational Details

We approximate the Gaussian filter by the same truncated quasi-Gaussian filter (TQGF) used in the analysis of sensor averaging errors, with length parameter s = R/Δ.

The sensor averaging errors were evaluated using a very high-resolution input dataset (grid spacing 2.34 km and 2.5 km), and filtered values were computed on a subsampled output grid (every 5 grid points, or roughly 12 km grid spacing). The desired output grid and the input grid of retrieved values have a spacing of 12.5 km, which approximately corresponds to the subsampling of every 5 grid point of the high-resolution input dataset used in the sensor-averaging error analysis.

We use the sensor-averaged (z_g) and true (square-averaged, z_s) quantities on this 12 km grid, and then apply the recursive filter to the sensor-averaged gridded values (z_g) to obtain interpolated values (z_{rf}). For single values, the aggregate error E_a of the combined effects of sensor averaging and interpolation (filtering) is then obtained as the difference E_a=(z_{rf} - z_s), which can be broken down into the components due to sensor averaging (E_g) and interpolation (E_{rf}) as $E_a = E_{rf} + E_g$ with $E_{rf} = z_{rf} - z_g$ and $E_g = z_g - z_s$.

Note that if a breakdown of the error budget in terms of mean square errors (MSE) or error variances is required, possible correlation between E_{rf} and E_g must be taken into account.

Therefore, defining:

$$MSE_a = \left(\frac{1}{N}\right)\sum_{k=1}^{N} E_{a,k^2} \tag{B7}$$

we could define:

$$MSE_a = C_g + C_{rf} \tag{B8}$$

with contribution C_g and C_{rf} given by:

$$C_g = MSE_g \tag{B9}$$

and:

$$C_{rf} = MSE_a - MSE_g \tag{B10}$$

For completeness, we also evaluate the statistics of the differences between z_g and z_{rf} (which allows determination of the correlation between E_{rf} and E_g).

B.2.3. Test Data

The test data were taken from the innermost grid of very high resolution mesoscale model forecasts. The model data and advice on how to use these data were graciously provided by Gregory J. Tripoli and Giulia Panegrossi of the University of Wisconsin. This model is derived in part from the Colorado State University Regional Atmospheric Modeling System (RAMS) model as described by Tripoli (1992a, 1992b). There are three cases, hereafter denoted Bonnie, Genoa, and Friuli. The Genoa and Friuli cases are described by Tripoli et al. (2000). The Bonnie case is described by Panegrossi et al. (2001).

Details for these cases are as follows:

- ***Bonnie:***

The Hurricane Bonnie forecast was made from initial conditions at 0000 UTC 26 August 1998. Data are used at 36, 38, 40, 42 and 44 forecast hours, valid at times ranging from 1200 to 2200 UTC 27 August. (The initial conditions for this forecast are from a lower resolution run that started at 0000 UTC 25 August 1998.) The horizontal resolution of the data is 2.5 km on the 200×200 innermost grid, and the 39 vertical levels are:

0, 200, 400, 600, 800, 1000, 1220, 1462, 1728, 2021, 2343, 2697, 3087, 3516, 3987, 4506, 5077, 5705, 6395, 7145, 7895, 8645, 9395, 10145, 10895, 11645, 12395, 13145, 13895, 14645, 15395, 16145, 16895, 17645, 18395, 19145, 19895, 20645, 21395 m.

The corresponding level increments are 200 m through 1000 m, then increasing smoothly to 750 m from 7000 m onward.

- ***Genoa:***

The Genoa flood forecast was made from initial conditions at 1200 UTC 26 September 1996. Data are used at 14, 20, 26, 32, 36 forecast hours, valid at times ranging from 0200 UTC 27 September to 0000 UTC 28 September. The horizontal resolution of the data is 2.34 km on the 180×180 innermost grid, and the vertical levels are equal to the first 37 levels of the Hurricane Bonnie case.

- ***Friuli:***

The Friuli flood forecast was made from initial conditions at 0000 UTC 5 October 1996. Data are used at 42, 48, 54, 60 forecast hours, valid at times ranging from 1800 UTC 6 October to 1200 UTC 7 October. The horizontal resolution of the data is 2.34 km on the 200×160 innermost grid, and the vertical levels are equal to the first 37 levels of the Hurricane Bonnie case. (The innermost grid was added at hour 39, at 1500 UTC 6 October.)

B.2.4. Example Calculations

We illustrate the filtering effects of the recursive filter interpolation for the case of the high-resolution low-level mixing ratio for Hurricane Bonnie at forecast hour 48. The raw field is shown in Figure B-1, the corresponding true retrieved (SURF, z_s) values in Figure B-2, and the sensor values (TQGF, z_g) on the input grid in Figure B-3. The recursive filter is applied to this field, with the result (z_{rf}) shown in Figure B-4. Results for the individual passes (not shown) show a progressively higher resolution analysis. In our simulation study, we assumed a uniform data density commensurate with the output grid resolution, in which case the first 4 passes of the recursive filter analysis have little effect (there is little difference between the final analysis of the 5-pass recursive filter outlined above and a single pass of the recursive filter analysis with the smallest length scale R). However, all results shown here use the final results of the 5-pass recursive filter analysis. The basic statistics for this case are given below.

Table B-2: Statistics for Bonnie low-level mixing ratio (see text).

	N	bar	sd	rms	min	max
z_s	1521	6.35e-03	2.26e-03	6.74e-03	0.001372	0.011280
z_g	1225	6.80e-03	2.15e-03	7.13e-03	0.001539	0.011058
z_{rf}	1225	6.80e-03	2.14e-03	7.13e-03	0.001604	0.010784
zae	1225	1.06e-06	6.39e-05	6.39e-05	-0.000236	0.000320
zre	1225	-3.90e-04	9.84e-03	9.85e-03	-0.045777	0.044760
rsae	1225	1.08e-06	1.35e-04	1.35e-04	-0.000491	0.000638
rsre	1225	-7.00e-04	2.03e-02	2.03e-02	-0.084502	0.085815
rgae	1225	2.00e-08	7.43e-05	7.43e-05	-0.000260	0.000340
rgre	1225	-3.18e-04	1.12e-02	1.12e-02	-0.048782	0.042979

Here ae and re represent the absolute and relative error, respectively, for the following difference fields:

zae, zre: z_s - z_g (error introduced through sensor averaging)

rsae, rsre: z_s - z_{rf} (aggregate error of sensor averaging and\ interpolation)

rgae, rgre: z_g - z_{rf} (error introduced through interpolation)

In this case the interpolation step introduces an additional error of approximately the same magnitude as the error introduced by the sensor averaging. The aggregate error rms (rsae=1.4 e^{-4}) is larger than the sum of individual contributions $\sqrt{rgae^2 + zae^2}$ = 9.8 e-5, indicating a positive correlation between the two error components.

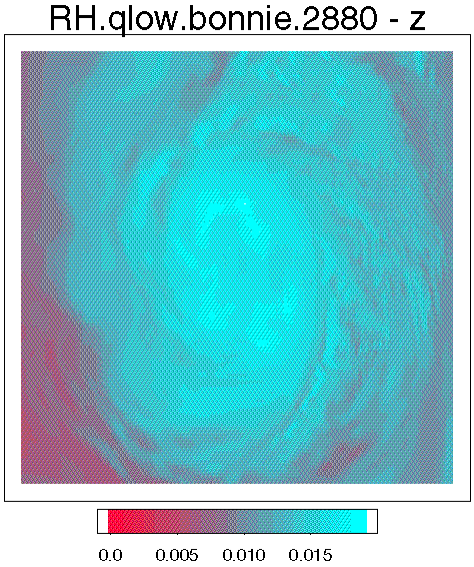

Figure B-1: The low-level mixing ratio field (kg/kg) for the Bonnie case at forecast hour 48.

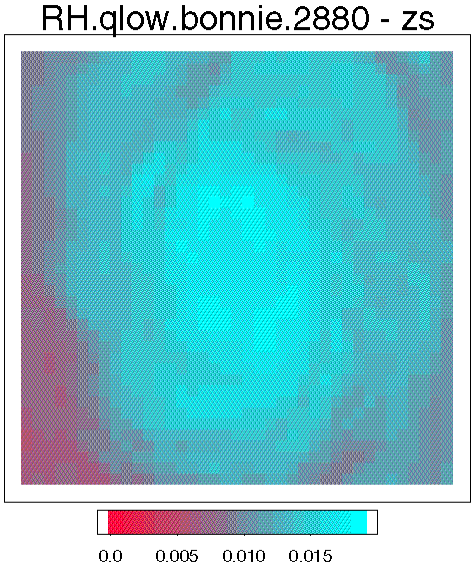

Figure B-2: The low-level mixing ratio field (kg/kg) for the Bonnie case at forecast hour 48, filtered with the SURF (true values, zs).

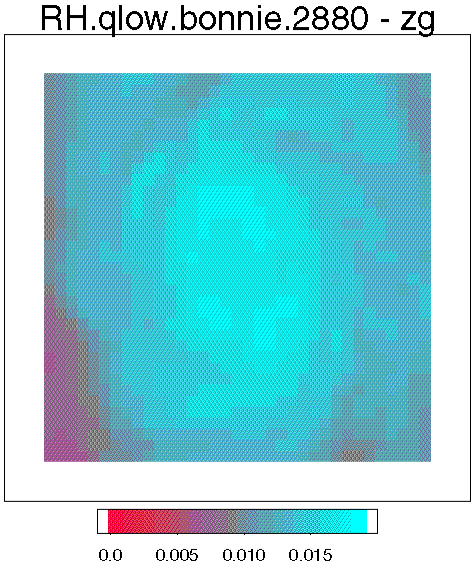

Figure B-3: The low-level mixing ratio field (kg/kg) for the Bonnie case at forecast hour 48, filtered with the TQGF (sensor values, zg).

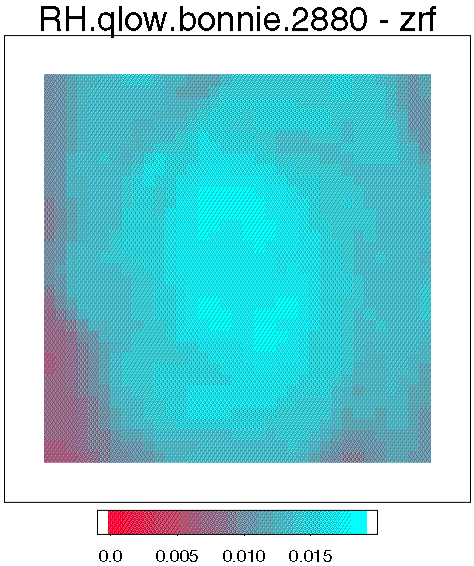

Figure B-4: The low-level mixing ratio field (kg/kg) for the Bonnie case at forecast hour 48, filtered with the TQGF and recursive filter (interpolated values, zrf).

B.2.5. Results for all Cases and Times

We extend the calculations shown in the previous section to all cases (Bonnie, Friuli, and Genoa) and times, for the layer-averaged mixing ratios (low, middle, and high), for the high-resolution (15 km) retrievals. The time-averaged error rms of the absolute error are shown side-by-side for the aggregate (rsae) and component (rgae and zae) errors in Table B-3. The corresponding time-averaged RMS statistics for the relative error are shown in Table B-4.

Table B-3: Time-averaged RMS absolution aggregate and component error, all cases.

	RMS AE			
Case	**Var**	**rsae**	**zae**	**rgae**
Friuli	r(mid)	3.31 e-05	1.71 e-05	1.70 e-05
	r(high)	8.94 e-06	4.67 e-06	4.55 e-06
	r(low)	5.44 e-05	2.78 e-05	2.84 e-05
Genoa	r(mid)	8.39 e-05	4.25 e-05	4.26 e-05
	r(high)	1.76 e-05	8.92 e-06	8.99 e-06
	r(low)	8.32 e-05	4.41 e-05	4.19 e-05
Bonnie	r(mid)	1.72 e-04	8.14 e-05	9.42 e-05
	r(high)	8.36 e-05	3.91 e-05	4.60 e-05
	r(low)	2.36 e-04	1.13 e-04	1.29 e-04

Table B-4: Time-averaged RMS relative aggregate and component error, all cases.

	100*RMS RE			
Case	**Var**	**rsae**	**zae**	**rgae**
Friuli	r(mid)	1.49	0.81	0.78
	r(high)	2.71	1.47	1.39
	r(low)	0.76	0.39	0.40
Genoa	r(mid)	2.48	1.29	1.28
	r(high)	4.21	2.33	2.22
	r(low)	0.94	0.51	0.48
Bonnie	r(mid)	2.54	1.23	1.42
	r(high)	4.93	2.47	2.93
	r(low)	1.83	0.89	1.01

Both tables show a pattern that is consistent with the results shown for the example calculation shown in the previous section: the interpolation errors as simulated here by the smoothing of the sensor-averaged fields are of the same order as those introduced by the sensor averaging pattern, leading to aggregate rms errors about twice as large as those resulting from the sensor averaging alone. Time-averaged rms aggregate errors are less than 5% for any of the cases for r(high), less than 3% for r(mid), and less than 2% for r(low).

B.2.6. Discussion/Conclusions

Regridding of retrieved values from a set of input FOVs to a set of output locations (at comparable spatial resolution) is performed using the recursive filter. The resulting interpolation errors are estimated here by approximating its effects as those of a Gaussian filter applied to the retrieved values.

We use the same datasets and methodology as what was used for the estimation of the errors introduced resulting from the mismatch of the (Gaussian) sensor and desired (boxcar) averaging patterns.

For the layer-averaged water vapor mixing ratios (r(mid), r(high), r(low)) retrieved at high resolution (15 km), the interpolation errors as simulated here by the smoothing of the sensor-averaged fields are of the same order as those introduced by the sensor averaging pattern, leading to aggregate rms errors about twice as large as those resulting from the sensor averaging alone. Time-averaged relative rms aggregate errors are less than 5% for any of the cases for r(high), less than 3% for r(mid), and less than 2% for r(low). These numbers are likely to be upper bounds (since they correspond to rather extreme meteorological situations). Moreover, as described in Section 2.3, the RMS errors in moisture are smaller when computed on simulated swath data based on the gridded data utilized in this appendix.