Clustering of Satellite Sounding Radiances to Enhance Mesoscale Meteorological Retrievals

DONALD W. HILLGER AND JAMES F. W. PURDOM

NOAA/NESDIS Regional and Mesoscale Meteorology Branch, Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, Colorado

2 August 1989 and 10 July 1990

ABSTRACT

Clustering is used to enhance mesoscale meteorological detail in retrievals produced from satellite sounding measurements. By placing sounding fields-of-view (FOVs) into groups of similar measurements, mesoscale details are reinforced, compared to arbitrary grouping of FOVs into a fixed block size. Clustering takes advantage of similarity among the measurements to avoid smearing gradient information. A case study is presented showing the advantage of clustering as applied to the satellite sounding problem.

1. Introduction

Satellite sounding measurements are known to contain high horizontal-resolution detail on atmospheric variability (Zehr and Green 1984). At the same time these high horizontal-resolution measurements need to be spatially averaged in order to increase their signal-to-noise. The normal procedure for spatial averaging is to arbitrarily group adjacent sounding fields-of-view (FOVs) into a fixed block size. This averaging, however, can smear gradient information in the sounding measurements, especially if the averaging is arbitrary and extends across gradients in the measurements. When the sounding FOVs are placed into groups of similar measurements, smearing of mesoscale gradient information is avoided. To accomplish this, a clustering technique is used to selectively group satellite sounding measurements. Patterns in the clusters of measurements reveal the extent and variability of air mass (temperature and water vapor) characteristics.

The noise levels in the various sounding channels are the size criteria for grouping the FOVs. Clusters of FOVs are, therefore, similar to within the noise levels of the measurements, and differences between the clusters represent significant changes above noise. Due to correlation between adjacent measurements, the clustered FOVs also have limited spatial continuity. This spatial continuity delineates and reinforces mesoscale features. The same process reduces noise by treating similar measurements in groups. Retrievals are then performed on the clustered FOVs, with only one retrieval necessary for each cluster of similar values.

Corresponding author address: Dr. Donald W. Hillger, NOAA/NESDIS, CIRA, Colorado State University, Fort Collins, CO 80523-0033.

2. Clustering of satellite sounding data

There are many variations on the theme of clustering as applied to satellite sounding data (Lipton et al. 1986; McMillin 1986). Clustering has been used mainly to classify synoptic-scale sounding measurements into preselected groups before doing retrievals, leading to an improvement in retrieved results (Ferraro 1986). Thompson et al. (1985) used pattern recognition and principal component analysis to predefine groups for satellite sounding data. Individual FOVs were then classified into the preset groups. Results showed a strong positive impact on physical-iterative retrievals of atmospheric variables. Uddstrom and Wark (1985) chose to classify satellite sounding data based on principal components, but using fixed class boundaries. Wark (1985) suggested that noise levels may be used to delineate the classes, as was tested in this study.

Clustering is the process of separating data points into groups of similar values. The terms classification and discrimination are used to describe the processes of putting data points into fixed or predefined groups. The authors above used classification, for example, to discriminate the type of retrieval based on predefined groups. The major difference with this work is that the clusters are not predefined. Rather, only the cluster size is predefined. The clustering process, as used here, does not anticipate the outcome of the retrievals. This is especially applicable to mesoscale changes that may be subtle. Therefore, the two characteristics that make this clustering technique unique are:

1) It is based on the *noise levels* of the measurements in each of the VAS channels that are clustered. This means that the profiles retrieved within a cluster vary only within the noise in the measurements, and dif-

ferences between clusters represent the minimum variation above noise.

2) It is applied to *mesoscale* VISSR Atmospheric Sounder (VAS) data in order to detect high-resolution airmass variations, by clustering single FOV (maximum resolution) measurements into groups of similar values. These two conditions make this technique useful for the mesoscale analysis of VAS data, in terms of the ability to detect high resolution variations above noise level.

3. Real data example

The case to which the clustering technique was applied is shown in Fig. 1. The satellite data cover southern Minnesota and much of Wisconsin and Iowa in a postfrontal situation on 7 September 1983. All FOVs are cloud free with some possible minor exceptions. A cloud-free situation was purposely chosen to avoid clusters based on cloud differences. Clustering was used to distinguish variations in clear VAS measurements and, therefore, in the temperature and water vapor structure in the area of concern. The extension to include cloudy FOVs is an obvious next step, which is being pursued.

The VAS FOVs are at a resolution of about 15 km in the east–west direction and about 23 km in the north–south direction. There are 19 lines of 45 elements each, or 855 FOVs (minus missing values), covering an area about 450 km by 650 km. The VAS FOVs are at much higher spatial resolution than the surrounding synoptic RAOB locations. This leaves large gaps for the satellite to measure subsynoptic thermal and moisture variations below the resolution of the RAOB network. The VAS data contain twelve spectral channels that are sensitive to temperature and water vapor variations throughout the troposphere. For this study, data from VAS channel 11 were missing, leaving eleven other VAS channels for analysis.

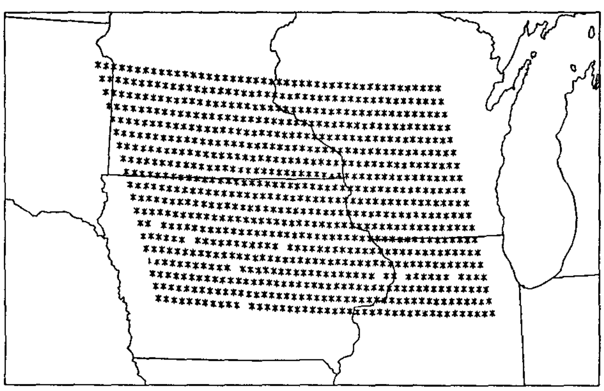

FIG. 1. Selected VAS FOVs at 15-km (east–west) resolution for 7 September 1983 at approximately 1400 UTC. Area is approximately 450 km by 650 km. Small gaps denote missing values in the grid of 19 lines by 45 elements.

4. The clustering process

Since VAS dwell-sounding data contain up to twelve spectral channels, there is a wide choice of variables (channels) that can be used to cluster the VAS FOVs. The clustering of VAS data is a computer-intensive process. It involves computing the "deviance" (Kendall 1975) between cluster members in measurement (effective blackbody temperature) space. The deviance is a multivariate measure similar to distance in Euclidean (x, y, z) space. But in this case the dimensions in cluster space are the various VAS channels.

The multivariate normalized deviance between the measurements in two FOVs in two or more channels (dimensions in cluster space) is

$$\text{dev} = (\Delta x/a)^2 + (\Delta y/b)^2 + \cdots \qquad (1)$$

where Δx and Δy are the differences in the measurements in the two channels, and a and b are the noise levels in the x and y channels, respectively. A constant deviance (e.g., dev = 1.0) can be visualized as an ellipse in two dimensions. A dev < 1.0 denotes that the measurements in two FOVs do not differ by more than the noise level in both the x and y channels, and a dev > 1.0 would occur if the measurements in the two FOVs differ by more than the noise level in one or more of the channels.

Two or more of the VAS channels are normally used for clustering. The difference in each channel is normalized by the noise level in that channel since the noise levels of the channels can vary widely. By normalizing the effective blackbody temperature differences by their noise levels, a deviance of 1.0 represents a signal equal to noise, and a value of 2.0 represents a signal twice that of noise, etc. The noise levels directly determine cluster size, with less noise resulting in smaller clusters. Noise levels were determined by structure function analysis of the VAS channels (Hillger and Vonder Haar 1988; Hillger et al. 1988).

The clustering process has the following steps (Hillger and Purdom 1988):

1) Pick a cluster "seed" by locating the most dense packing of FOVs in effective blackbody temperature space. This is done by computing the normalized deviance between all possible pairs of FOVs and determining the FOV coupling the largest number of other FOVs with dev ≤ 1.0.

2) Flag the other FOVs within the maximum allowable deviance (=1.0) as belonging to the cluster. FOVs with dev > 1.0 from the cluster seed remain unclustered.

3) Repeat steps 1 and 2 until no new cluster seeds can be found. New clusters seeds however have to be at dev ≥ 2.0 in order to make new clusters unique and nonoverlapping.

4) Force remaining FOVs into existing clusters (optional).

5. Principal component analysis

The process of clustering VAS measurements can be based on either all or a subset of the VAS channels. Results of clustering on VAS channels directly are not shown. Rather, to reduce the number of independent variables, a principal component analysis was applied to the VAS channels. The well-known technique of eigenvalue/eigenvector analysis was used to transform the VAS channels into VAS principal components (PCs), thereby packing the information content of the VAS channels into fewer components. The transformation from effective blackbody temperatures to principal components follows the formula

$$\mathbf{P} = \mathbf{E}\mathbf{B} \tag{2}$$

where **B** is an effective blackbody temperature vector being transformed by the eigenvector matrix **E** into principal component vector **P**. This represents a one-to-one transformation between **B** and **P** vectors. The eigenvectors of the covariance matrix of VAS channels were determined using standard math/statistics software.

The clustering technique is equally applicable to VAS effective blackbody temperatures or to VAS PCs. However, if PCs are used instead of VAS channels, the PC noise levels must be substituted for the VAS channel noise levels in order to determine the deviances in Eq. (1). The PC noise levels are available by structure function analysis on the VAS PCs, in a manner similar to determining the noise level of VAS effective blackbody temperatures (Hillger and Vonder Haar 1988), or by use of Eq. (2).

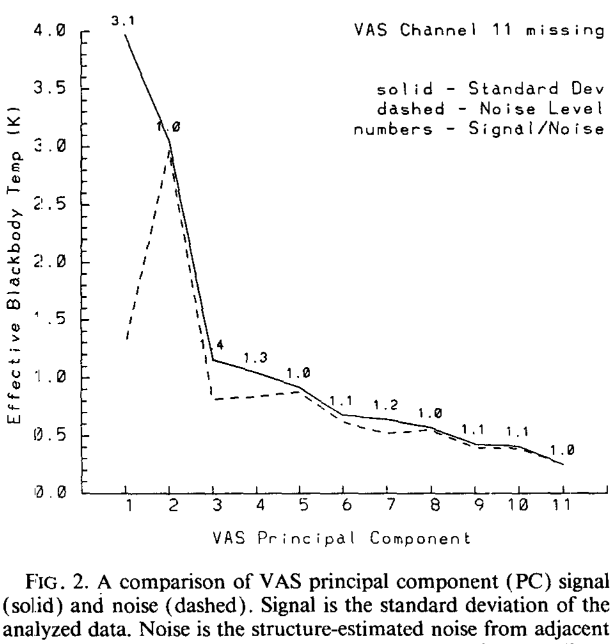

FIG. 2. A comparison of VAS principal component (PC) signal (solid) and noise (dashed). Signal is the standard deviation of the analyzed data. Noise is the structure-estimated noise from adjacent data pairs. Signal-to-noise (numbers) are the ratios of the two lines. Note the low signal-to-noise (high noise level) for PC 2, especially compared to higher-ordered PCs 3 and 4.

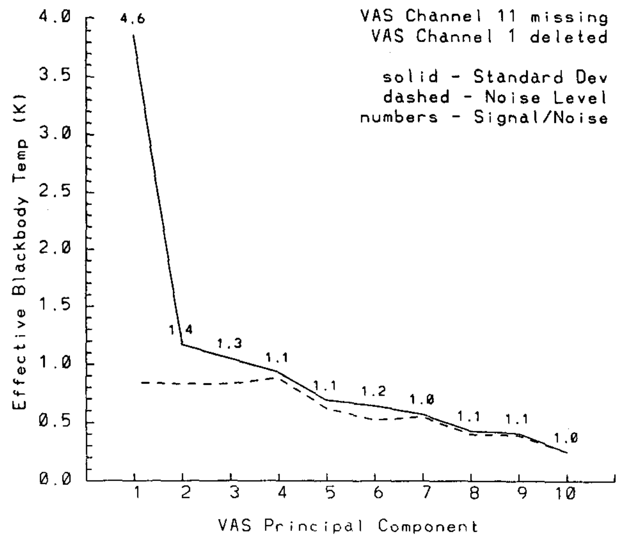

FIG. 3. Same as Fig. 2, but with VAS channel 1 removed from the principal component analysis, resulting in only ten PCs. Note the nearly steady decrease in signal-to-noise with increasing PC number.

The PCs are, by definition, ordered in terms of descending amount of explained variance, with each PC explaining more variance than following components. However, the same descending order was not found in the case of PC noise levels. The structure-estimated PC noise did decrease in general with increasing PC number, *except* for the second PC, as shown in Fig. 2. For the second PC the noise variance is nearly as large as the variance in that PC, indicating that the second PC is mainly composed of noise. The third and fourth PCs actually have larger signal-to-noise ratios than the second PC, indicating that they contain more information. Numbers to support this for VAS have not been seen in the published literature, but analyses of multispectral LANDSAT-type data have indicated that higher-order components may be more useful than lower-ordered components, especially when the noise levels of the original measurements (channels) vary widely (Green et al. 1988).

The noisy second PC, as indicated in Fig. 2, results from a large noise level and low signal-to-noise in VAS channel 1. This was determined by removing channel 1 from the principal component analysis, resulting in the PCs being ordered more appropriately according to decreasing signal-to-noise, as shown in Fig. 3. PCs beyond the sixth are almost completely noise dominated according to this analysis, indicating that the ten VAS channels (without channels 1 and 11) contain only six pieces of information. This redundancy in information content is a well known fact for remote sounding measurements. Figure 4, in a manner similar to Figs. 2 and 3, shows a signal-to-noise analysis of the

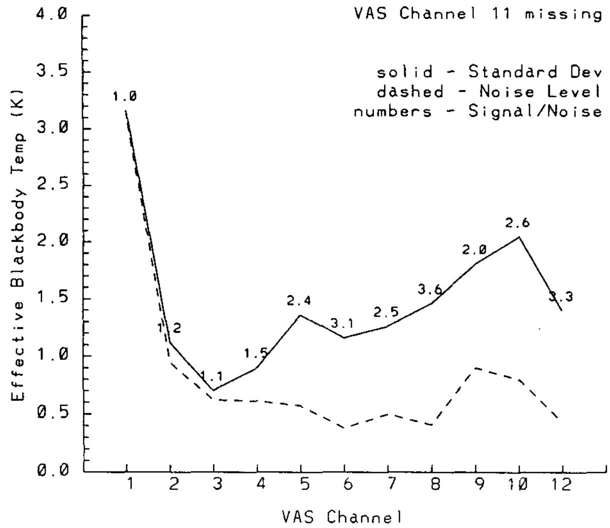

FIG. 4. Same as Figure 2, but for VAS channels, showing that the stratospheric and upper-tropospheric channels (1–3) contain little signal above noise.

VAS channels. Channel 1 has a large noise level, equal to the signal and therefore contains virtually no information in this case.

Because the PCs are now ordered in terms of information content, information can be added by clustering on increasing numbers of PCs. Clustering on one PC alone results in clusters (A–I) as shown in line-element (real) space in Fig. 5. Adding a second PC will result in more and smaller clusters. Since the first two PCs contain over 80% of the signal in the VAS channels, clustering on two PCs is similar to clustering based on nearly all of the information in the VAS channels. With two PCs, the clustering technique produced the clusters in Fig. 6 (a and b) as shown in PC space. In Fig. 6a the individual cluster members are designated by the letters A–P (16 clusters). The ellipses in Fig. 6b show the cluster extent, based on PC noise levels. The same clusters are shown in line-element (real) space in Fig. 7. A very slight amount of smoothing of adjacent FOV measurements was used to give the clusters additional spatial continuity. The cluster pattern shown in Fig. 7 indicates the predominance of east–west features in the field, with stronger (nonisotropic) gradients oriented in the north–south direction. Results for three PCs (not shown) are similar to those for two PCs with only a limited number of additional clusters.

6. Clustering versus blocking

To increase the signal-to-noise ratio of the VAS measurements, multiple FOVs need to be averaged together. The amount of spatial averaging varies, depending on a tradeoff between increased signal-to-noise and the desire for high resolution. The single FOV resolution of VAS large detectors is approximately 15 km. In mesoscale applications, a resolution of 20–75 km is typically needed (Jedlovec 1984, 1985). Operational VAS retrievals on the VAS Data Utilization Center (VDUC) averaged together 11 × 11 *small* (8 km) detector FOVs to achieve a resolution of about 75 km (Hayden 1988). That is about equivalent to 5 × 5 *large* (15 km) detector FOVs. Therefore, in this study retrievals using clustered FOVs are compared to retrievals using spatial blocking of 5 × 5 large FOVs.

By clustering on two PCs there were fewer (16) clusters than the same 855 FOVs blocked into 36 groups of 5 × 5 or 25 FOVs. Most clusters contain more FOVs than the arbitrary blocks. Thus clustering achieves economy in the retrieval, as well as improved signal-to-noise through averaging of similar measurements. Note further that all of the equal-sized blocks in Fig.

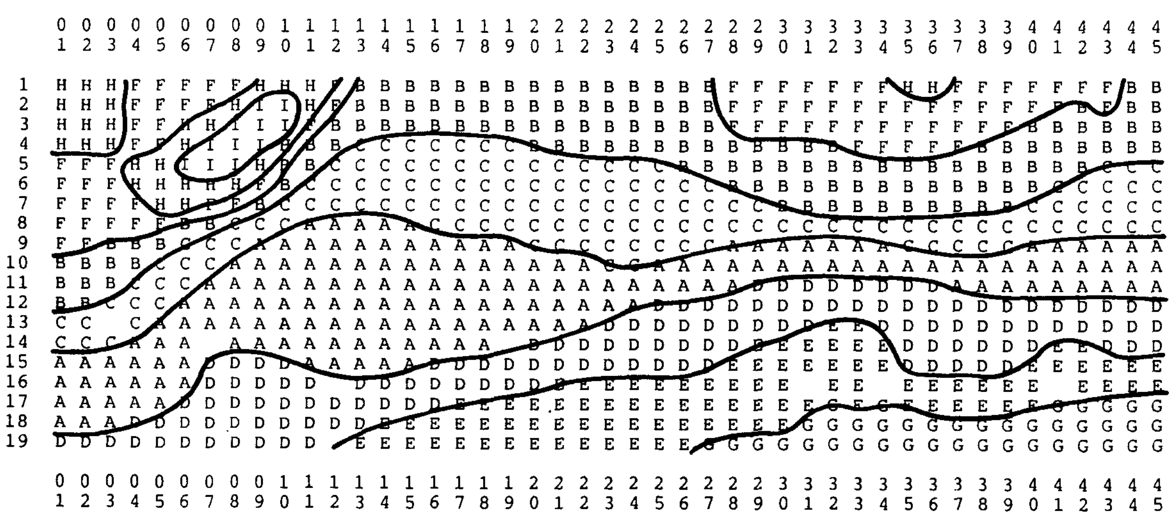

FIG. 5. Clusters of VAS FOVs in line-element (real) space based on variations in PC 1 only.

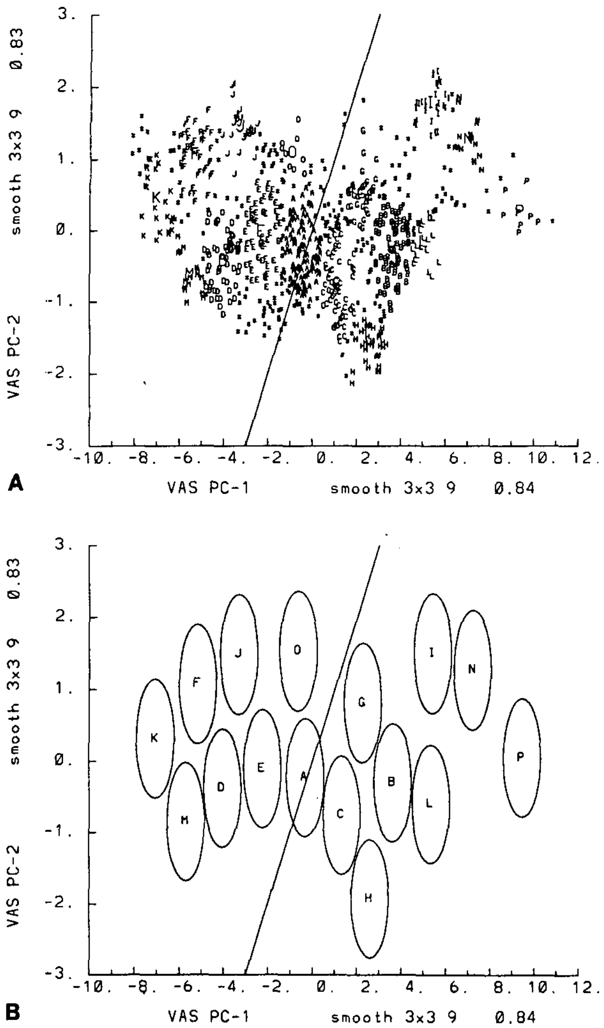

FIG. 6. Clusters of VAS FOVs in PC space based on variations in PCs 1 and 2. (a) Letters represent the clusters, with stars for unclustered FOVS. A slight spatial smoothing was applied to the data prior to clustering. (b) The ellipses show cluster limits determined by PC noise levels. The length of the semimajor and semiminor axes are given by the numerical values on each axis.

8, which are drawn over the cluster designations for two PCs, contain FOVs from more than one cluster. Thus, soundings retrieved from measurements in a given block should vary more than those retrieved from measurements in a given cluster. The variability in the clusters is limited to the noise level in the radiances, while the variability within blocks is not limited.

7. VAS retrievals

Retrievals are performed using the Smith et al. (1985) simultaneous temperature–moisture retrieval algorithm so that comparisons can be made between soundings retrieved on clustered and spatially blocked FOVs. In all cases the same first-guess sounding was used for the entire area, with the exception of the surface temperature being modified to better match the window channel measurements. Before soundings are retrieved, the PCs representing the clusters are changed back into effective blackbody temperatures using the inverse of Eq. 2.

$$\mathbf{B} = \mathbf{E}^{-1}\mathbf{P}. \tag{3}$$

Five retrievals have been produced for each cluster, one representing the cluster center, and one for the minimum and one for the maximum in each of the first two PCs. These four retrievals, which represent the four points at the ends of the semimajor and semiminor axes of the cluster ellipses, represent the variability within a given cluster. Examples of retrievals for cluster L as shown on the skew T–log P plot in Fig. 9 indicate the spread in retrievals allowed by variations within noise. In contrast, a much larger spread is evident in Fig. 10 for the retrievals from block 7 (counting from the upper left in Fig. 8), which has only about half of its area covered by FOVs from cluster L. Again five retrievals are shown, equivalent to those in Fig. 9. The larger spread in retrievals is due to the arbitrary spatial blocking allowing values from two clusters (B and L). As anticipated, arbitrary averaging results in more variability in the blocks than in the "more natural" clusters. Since some spatial averaging is required to increase the signal-to-noise ratio of VAS measurements, it makes sense to average the measurements in a manner that minimizes the variance of the combined values by clustering and maximizes the variances between clusters. How this affects horizontal fields from the retrievals is shown in section 8.

8. Horizontal cross sections

Horizontal cross sections were produced from retrievals on both the clustered and blocked VAS data. Objective analysis schemes were used to interpolate values at all FOVs. The weighting applied to each retrieved cluster value is based on the normalized-deviance (in cluster space) of the FOV from the retrieved values at the three (3) nearest cluster centers. Retrieved values at cluster centers with less deviance from the FOV are given more weight, retrieved values with greater deviance are given less weight. The weighting used to determine the temperature at any point in the field is

$$T = \sum_i^3 w_i T_i / \sum_i w_i. \tag{4}$$

Here T_i is the retrieved temperature for cluster i using the weight

$$w_i = \sum_j^3 \mathrm{dev}_j - \mathrm{dev}_i \tag{5}$$

where the deviances are subtracted from their sum in

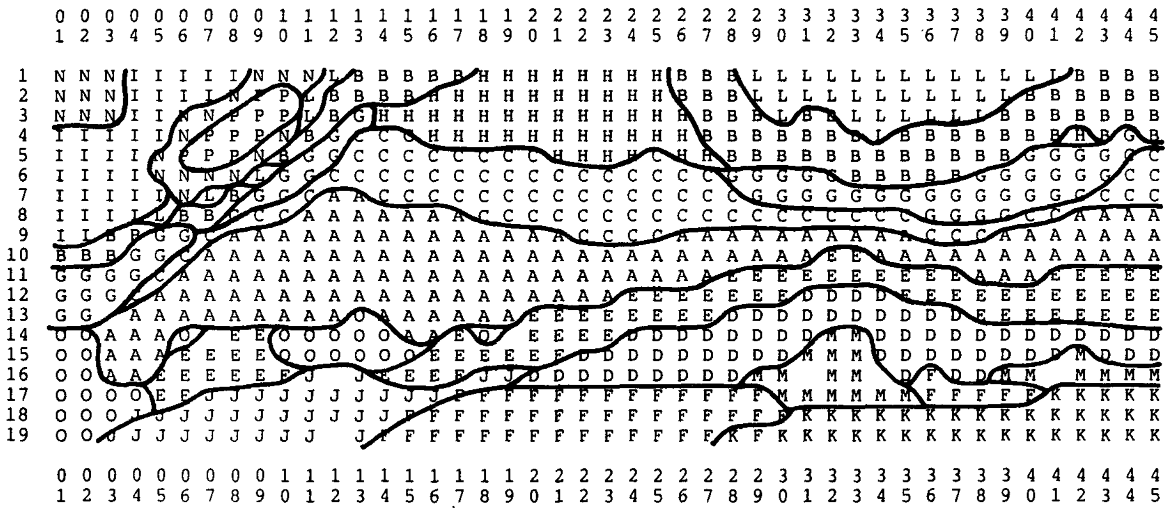

FIG. 7. The same clusters as in Fig. 6 based on PCs 1 and 2 in line-element (real) space. The addition of PC 2 results in more clusters than in Fig. 5.

order to more heavily weight closer values in cluster space. The number of cluster centers in the summation can be variable, depending on the range of influence wanted. In this case, three centers were used to influence each FOV. Note that the objective analysis for clustered retrievals does not consider their spatial separation, rather the deviance in cluster space (similarity in value), thus avoiding the smearing of gradients. The atmosphere is seldom homogeneous or isotropic enough to assume that adjacent measurements are the same within their noise level. Clustering tells us when or when not to consider adjacent measurements.

An example of a horizontal field based on clustered retrievals is shown in Fig. 11. The 700-hPa temperature field shows mesoscale features that might not be expected on such a small scale. The cooler temperatures in the northwest corner of the field are most likely due to subresolution clouds in the VAS data. Fractional-FOV clouds could lower the retrieved temperatures, since this area was treated as cloud-free-based window-channel measurements in VAS channel 8 (11 μm). Clustering, therefore, may be very useful for detecting and eliminating areas of subresolution cloudiness.

The same field produced from blocked retrievals is shown in Fig. 12. Unlike the clustered retrievals, the objective analysis between the retrieved values for the blocks considers only spatial separation rather than similarity in value. Two-dimensional (three-point) linear interpolation is used to generate the value at any FOV from the three nearest block centers in real space. Note the smoother horizontal features, lacking mesoscale detail, with gradients not as strong or localized

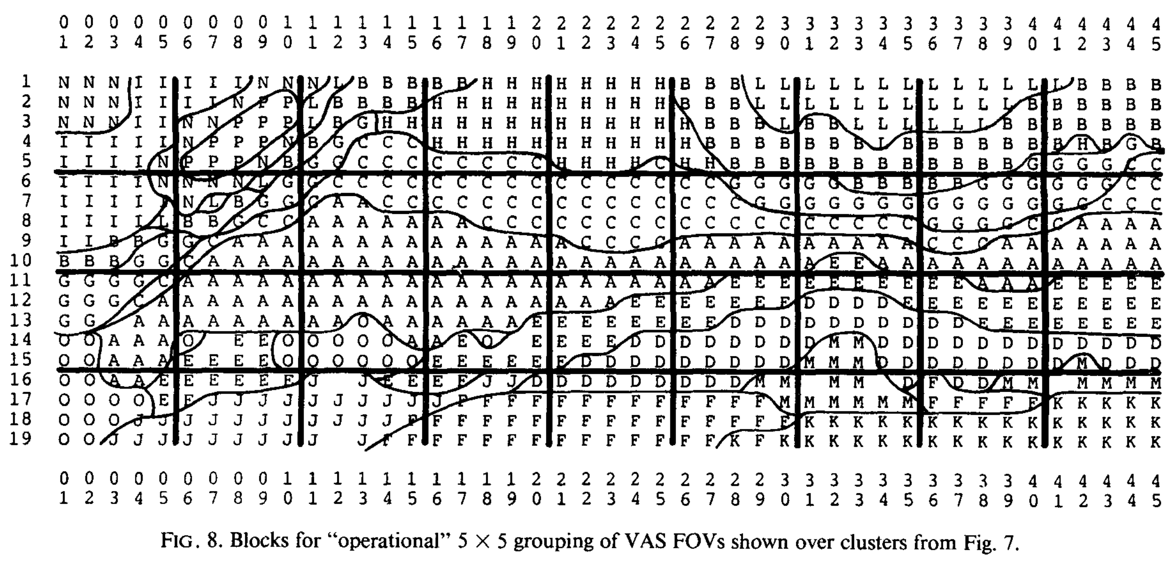

FIG. 8. Blocks for "operational" 5 × 5 grouping of VAS FOVs shown over clusters from Fig. 7.

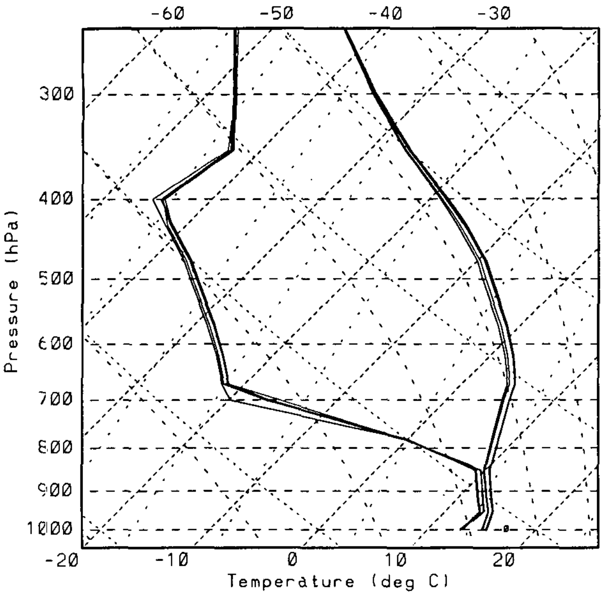

FIG. 9. Examples of soundings retrieved for cluster L. Shown are five retrievals; one for the cluster mean, and one for the minimum and one for the maximum in each of PCs 1 and 2.

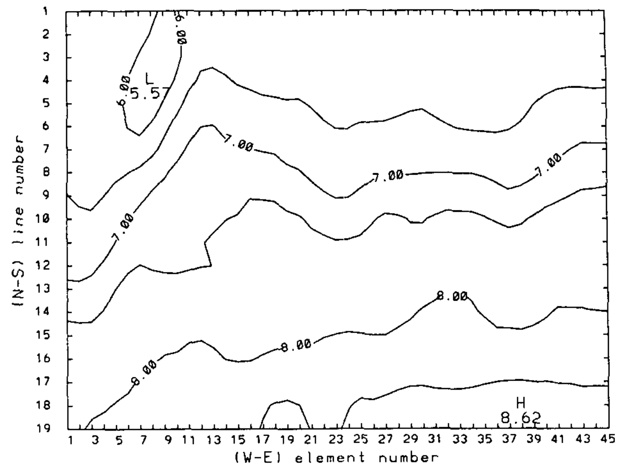

FIG. 11. Horizontal field of 700-hPa temperatures derived from soundings retrieved on clustered FOVs.

as those in the clustered retrivals. Some of the contour lines at small scales are at right angles between the two methods.

Are the small-scale features in Fig. 11 real? Since high-resolution conventional RAOB data are not available for verification of this case, this is a difficult question to answer. Mostek et al. (1980) similarly cited the lack of high-resolution RAOBs as a major limitation in verifying VAS soundings in mesoscale studies. However, previous work in which VAS retrievals for this case were compared with synoptic-scale RAOBs (Hillger and Purdom 1988) showed that the large-scale features were faithfully reproduced. Thus, it seems reasonable that smaller-scale features are similarly reproduced. This question is currently being investigated using VAS datasets coincident with some mesoscale conventional observations.

9. Summary and conclusions

A clustering technique is described and applied to satellite sounding measurements prior to retrieval of atmospheric profiles. Cluster size is based on the noise

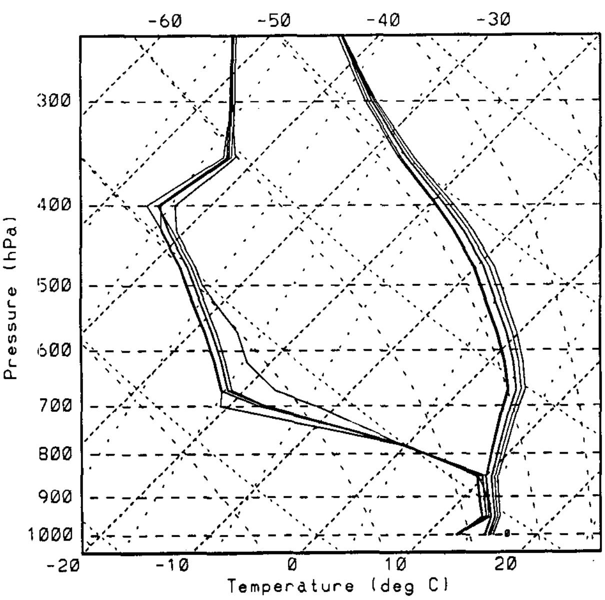

FIG. 10. Examples of soundings retrieved for block 7. Shown are five retrievals equivalent to those shown in Fig. 9. Note the larger variability in soundings, due to FOVs from two clusters (B and L).

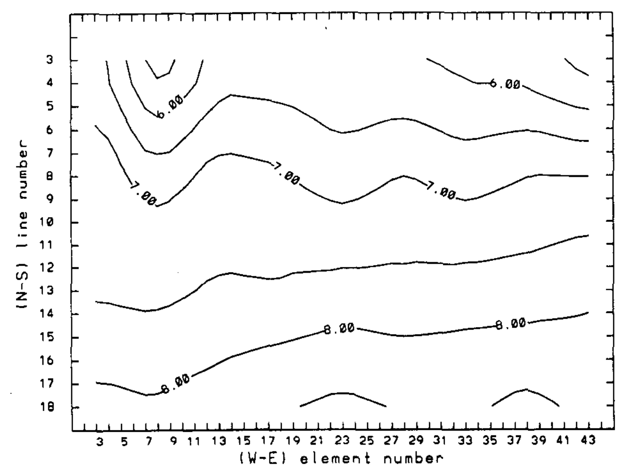

FIG. 12. Horizontal field of 700-hPa temperatures derived from soundings retrieved on blocked FOVs. Note the smoother features due to smearing of gradient information.

levels of the measurements, with variations between clusters then due to signal-above-noise. To see the effect of clustering, soundings within clusters show less variability than soundings within arbitrary spatial blocks with resolution similar to that used for operational VAS retrievals. On the other hand, horizontal fields produced from clustered retrievals show more mesoscale detail due to the natural grouping of the sounding FOVs. Mesoscale features are reinforced by grouping together similar FOVs, and gradient information is maintained. Although applied to clear FOVs, this technique also holds potential for cloud-clearing by detecting FOVs with subresolution clouds as clusters separate from other clusters of FOVs.

Acknowledgments. This work was supported by NOAA Contract NA85RAH05045 (5-31209 and in part by 5-31208). Thanks to Mrs. Debbie Eckdahl for typing the manuscript.

REFERENCES

Ferraro, R. R., Jr., N. C. Grody and J. A. Kogut, 1986: Classification of geophysical parameters using passive microwave satellite measurements. *IEEE Trans. Geosci. Remote Sens.*, **GE-24**, 1008–1013.

Green, A. A., M. Berman, P. Switzer and M. D. Craig, 1988: A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Trans. Geosci. Remote Sens.*, **26**, 65–74.

Hayden, C. M., 1988: GOES–VAS simultaneous temperature–moisture retrieval algorithm, *J. Appl. Meteor.*, **27**, 705–733.

Hilger, D. W. and J. F. W. Purdom, 1988: Clustering of VAS radiances to distinguish air mass characteristics. *International Radiation Symp.*, Lille, France, IAMAP, 523-527.

——, and T. H. Vonder Haar, 1988: Estimating noise levels of remotely-sensed measurements from satellites using spatial structure analysis. *J. Atmos. Oceanic. Technol.*, **5**, 206–214.

——, A. S. Jones, J. F. W. Purdom and T. H. Vonder Haar, 1988: Spatial and temporal variability of VAS radiance measurements by structure and correlation analysis. *Third Conf. on Satellite Meteorology and Oceanography* Anaheim, CA, Amer. Meteor. Soc., 51–54.

Jedlovec, G. J. 1984: A statistical evaluation and comparison of VISSR Atmospheric Sounder (VAS) data and corresponding rawinsonde measurements. NASA Technical Memorandum 82575, NASA/MSFC, 11 pp.

——, 1985: An evaluation and comparison of vertical profile data from the VISSR Atmospheric Sounder (VAS). *J. Atmos. Oceanic Technol.*, **2**, 559–581.

Kendall, M., 1975: *Multivariate Analysis.* Hafner Press, 31–46.

Lipton, A. E., D. W. Hilger and T. H. Vonder Haar, 1986: Water vapor vertical profile structures retrieved from satellite data via classification and discrimination, *Mon. Wea. Rev.*, **114**, 1103–1111.

Menzel, W. P., W. L. Smith, G. S. Wade, L. D. Herman and C. M. Hayden, 1983: Atmospheric soundings from a geostationary satellite. *Appl. Opt.*, **22**, 2686–2689.

McMillin, L. M., 1986: The use of classification procedures to improve the accuracy of satellite soundings of temperature and moisture. *Second Conf. on Satellite Meteorology/Remote Sensing and Applications,* Williamsburg, VA, Amer. Meteor. Soc., 1–4.

Mostek, A., L. W. Uccellini, R. A. Petersen and D. Chesters, 1986: Assessment of VAS soundings in the analysis of a preconvective environment. *Mon. Wea. Rev.*, **114**, 62–87.

Smith, W. L., H. M. Woolf and A. J. Schreiner, 1985: Simultaneous retrieval of surface atmospheric parameters: a physical and analytically direct approach. *Advances in Remote Sensing.* A. Deepak, H. E. Fleming and M. T. Chahine, Eds., A. Deepak Pub., 221–232.

Thompson, O. E., M. D. Goldberg and D. A. Dazlich, 1984: Pattern recognition in the satellite temperature retrieval problem. *J. Climate Appl. Meteor.*, **24**, 30–48.

Uddstrom, M. J., and D. Q. Wark, 1985: A classification scheme for satellite temperature retrivals. *J. Climate Appl. Meteor.*, **24**, 16–29.

Wark, D. Q., 1985: Temperature profiles derived from a restricted range of solutions using satellite measurements. *Advances in Remote Sensing.* A. Deepak, H. E. Fleming and M. T. Chahine Eds., A. Deepak, Pub., 473–478.

Zehr, R. M., and R. N. Green, 1984: Mesoscale applications of VAS imagery. *Conf. on Satellite/Remote Sensing and Applications,* Clearwater Beach, FL, Amer. Meteor. Soc., 94–98.