Automated identification of characteristic droplet size distributions in stratocumulus clouds utilizing a data clustering algorithm - Supplement

Nithin Allwayin, [a] , Michael L. Larsen , [a,b] Alexander G. Shaw, [c] and Raymond A. Shaw [a]

[a] *Michigan Technological University, Houghton, Michigan*

[b] *College of Charleston, Charleston, South Carolina*

[c] *Brigham Young University, Provo, Utah*

Corresponding author: Michael L. Larsen, LarsenML@cofc.edu

Corresponding author: Raymond A. Shaw, rashaw@mtu.edu

1. Selection of DBSCAN input parameters: Additional Information

DBSCAN input parameters are selected to maximize the number of clusters in each segment. As discussed in the main paper, the results are less sensitive to "epsilon" values and hence it is chosen to be a constant in our analysis. The "min points" values are iterated in steps of 5 for an "epsilon" value of 0.1. The lower cutoff value for "min points" is set as 10 in the paper. For more complete illustration, we plot the dependence of the number of identified clusters on the minimum number of points in a cluster in Figure 1. This demonstrates that even for smaller "min points", the algorithm identifies finite number of characteristic distributions.

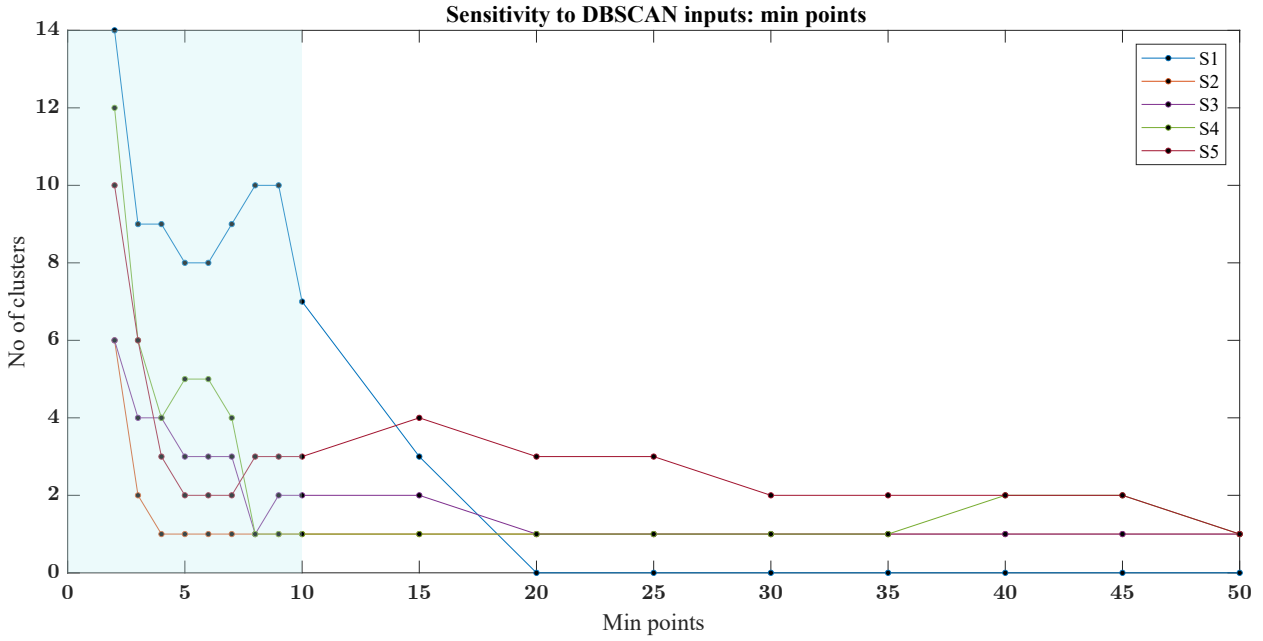

FIG. S 1. Sensitivity of the DBSCAN results to the input parameter: "min points". The number of clusters for the different cloud segments are shown. Here the value of "epsilon" is fixed to be 0.1.

2. Results for segment S3 & S5

Two clusters are identified for the segment S3. This segment is from the mid cloud region. The larger cluster has 715 members and the smaller one has 18. The results are shown in Figure S 2. Segment S5 is the transect near the cloud base. Four clusters with sizes 255,48,116 and 17 are identified for this segment and are shown in Figure S 3.

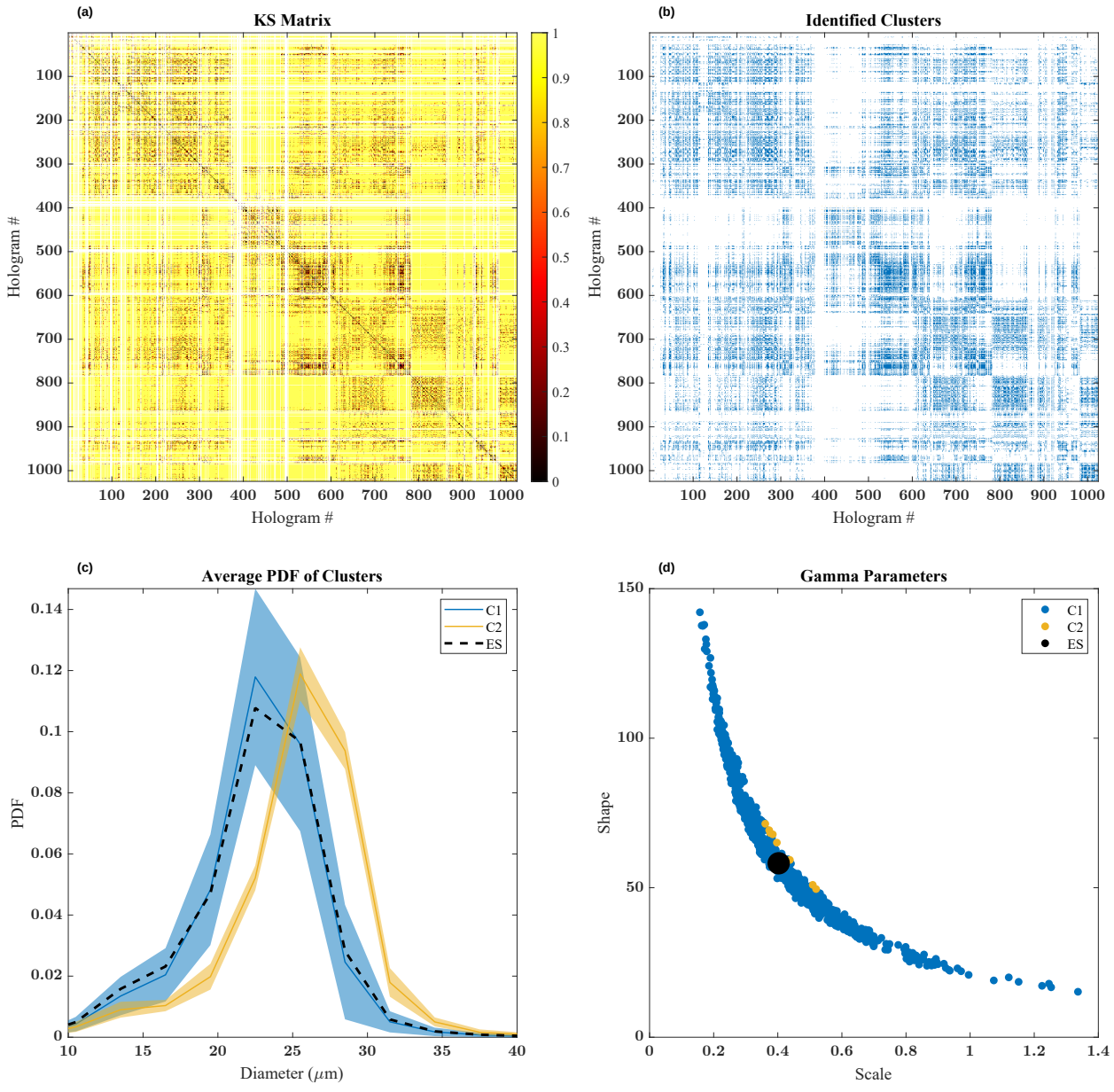

FIG. S 2. Results from Segment S3 (a) KS Matrix for the segment. (b) Clusters identified by the algorithm. The clusters are depicted by different colours. (c) Average PDFs of the different clusters. The shaded portion represents one standard deviation. The dashed black line shows the PDF for the entire segment of holograms above the cutoff. (d) The fitted shape and size parameters of the modified gamma distribution for different clusters. The large black dot gives the shape and size parameter for the entire segment.

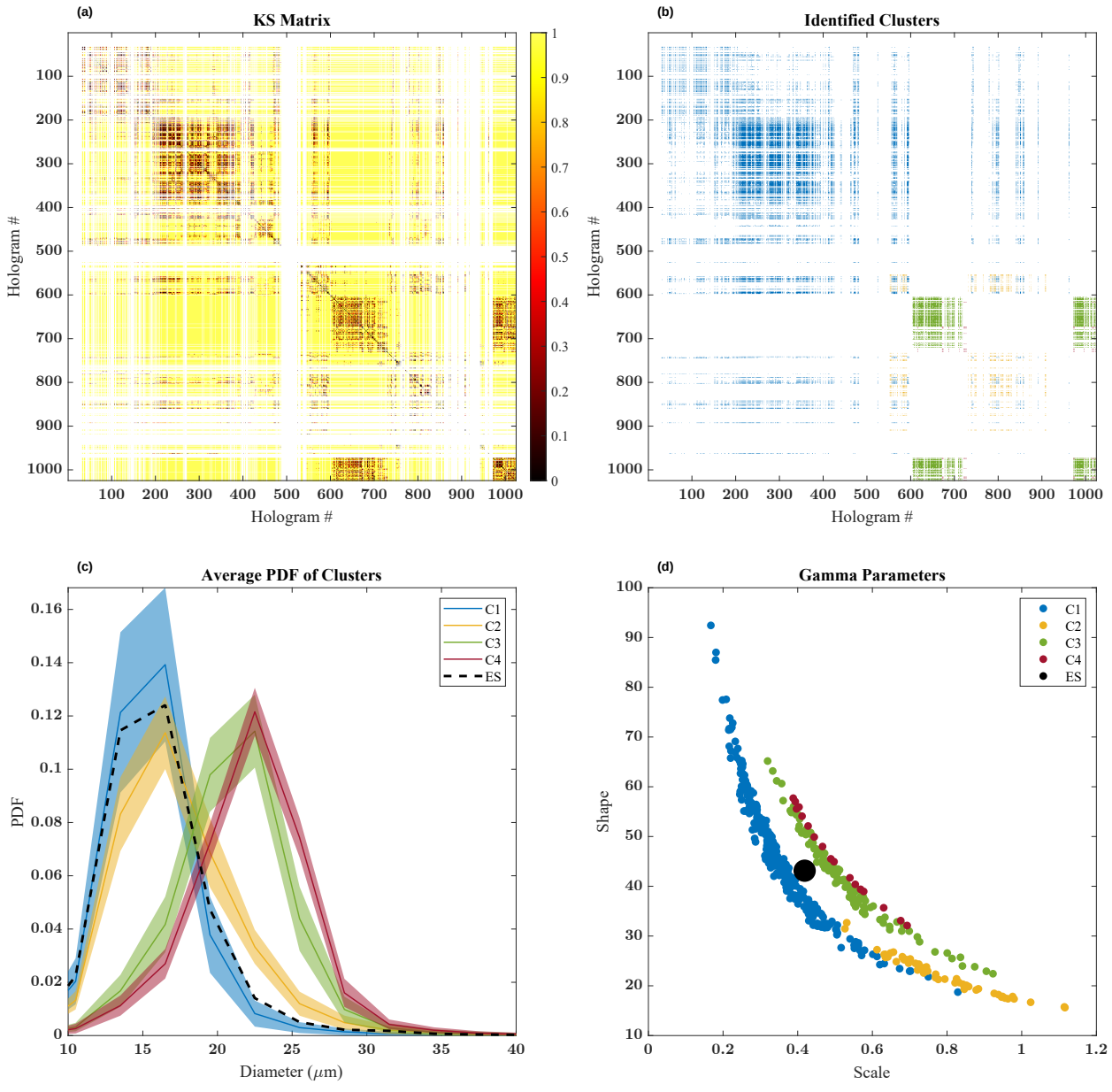

FIG. S 3. Results from Segment S5 (a) KS Matrix for the segment. (b) Clusters identified by the algorithm. The clusters are depicted by different colours. (c) Average PDFs of the different clusters. The shaded portion represents one standard deviation. The dashed black line shows the PDF for the entire segment of holograms above the cutoff. (d) The fitted shape and size parameters of the modified gamma distribution for different clusters. The large black dot gives the shape and size parameter for the entire segment.

3. Results of synthetic holograms-Segment SD3

The results for the set of synthetic holograms labelled SD3 is illustrated in Figure S 4. This set is one with the largest noise content among the three synthetic data sets.

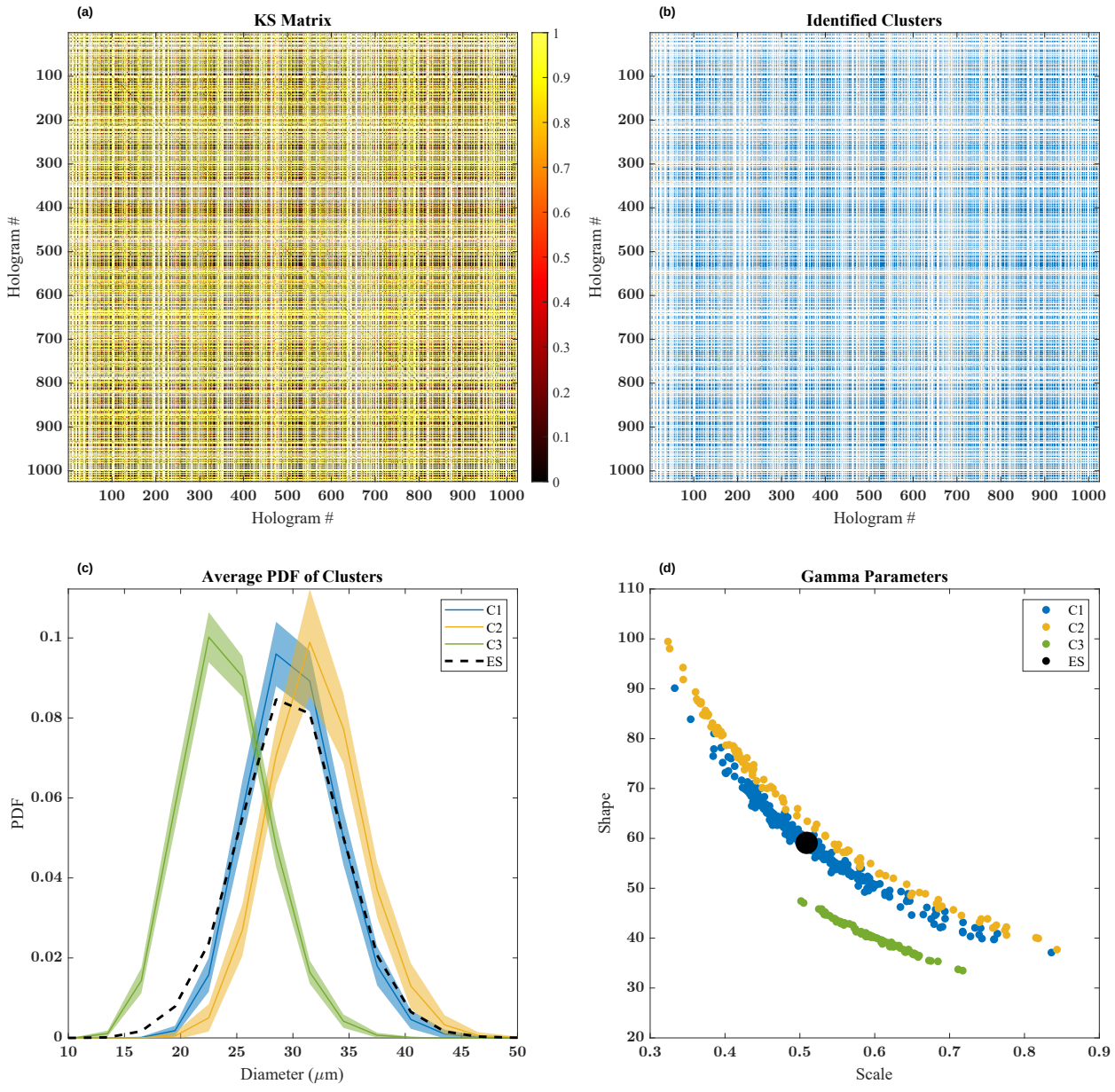

FIG. S 4. Results for the synthetic hologram set: SD3 (a) KS Matrix for the segment. (b) Clusters identified by the algorithm. The clusters are depicted by different colours. (c) Average PDFs of the different clusters. The shaded portion represents one standard deviation. The dashed black line shows the PDF for the entire segment of holograms above the cutoff. (d) The fitted shape and size parameters of the modified gamma distribution for different clusters. The large black dot gives the shape and size parameter for the entire segment.