

BIC Scores calculated for clusters received for different pairs of $\Delta Dist$ and ΔVar values

While traversing down the dendrogram, it is quite evident that the ward distance of the subtrees or sub-dendrograms decreases, but the variance of the subtrees do not decrease monotonically, as shown in fig-1. To comprehend the quality of the clusters received via the Distance - Variance Selection method, we applied BIC scores for a set of $\Delta Dist$ and ΔVar values. In fig-2 we see that

1. for very low values of $\Delta Dist$ and ΔVar , we received very high BIC scores, which does not indicate a suitable clustering model, and
2. for very high values of $\Delta Dist$ and ΔVar values, we detect very few clusters or carbon uptake regimes that are not expected according to domain science.

Hence, to acquire lower BIC and the expected number of clusters, we set the range of $\Delta Dist$ to [20, 40] and ΔVar to [0.1, 1.0].

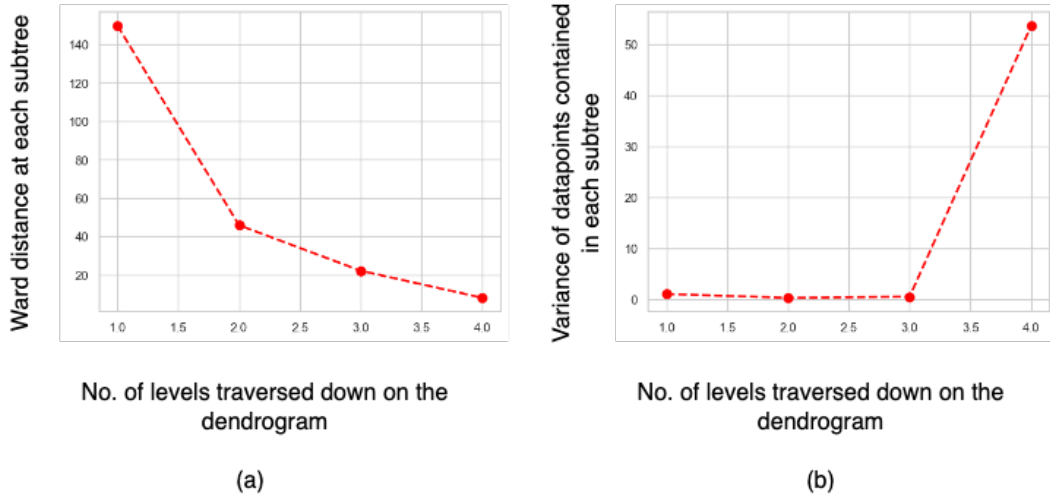


Figure 1: This figure shows how Ward distance (a) and total variance of all points contained (b) changes in the leftmost branch of the dendrogram of January 2009 (shown in fig-5 of main paper) at each subtree while traversing the dendrogram from top to bottom.

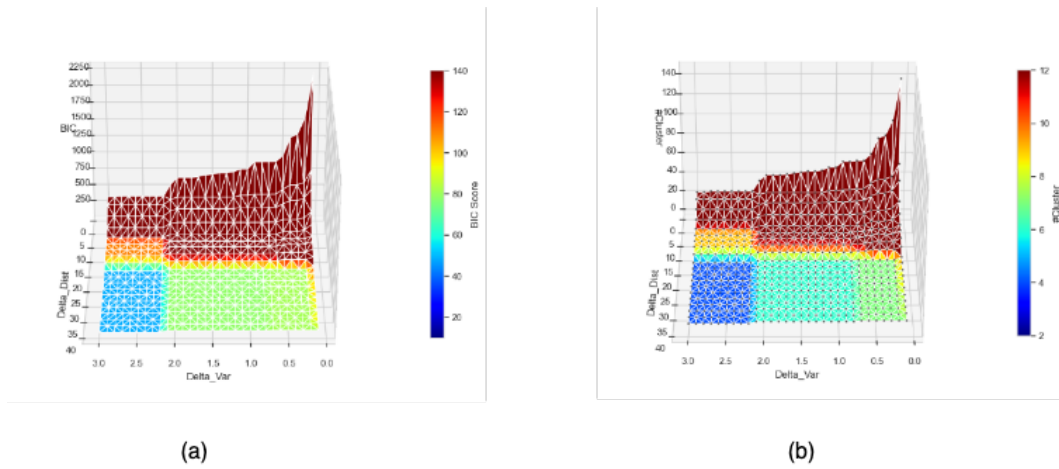


Figure 2: (a) BIC score wrt different sets of thresholds for change in distance (0-40) and variance(0.0 - 3.0) for the month of January in 2009, (b) number of clusters generated for change in distance (0-40) and variance(0.0 - 3.0) for the month of January in 2009