Thank you for your thoughtful comments and concerns about the methodology, in particular use of MDS + cluster analysis +Random Forest. In response to your concerns we added details/limitations of the cluster approach and an Appendix on the cluster solutions.

Once we created the dissimilarity matrix, we used MDS to generate a two-dimensional ordination showing landscape dissimilarity with the *MASS* package in R (Venables and Ripley, 2002). The MDS makes it possible to evaluate dissimilarity in two dimensions, which is essential to our goal to evaluate representativeness. We used the Kruskal method of non-metric scaling with the IsoMDS function in the *MASS* package (Venables and Ripley, 2002). IsoMDS works best when applied to metric variables (Torgerson, 1958). Torgerson (1958) initially developed this method, which assumes that the data obey distance axioms. It uses eigendecomposition of the dissimilarity to identify major components and axes, and represents any point as a linear combination of dimensions. This is very similar to principal component analysis (PCA) or factor analysis, but it uses the dissimilarity matrix rather than a correlation matrix as input. Furthermore, the included dimensions are the most important dimensions produced, like PCA which is able to identify all of the dimensions that exist in the original data up to N−1, but will retain only the most important ones.

Knowing that regional patterns in climate and land cover will be important for scaling CH_4 to the regional and national scale, we divided the US into clusters to evaluate representativeness using the first and second dimension from the MDS. Cluster analysis is an objective method of producing meaningful, mutually exclusive groups based on similarities among entities (Balijepally et al. 2011). This approach is descriptive, a-theoretical, and non-inferential with sound mathematical support (Balijepally et al. 2011). Clustering outcomes are driven by large effect sizes or the accumulation of many smaller effects across features, and are mostly unaffected by differences in covariance structure (Dalmaijer et al. 2020). Sufficient statistical power is achieved with relatively small samples (Dalmaijer et al. 2020), provided cluster separation is sufficient. Traditional notions about statistical power only partially apply to cluster analysis (Dalmaijer et al. 2020). Increasing the number of sample points above a sufficient sample size does not improve power, but effect size is important (Dalmaijer et al. 2020). Clustering is useful when large subgroup separation is expected, and when MDS improves cluster separation (Dalmaijer et al. 2020).

We used cluster analysis to summarize our results within a geographical context, an approach that has been used to delineate spatial sampling domains, to assess the spatial representativeness of network infrastructure, and to suggest arrangements of study sites (Sulkava et al., 2011; Kumar et al., 2016). We determined the optimal number of clusters using the library *cluster* and the function pam in R (Reynolds et al., 2006; Schubert and Rousseeuw, 2019, 2021). This approach uses the k-medoids algorithm, which partitions a data set into k groups or clusters and is a robust alternative to k-means clustering (Kaufman and Rousseeuw, 2009). The k-medoid algorithm is less sensitive to noise and outliers, compared to k-means, because it uses medoids as cluster centers. The k-medoids algorithm requires the user to specify k, the number of clusters to be generated. A useful approach to determine the optimal number of clusters is the **silhouette** method. We fit an increasing number of clusters from 2 to 50 to construct a silhouette plot and choose the number of clusters that maximized the average silhouette width (Figure S2).

While useful, there are limitations to cluster analysis that can affect cluster patterns and the stability of clusters. The cluster solution is dependent upon the clustering variables, the similarity/dissimilarity measure used, the clustering algorithm, and the data used to estimate clusters. Therefore, varying elements of clustering methods can lead to many alternative cluster solutions (Balijepally et al. 2011). Cluster solutions can be produced in the absence of natural structure in the data, and there is no statistical basis to reject the null hypothesis that there are no natural groupings in the data (Balijepally et al. 2011). Cluster algorithms also cannot differentiate between relevant versus irrelevant variables. Therefore, only the variables expected to be influential should be used (Balijepally et al. 2011). Variables should emanate from past research or explicit theory, and be consistent with the objectives of the study.

Due to the limitations of this approach, it is important to validate the cluster solution to ensure its meaningfulness and utility (Punj and Stewart 1983; Balijepally et al. 2011). Consistency is established by checking the stability of cluster solutions obtained by using multiple algorithms (Punj and Stewart 1983) or through splitting a sample, analyzing the cluster solutions for the two halves separately, and checking their consistency. After checking for reliability, the validity of a cluster solution is established through external validity and criterion-related validity. External validity ensures that clusters are representative of the actual population (Cook & Campbell, 1979) and can be verified by clustering on a hold-out sample using the same variables and assessing the similarity of the two solutions. This analysis was repeated 5 times to ensure that the 20,000 pixel subsample would produce similar results in the dimensions and clustering. For simplicity, we show the results of the first analysis, and a comparison of clustering methods and measures of stability are available in Appendix 2.

Once defined, each cluster was represented by one of the data points in the cluster named the cluster medoid. The medioid was the lowest average dissimilarity between it and all other objects in the cluster. The medoid can be considered a representative example of the members of that cluster. We calculated the dissimilarity between every location within the cluster to the medoid to create a measure of how different each location was from the medoid condition of each cluster. We utilized the pointDistance function in the *raster* package, which provided a unit-less relative measure of dissimilarity that was determined by measuring the difference between the first and second dimensions produced by the isoMDS of each point in a cluster to the dimensions of the medoid.

To extrapolate the cluster and dissimilarity layers across the entire US beyond the 20,000-pixel subsample and to show the predictive validity (Kerlinger, 1986) of the cluster solution we employed the machine learning algorithm Random Forest (RF) with the package *randomForest* (Liaw and Wiener, 2002) to model the first and second dimensions using the land cover and climate layers as predictors. We then created a Random Forest model of the cluster layer using the first and second dimension as the explanatory variables. All models were then projected spatially to produce a spatially explicit cluster layer and a dissimilarity layer beyond the 20,000 sample points that were used in the MDS analysis. The RF algorithm was first introduced by (Breiman, 2001), and uses an ensemble of regression trees to predict target values. In RF, a series of bootstrapped datasets are used to generate independent regression trees; at each node, a random sample of predictor variables is selected for use. The RF prediction is the ensemble of multiple individual trees. We created 500 trees for each year and site, using 80% of the data for

model fitting and 20% for model validation. The fit of each RF was evaluated with the out-of-bag mean square error (OOB MSE), and variable importance was computed as the amount of the prediction error increased when a particular predictor was permuted. Initially, 500 RF trees were generated. Overall model fit was evaluated with the average of the 500 OOB MSEs from the final model for each year and site, and variable importance was calculated as the average rank of each predictor variable for the 500 models. This approach allowed us to measure the importance of the original data on the first and second dimensions defined by the MDS and how the MDS leads to cluster and dissimilarity patterns. This step was essential to producing a spatially explicit cluster and dissimilarity layers for the entire US, since the MDS analysis limits the number of observations that can be analyzed. This is also important for evaluating the meaningfulness of the cluster by using the original variables used in the development of the distance matrix to predict clusters.

Supplement:

We evaluated the stability of the cluster solution presented (Figure 2) by comparing it to the cluster solutions of 6 example cases (Table S1). The example cases (Samples 1-5) use the same analysis approach as the presented solution (Figure 2). While Sample 1 is from a systematically random sample, samples 2-5 are from a randomly sampled dataset. To understand the impacts of subsampling, we derived a cluster solution for each sample and compared it to the standard solution (Figure 2). The package *bigmds* uses the divide-and-conquer MDS approach (Delicado and Pachon-Garcia 2020). This algorithm partitions the data into subsamples (n=60,000), where classical methods can work. In order to align all the solutions, the Procrustes formula is used (Borg and Groenen 2005). Following the MDS we used the kmeans function to cluster the bigmds (Forgy 1965; Hartigan and Wong 1979; Lloyd 1982).

Table S1: Example cases to test the stability of the cluster solution.

Example	**Description**	**Sample Size**	**R Packages**
Sample 1	Systematic random sample across all climate and land cover classes.	20,000	*kmed. MASS, PAM*
Sample 2	Random Sample	20,000	*kmed. MASS, PAM*
Sample 2	Random Sample	20,000	*kmed. MASS, PAM*
Sample 3	Random Sample	20,000	*kmed. MASS, PAM*
Sample 4	Random Sample	20,000	*kmed. MASS, PAM*
Sample 5	Random Sample	20,000	*kmed. MASS, PAM*

BigMDS	divide_conquer_mds	8,268,498	*bigmds, stats*

To compare cluster solutions we measured the stability of a cluster (Figure S3 and Figure S4a.). Stability was measured as the consistency of clustering within the presented solution/standard (Figure 2). This approach is not influenced by the cluster label changing, but by the presence of multiple clusters occurring within the standard cluster. A stability of 100% indicates that all samples within the standard cluster belong to the same cluster in the sample solution.

Regardless of the sampling, similar clustering patterns were obtained (Figure S4) and the mean stability was 82% across the six examples explored (Appendix. 2). The mean stability of the bigmds example was 65%. The cluster stability was low between neighboring clusters NEb and NEa, and between Eb and the SE cluster. Altogether these results show that failing to subsample the data properly leads to declines in cluster stability and changes in the clustering method can lead to differences in the final cluster solution (Figure S4).

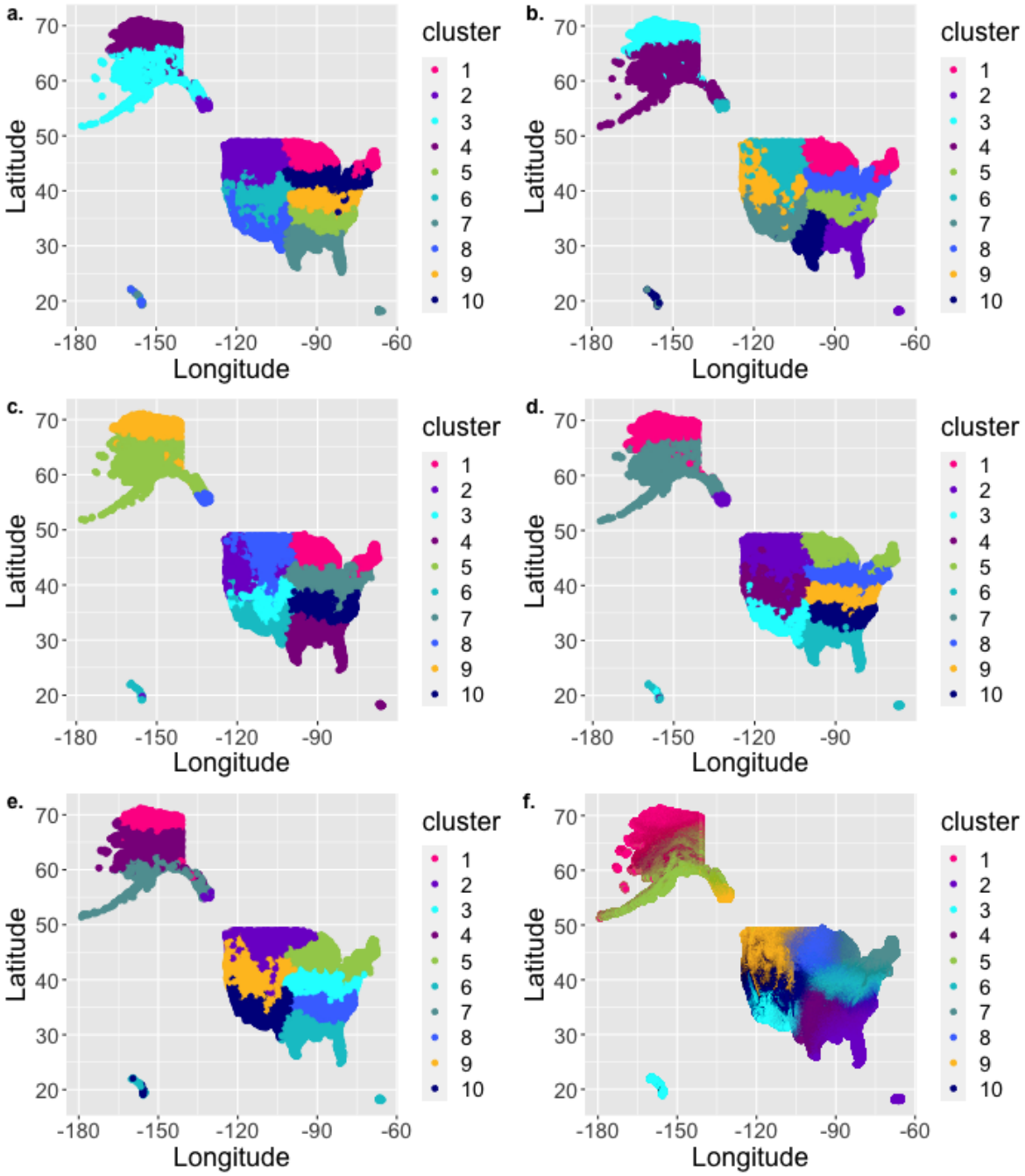

Figure S3. The cluster solution for a. sample 1, b. sample 2, c. sample 3, d. sample 4, e. sample 5, and d. the Bigmds (Table S1).

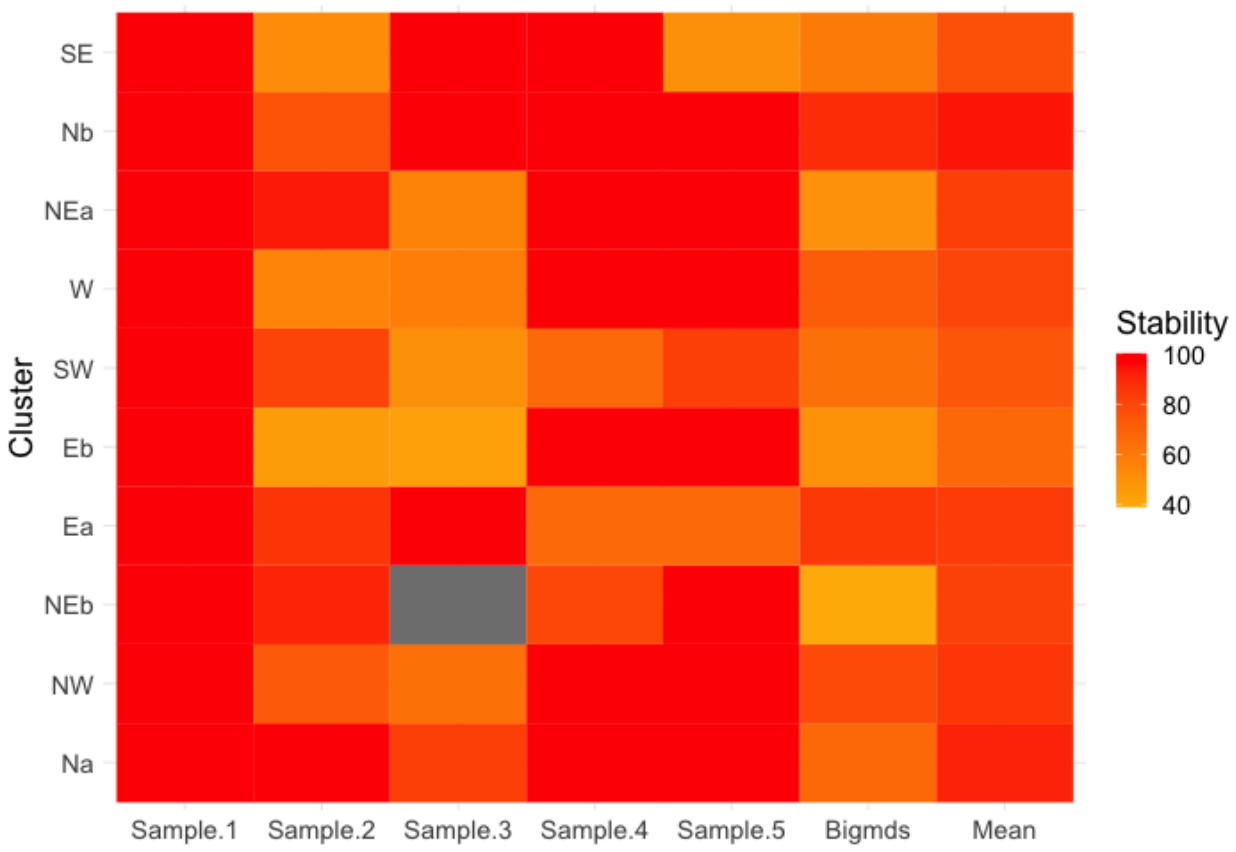

Figure S4. The stability of the example cases compared to the standard cluster solution (Figure 2). The mean stability is 82% across all cluster solutions. The gray region was not sampled by sample 3.