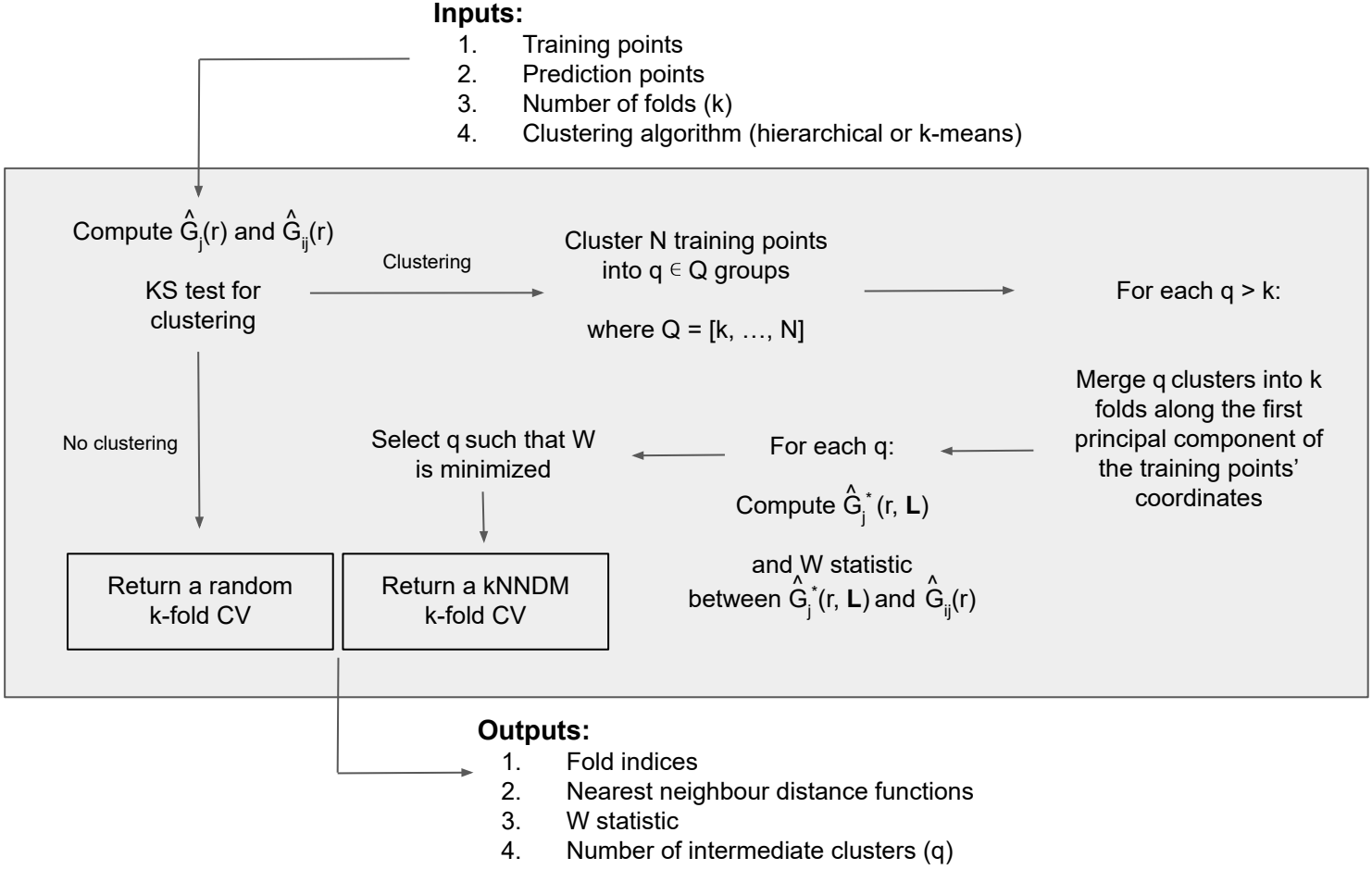
Inputs:
1. Training points
2. Prediction points
3. Number of folds (k)
4. Clustering algorithm (hierarchical or k-means)
Compute $\hat{G}_j(r)$ and $\hat{G}_{ij}(r)$
KS test for clustering
Clustering
No clustering
Cluster N training points into q ∈ Q groups
where Q = [k, …, N]
For each q > k:
Merge q clusters into k folds along the first principal component of the training points' coordinates
For each q:
Compute $\hat{G}_j^*(r, \mathbf{L})$
and W statistic between $\hat{G}_j^*(r, \mathbf{L})$ and $\hat{G}_{ij}(r)$
Select q such that W is minimized
Return a random k-fold CV
Return a kNNDM k-fold CV
Outputs:
1. Fold indices
2. Nearest neighbour distance functions
3. W statistic
4. Number of intermediate clusters (q)