Description of algorithm

1. For each column impute missing values using cubic splines. Any missing values coming before the first non-missing value or after the last non-missing value are not imputed since the behavior of extrapolating cubic splines before the first spline or after the last spline can sometimes be erratic.
2. For each unique pair and of columns:
   1. Center and by subtracting the mean/average of each.
   2. Calculate the following distance metric based on crosscorelation (taken from the function function CCorDistance in R’s [Tdist](https://cran.r-project.org/web/packages/TSdist/index.html) package (author Usue Mori).

(1)

where is the normalized cross-correlation between centered column and centered column at lag .

1. Do agglomerative hierarchical clustering The distance metric between individual columns and is given in (1).

For groups and of multiple columns the distance is calculated as the max of all distances between individual columns of u and v.

for any columns .

1. The selected criterion for stopping the agglomeration of different groups is that in order for two groups to be joined the distance must be less than the 20 % quantile of all distances between any two unique columns in the data.