

Simulation with algorithm 2

Sayani Gupta

Simulation design

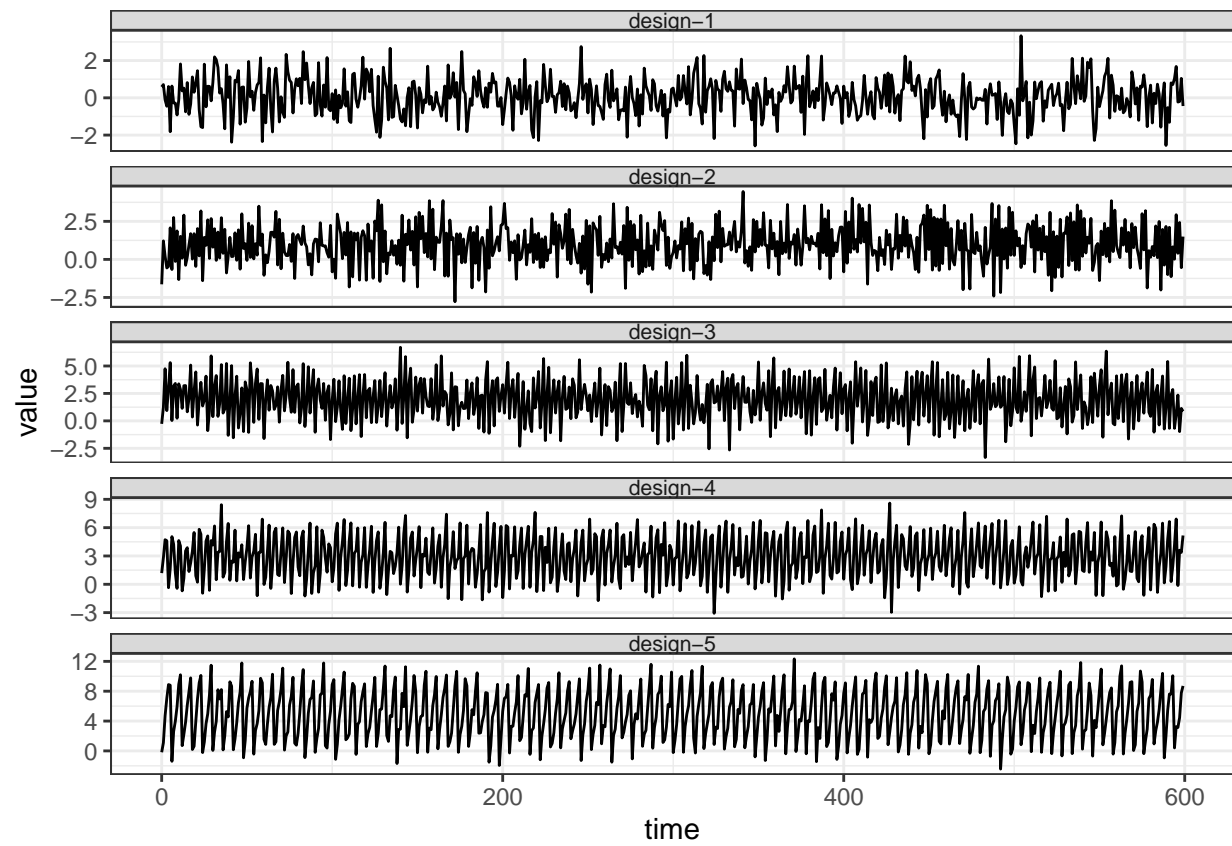
Three circular granularities g1, g2 and g3 are considered with levels 2, 3 and 4 respectively. Many time series with 600 observations are created using the five designs below, each of which is iterated four times. We anticipate to have five clusters, each with four time series conforming to the same design, once we execute the clustering.

Algorithm

Compute wpd for each granularity. The distance matrix is then computed with 20 time series as cases and three granularities as variables. Manhattan distances are computed and then hierarchical clustering is applied on them.

design	g1	g2	g3
design-1	no	no	no
design-2	yes	no	yes
design-3	no	yes	no
design-4	yes	no	yes
design-5	yes	yes	yes

Raw plots



Designs (Distribution of simulated data across different granularities)

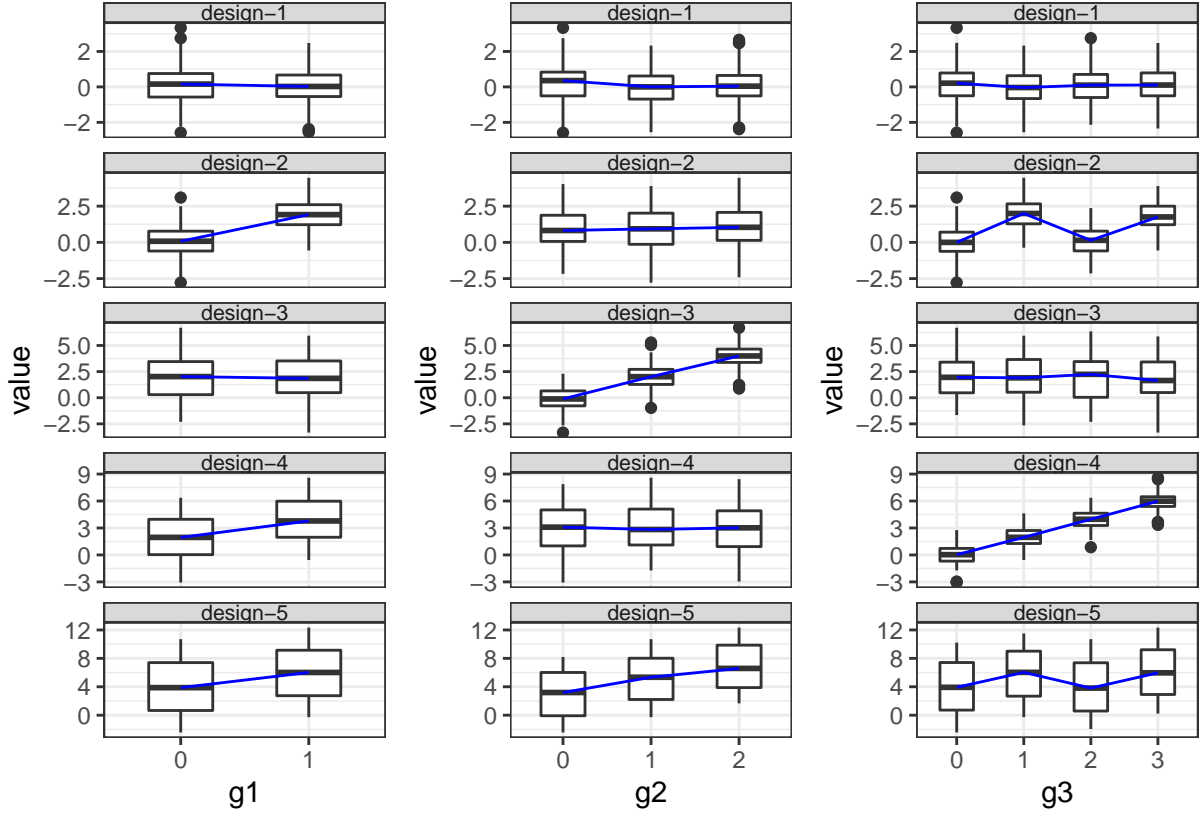


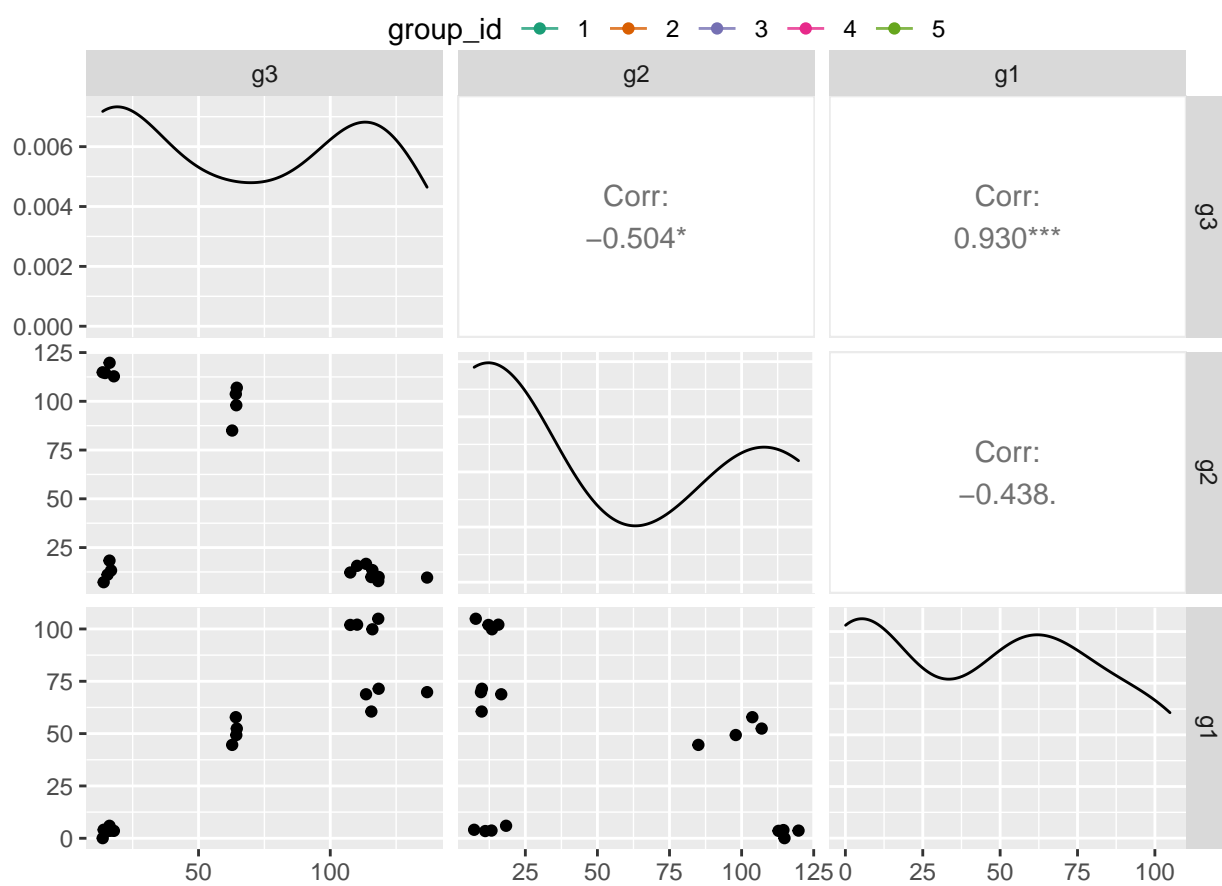
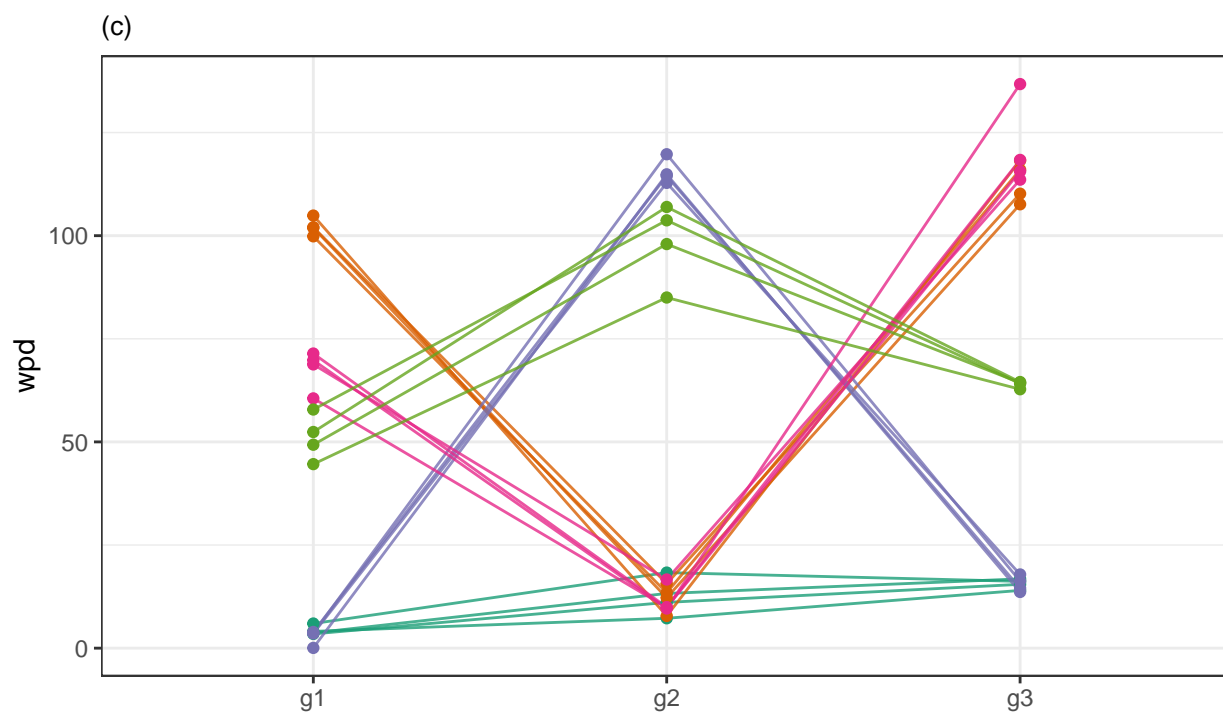
Table of wpd across designs and wpd

data_id	g3	g2	g1
design-1-1	15.471	11.095	3.450
design-1-2	13.994	7.241	4.058
design-1-3	16.196	18.318	5.966
design-1-4	16.795	13.256	3.692
design-2-1	107.613	12.214	101.916
design-2-2	110.182	15.639	102.042
design-2-3	118.215	7.807	104.875
design-2-4	115.995	13.435	99.841
design-3-1	14.546	114.497	3.924
design-3-2	13.623	114.872	0.062
design-3-3	17.866	112.808	3.554
design-3-4	16.243	119.741	3.665
design-4-1	118.368	9.938	71.425
design-4-2	115.595	9.859	60.551
design-4-3	113.580	16.643	68.790
design-4-4	136.751	9.612	69.808
design-5-1	64.519	106.959	52.402
design-5-2	64.154	103.724	57.853
design-5-3	64.341	97.969	49.319

data_id	g3	g2	g1
design-5-4	62.766	85.010	44.611

Clustering of designs

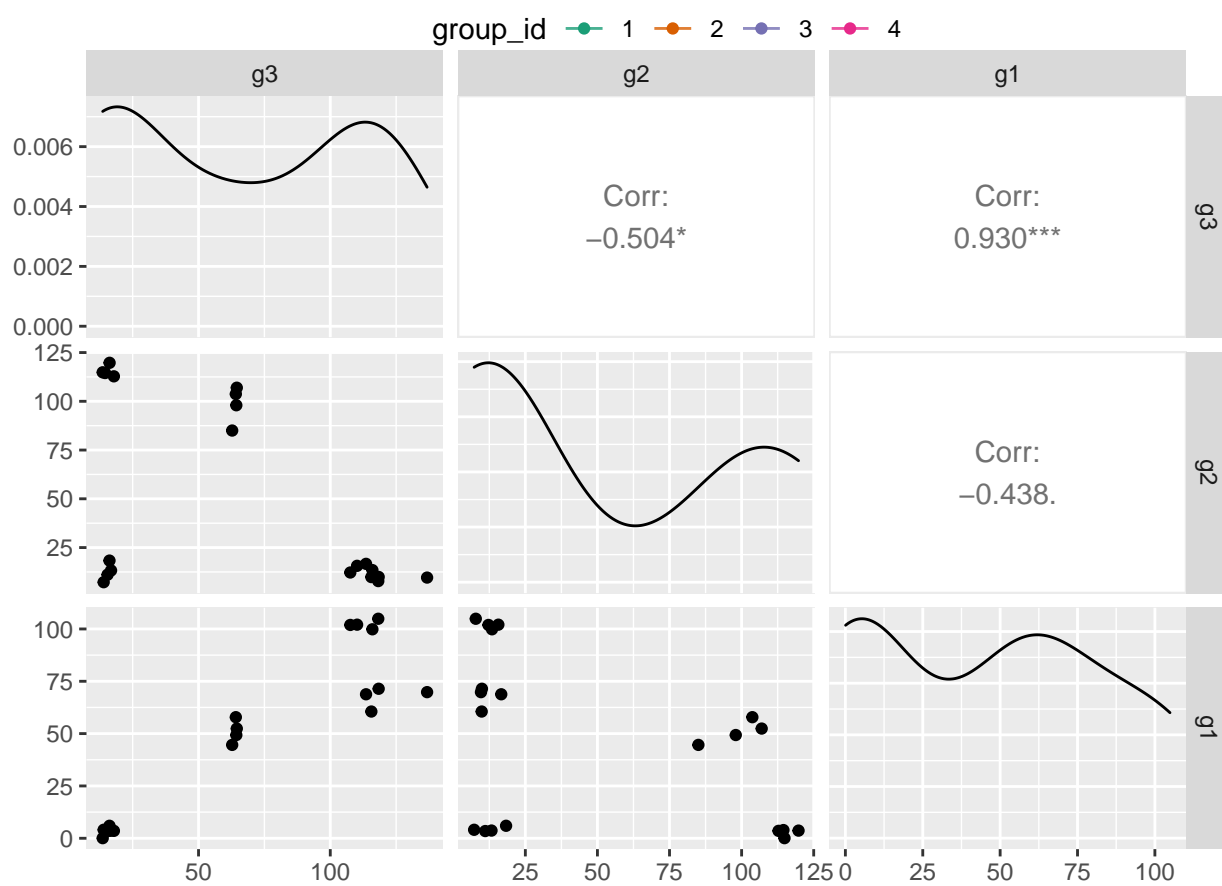
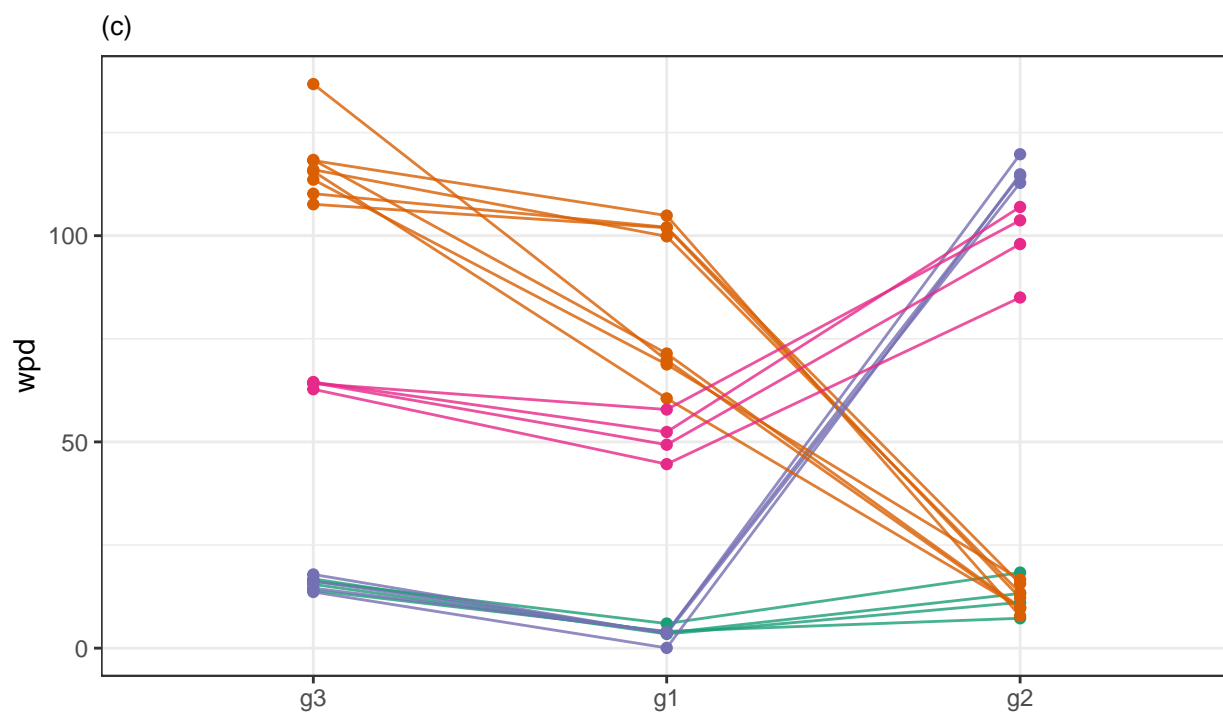
group	data_id
1	design-1-1
1	design-1-2
1	design-1-3
1	design-1-4
2	design-2-1
2	design-2-2
2	design-2-3
2	design-2-4
3	design-3-1
3	design-3-2
3	design-3-3
3	design-3-4
4	design-4-1
4	design-4-2
4	design-4-3
4	design-4-4
5	design-5-1
5	design-5-2
5	design-5-3
5	design-5-4



With optimal number of clusters

Design 2 and 4 form a part of the same cluster with the optimal number of clusters chosen as 4. Design 2 and 4 are similar in the sense that both of them changes across categories for g1 and g3, and remains constant for g2.

group	data_id
1	design-1-1
1	design-1-2
1	design-1-3
1	design-1-4
2	design-2-1
2	design-2-2
2	design-2-3
2	design-2-4
3	design-3-1
3	design-3-2
3	design-3-3
3	design-3-4
2	design-4-1
2	design-4-2
2	design-4-3
2	design-4-4
4	design-5-1
4	design-5-2
4	design-5-3
4	design-5-4



PCA based clustering

