

Simulation with algorithm 2- different means and one nuisance granularity

Sayani Gupta

Simulation design

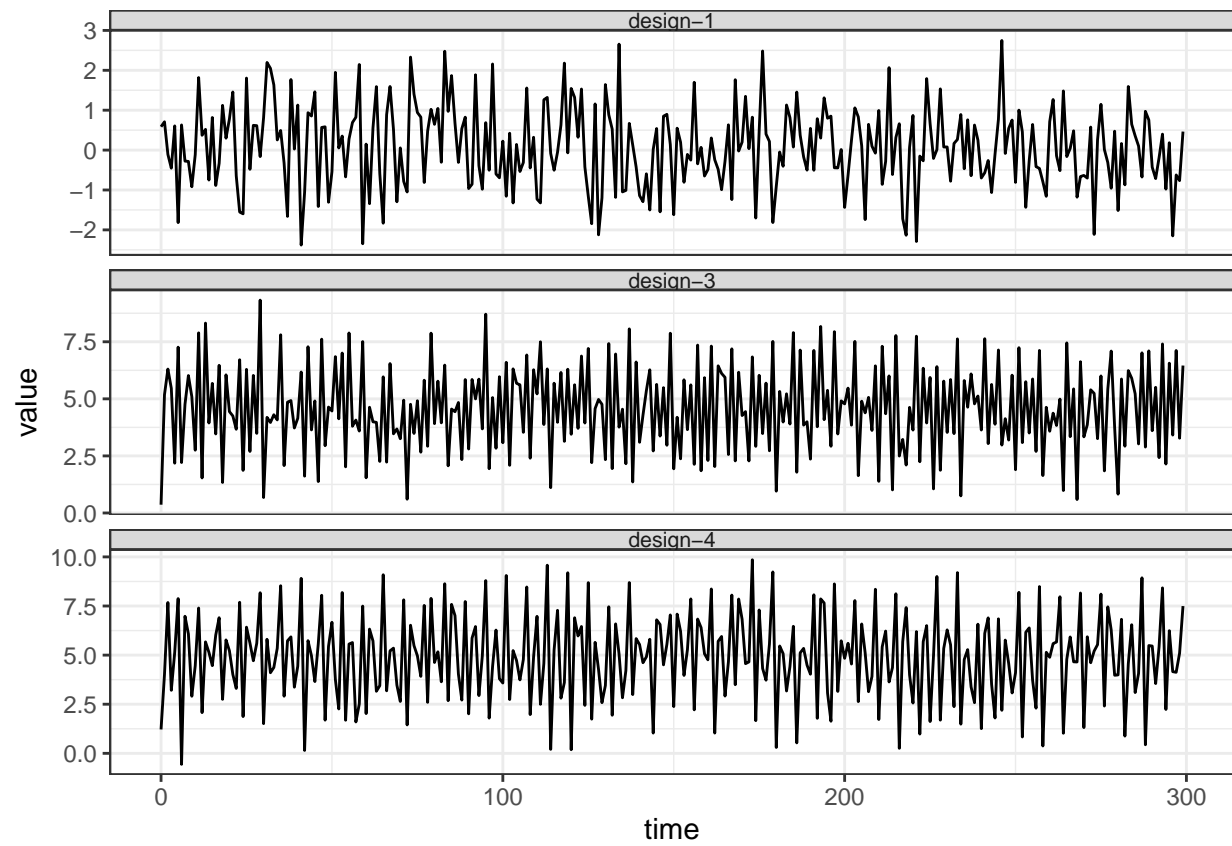
Three circular granularities g1, g2 and g3 are considered with levels 2, 3 and 4 respectively. Many time series with 600 observations are created using the four designs below, each of which is iterated five times. We anticipate to have four clusters, each with five time series conforming to the same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

Algorithm

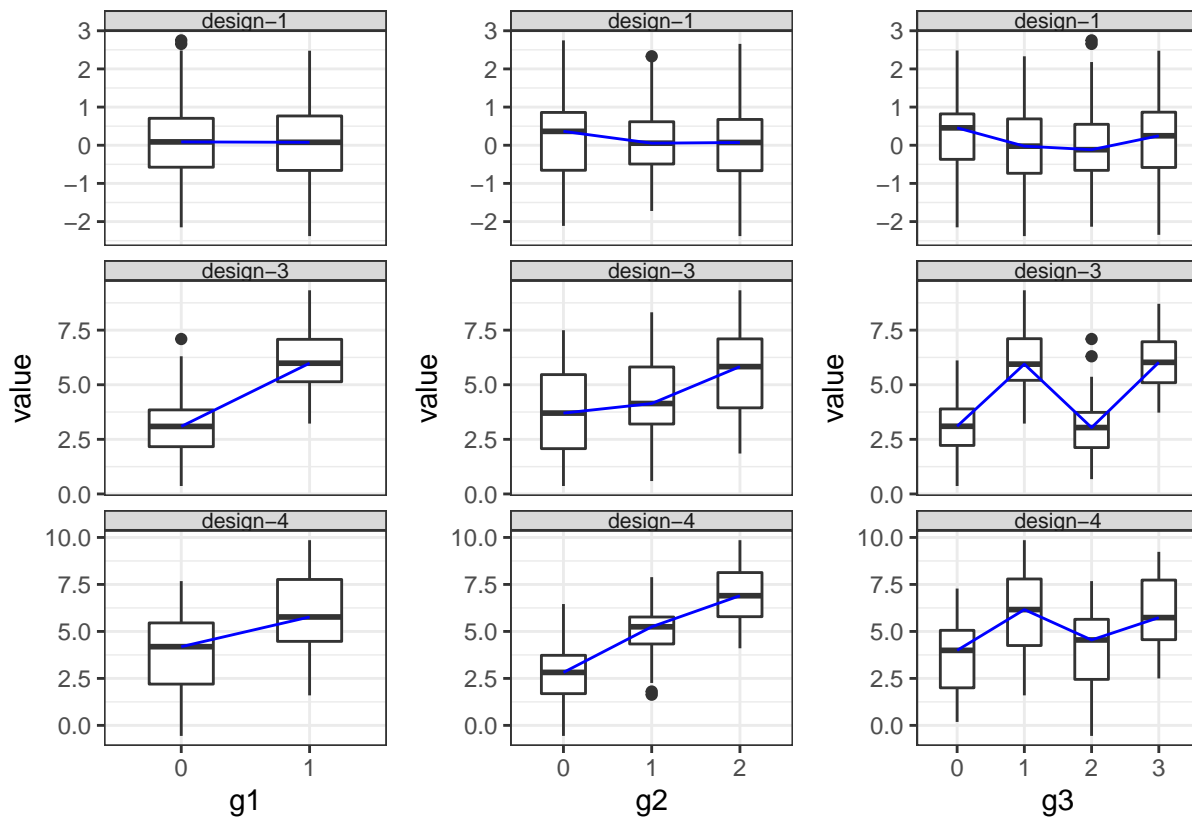
Compute wpd for each granularity. The distance matrix is then computed with 20 time series as cases and three granularities as variables. Manhattan distances are computed and then hierarchical clustering is applied on them.

```
## # A tibble: 16 x 2
## # Groups:   design [4]
##   design      s
##   <chr>   <dbl>
## 1 1      0
## 2 2      1
## 3 2      7
## 4 2      3
## 5 2      5
## 6 2      9
## 7 3      2
## 8 3      6
## 9 3      4
## 10 3     5
## 11 3     3
## 12 3     7
## 13 4     2
## 14 4     6
## 15 4     4
## 16 4     8
```

Raw plots



Designs (Distribution of simulated data across different granularities)



```
## # A tibble: 3 x 2
##   group      n
##   <int> <int>
## 1     1     5
## 2     2     5
## 3     3     5

## # A tibble: 5 x 2
##   group customer_id
##   <int> <chr>
## 1     1 design-1-s-1
## 2     1 design-1-s-2
## 3     1 design-1-s-3
## 4     1 design-1-s-4
## 5     1 design-1-s-5

## # A tibble: 5 x 2
##   group customer_id
##   <int> <chr>
## 1     2 design-3-s-1
## 2     2 design-3-s-2
## 3     2 design-3-s-3
## 4     2 design-3-s-4
## 5     2 design-3-s-5

## # A tibble: 5 x 2
```

```
##      group customer_id
##      <int> <chr>
## 1      3 design-4-s-1
## 2      3 design-4-s-2
## 3      3 design-4-s-3
## 4      3 design-4-s-4
## 5      3 design-4-s-5
```