

# kNNDM: k-fold Nearest Neighbour Distance Matching Cross-Validation for map accuracy estimation

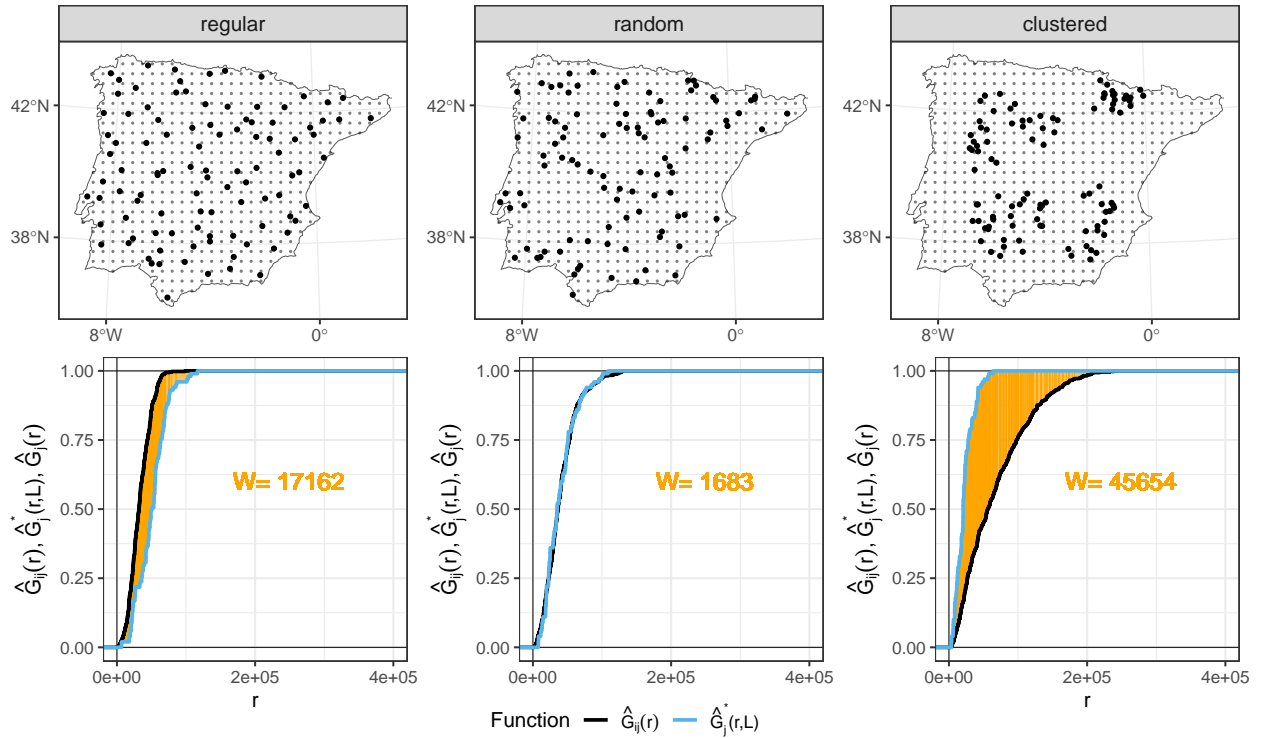
Jan Linnenbrink & Carles Milà

2023-02-17

This file contains the code to reproduce figures 1, 3, 5 & 6 of the paper “kNNDM: k-fold Nearest Neighbour Distance Matching Cross-Validation for map accuracy estimation” by J Linnenbrink, C Milà, M Ludwig & H Meyer.

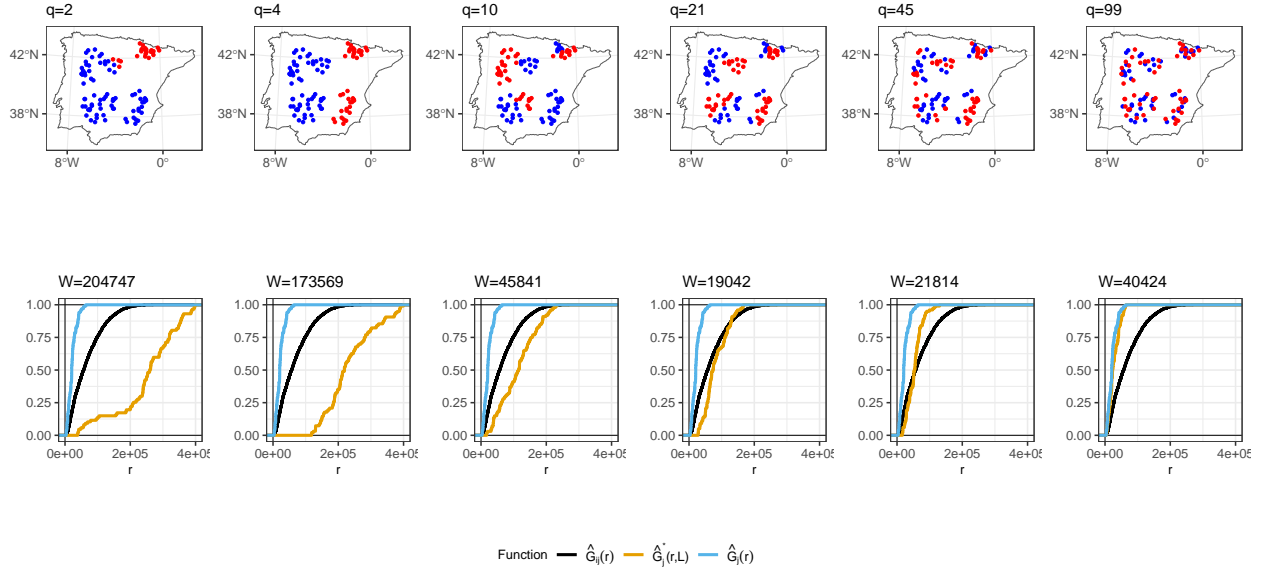
## Example: different clustering, NND ECDFs and W statistics

This code reproduces figure 1, where different configurations of training samples and their corresponding nearest neighbour distance (NND) empirical cumulative distribution functions (ECDF) are shown. Also, the Wasserstein statistic is shown in orange.



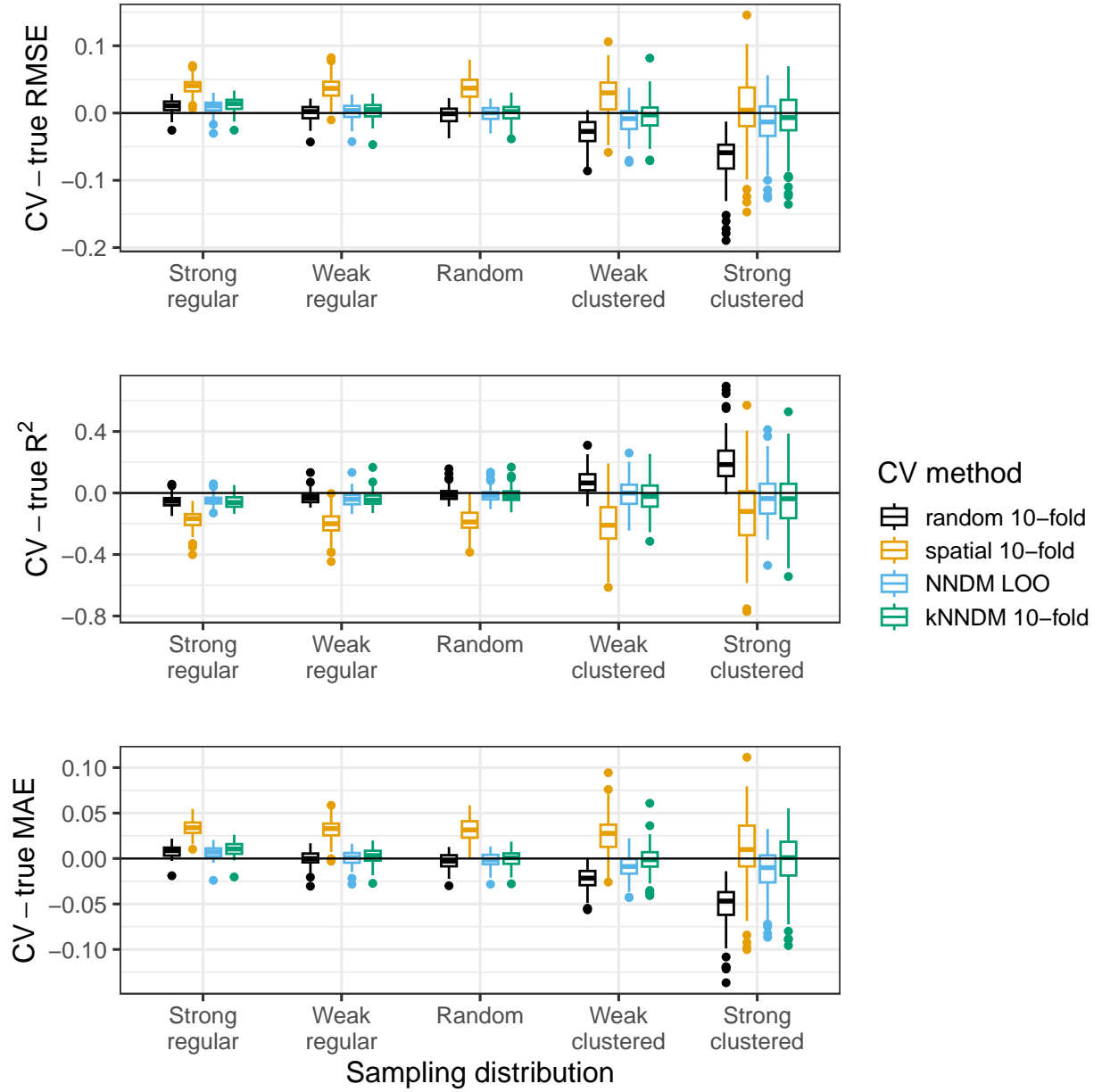
## kNNDM workflow

This code reproduces the kNNDM workflow shown in figure 3. Several numbers of clusters  $q$  are compared regarding their  $W$  statistic (bottom row).



## Simulation results

The following code reproduces figure 5, and shows the differences between cross-validated and true RMSE,  $R^2$  and MAE for different sampling distributions.



### Association CV - True error and W statistic

The following code reproduces figure 6, which shows the association between the absolute value difference between the CV and true map accuracy statistics and W statistic.

```
## [1] "Rsquared for Rsquared: 0.6 ; Rsquared for MAE: 0.73 ; Rsquared for RMSE: 0.66"
```

