Results of simulation 2: above-ground biomass and soil organic carbon stock

Jan Linnenbrink

2023-06-04

This file contains the code to reproduce the figures of section three of the master thesis "Development of a new spatial cross-validation strategy and its application for machine learning based modelling of ecological data" by Jan Linnenbrink
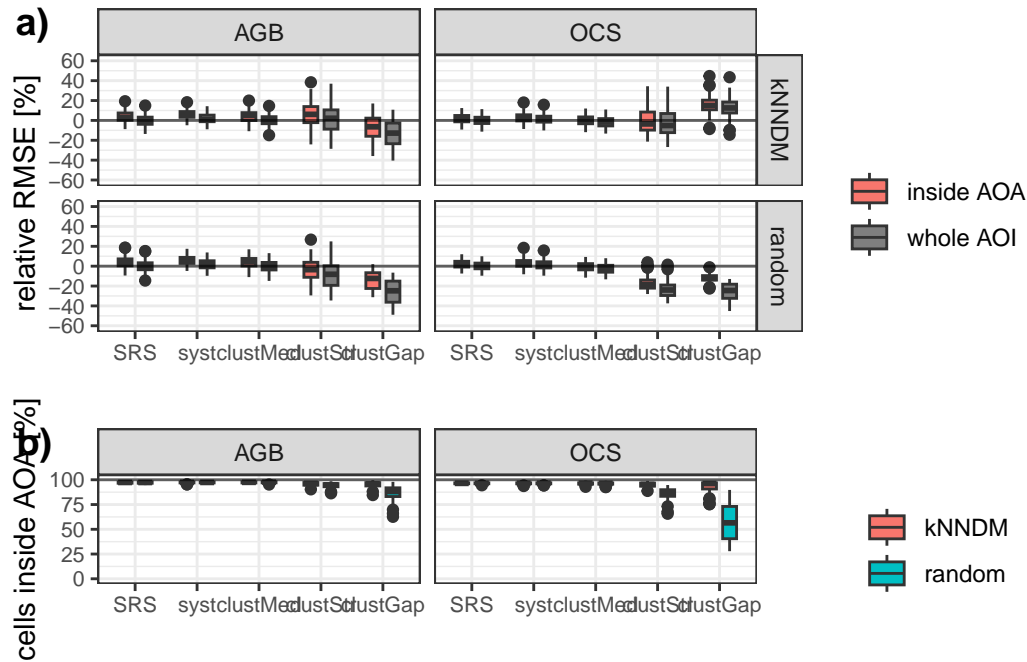
Comparison of the performance of different CV methods

The performance of the different CV methods measured as the percentual deviation of the CV estimated error to the true error. The two different prediction scenarios AGB (above ground biomass) and OCS (organic carbon stock) are shown in two different rows. The different sampling designs, ranging from simple random sampling (SRS) over regular (syst), medium clustered (clustMed), strong clustered (clustStrong) to gapped clustered (clustGap), are shown at the x-axis, while different CV methods are indicated by colours. Note that k=10 for all CV methods but NNDM, where k=N.
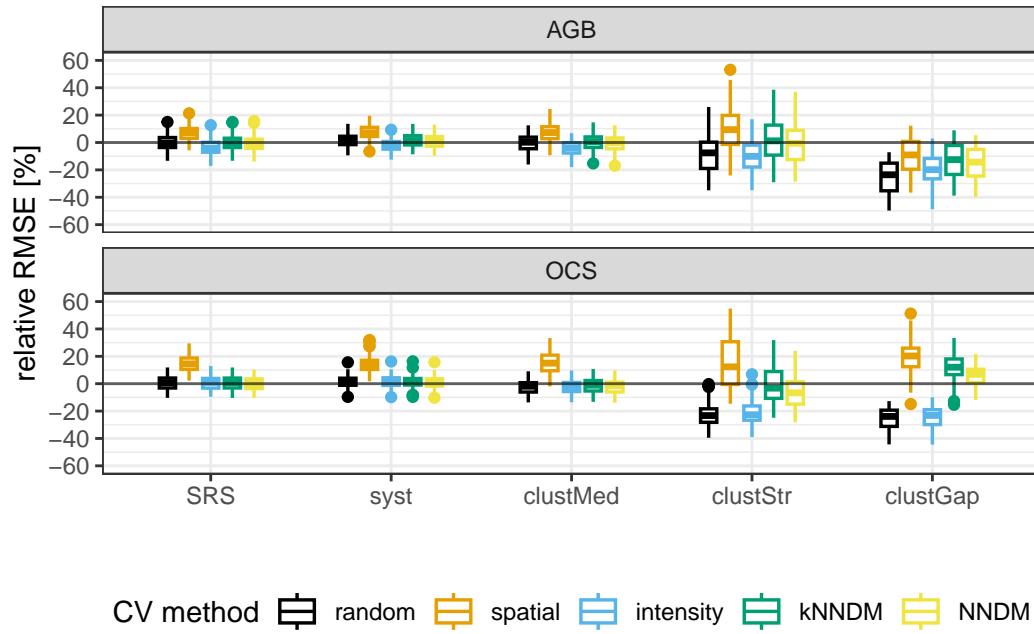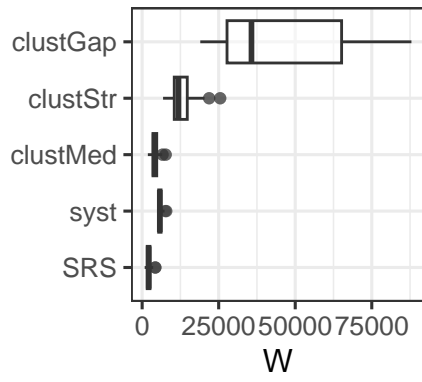
CV error estimates when AOA is taken into account

a) shows the performance of the different CV methods when the area of applicability (AOA, Meyer & Pebesma (2021)) is taken into account. The different colours indicate if the true RMSE was calculated for the whole area, or if cells outside the AOA were masked out. b) shows the percentage of grid cells that are inside the AOA. Colours indicate the different algorithms used. The two different prediction scenarios AGB (above ground biomass) and OCS (organic carbon stock) are shown in two different columns. The different sampling designs range from simple random sampling (SRS) over regular (syst), medium clustered (clustMed), strong clustered (clustStrong) to gapped clustered (clustGap) and are shown at the x-axis.

W stat and association with error

a) shows the performance of the different CV methods when the area of applicability (AOA, Meyer & Pebesma (2021)) is taken into account. The different colours indicate if the true RMSE was calculated for the whole area, or if cells outside the AOA were masked out. b) shows the percentage of grid cells that are inside the AOA. Colours indicate the different algorithms used. The two different prediction scenarios AGB (above ground biomass) and OCS (organic carbon stock) are shown in two different columns. The different sampling designs range from simple random sampling (SRS) over regular (syst), medium clustered (clustMed), strong clustered (clustStrong) to gapped clustered (clustGap) and are shown at the x-axis.

a)

b)

4