# 4. Validation Strategies

## 4.2 Standard and spatial cross-validation

| **Guidelines**   * Use the randomForest and caret packages in R * randomForest must be run with the same settings used in GEE   + 10-fold cross-validation repeated 100 times   + How well do we make predictions in the regions that we have samples? (remember that our samples are clustered and we need to take this into account) * leave-cluster-out cross-validation repeated 100 times   + create clusters in using spatial and temporal coordinates   + clusters can be created using k-means   + number of clusters: 10   + Example code: <https://github.com/AlexandreWadoux/SpatialValidation>   + How well do we make predictions in the regions that we DO NOT have samples? (remember that our samples are clustered and we need to take this into account) * Equations for validation statistics:   + mean error   + mean absolute error   + mean square error   + model efficiency |
| --- |

# 5. Map Collections and Analysis

## 5.2 Standard and spatial cross-validation

| **Guidelines**   * Compute global and biome-wise statistics * Validation statistics: present in box-plots and discuss * Present plot with predicted against observed |
| --- |