rs-local

Tuning of the resampling-local (rs-local) transfer learning algorithm to choose a SSL subset of size *k* optimized for the local prediction data set

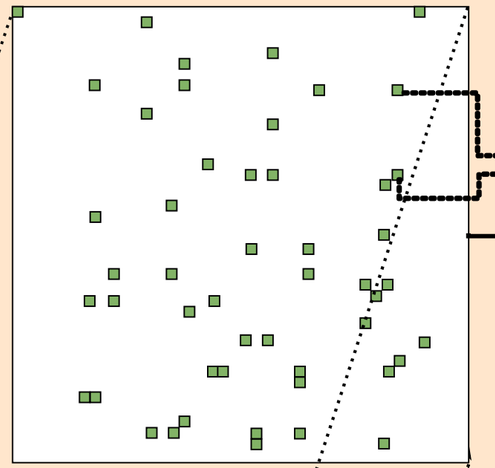
Tuning paramters:
– *r*
– *b*
– *k*

Weighted ranking based on RMSE:
– Rank samples based on how **frequently** they **appear in models that perform well** (RMSE) on site-specific samples
– Weight the ranks by considering the number of times a sample is selected in B repeats

Repeat the sampling procedure ***b = 50*** times

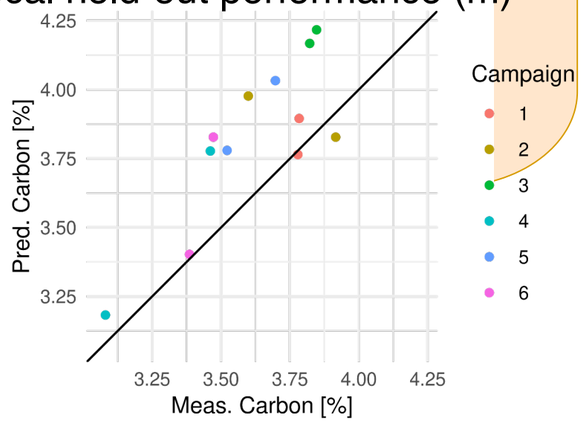
At each iteration *i* + 1, randomly draw ***k* = 50** samples **without replacement** from the current SSL subset

x 50



Tune a partial least squares regression (PLSR) model using *k* samples

local hold-out performance (m)



n_{SSL} = 2000

K: Target (number of) spectral library samples selected by rs-local

k = 50

m = 12

M: site-specific samples with reference analyses

Tune model on final set

Prediction on local test set with final PLSR model developed on *k* subset + *m* local calibration samples

Collect row indices of k selected (**idx$_k$**) sampled observations together with RMSE$_m$ *(local hold-out set)*

B	**RMSE**	**idx$_k$**
1	0.23	c(4, 11, 23, ...)
2	0.11	c(1, 3, 222, ...)
---	---	...
50	0.15	c(14, 45, 99, ...)

Performance-driven library reduction

At each iteration ***i + 1,*** ***remove*** ***r * n$_{SSL, i}$*** samples that are consistently in weakest models

Iteration i = 0: *remove* 0.05 * 2000 **= 100 samples**

iteration *i* = 1*:*** *remove* 0.05 * 1900 **= 95 samples**