Choosing representative BRT-models from sets of replicates

Supplementary Information
Cape vs SWA publication

Ruan van Mazijk

2019-01-31

A representative model from a set of replicates is one which falls within the 5% quantile bound about the set's median values for *all* the various model-quality statistics (nt, R^2_{pseudo}, R^2_{E-O}). When there is more than one model that meets this criterion, one is chosen at random.

```
is_between <- function(x, a, b) {
  (x >= a) & (x <= b)
}
determine_bounds <- function(reps, bound_quantile = 0.05) {
  columns <- c("nt", "pseudo_r2", "pred_obs_r2", "pred_obs_r2_exp")
  summarise_at(reps, columns, funs(
    # E.g. 0.025 + 0.025 = 5% quantile about the median
    lwr = quantile(., (0.5 - (bound_quantile / 2))),
    upr = quantile(., (0.5 + (bound_quantile / 2)))
  ))
}
filter_reps_by_bounds <- function(reps, bounds) {
  bounds %$% filter(reps,
    is_between(nt,          nt_lwr,          nt_upr),
    is_between(pseudo_r2,   pseudo_r2_lwr,   pseudo_r2_upr),
    is_between(pred_obs_r2, pred_obs_r2_lwr, pred_obs_r2_upr)
  )
}
```

```
reps <- read.csv(...)
bounds <- determine_bounds(reps)
filter_reps_by_bounds(reps, bounds)
```