

## WeightedLinearEnsemble evaluation

```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.0.5

library(dplyr)

## Warning: package 'dplyr' was built under R version 4.0.5

## 
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
## 
##     filter, lag

## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union

library(tidyr)

## Warning: package 'tidyr' was built under R version 4.0.5

library(comprehendr)

## Warning: package 'comprehendr' was built under R version 4.0.5

library(stringr)
library(ungeviz)
library(relayer)

## Note: The package "relayer" is highly experimental. Use at your own risk.

library(patchwork)

## Warning: package 'patchwork' was built under R version 4.0.3
```

```

base_dir <- "D:/skola/1/weighted_ensembles/tests/test_cifar_2021/data/data_tv_5000_c100/0/complete"
net_df <- read.csv(file.path(base_dir, "net_metrics.csv"))
ens_df_cal <- read.csv(file.path(base_dir, "ens_cal_metrics.csv"))
ens_df_pwc <- read.csv(file.path(base_dir, "ens_pwc_metrics.csv"))

net_long <- pivot_longer(net_df,
  cols = c("accuracy", "nll", "ece"),
  names_to = "metric", values_to = "value"
)
ens_cal_long <- pivot_longer(ens_df_cal,
  cols = c("accuracy", "nll", "ece"),
  names_to = "metric", values_to = "value"
)
ens_pwc_long <- pivot_longer(ens_df_pwc,
  cols = c("accuracy", "nll", "ece"),
  names_to = "metric", values_to = "value"
)

networks <- net_df$network

comb_stats_df <- data.frame(matrix(
  ncol = 14, nrow = 0,
  dimnames = list(NULL, c(
    "combination_size", "combination_id",
    "acc_min", "acc_max", "acc_avg", "acc_var",
    "nll_min", "nll_max", "nll_avg", "nll_var",
    "ece_min", "ece_max", "ece_avg", "ece_var"
  )))
))

for (sss in unique(ens_df_cal$combination_size))
{
  for (ssi in unique(ens_df_cal %>%
    filter(combination_size == sss) %>%
    pull(combination_id)))
  {
    cur_nets_vec <- to_vec(
      for (net in networks) {
        if (str_replace_all(net, "-", ".") %in% colnames(ens_cal_long) &&
          (ens_cal_long %>%
            filter(combination_size == sss & combination_id == ssi) %>%
            pull(str_replace_all(net, "-", ".")))[1] == "True") {
          net
        }
      }
    )
    cur_nets <- net_df %>% filter(network %in% cur_nets_vec)
    comb_stats_df[nrow(comb_stats_df) + 1, ] <- c(
      sss, ssi,
      min(cur_nets$accuracy), max(cur_nets$accuracy), mean(cur_nets$accuracy), var(cur_nets$accuracy),
      min(cur_nets$nll), max(cur_nets$nll), mean(cur_nets$nll), var(cur_nets$nll),
      min(cur_nets$ece), max(cur_nets$ece), mean(cur_nets$ece), var(cur_nets$ece)
    )
  }
}

```

```

    }

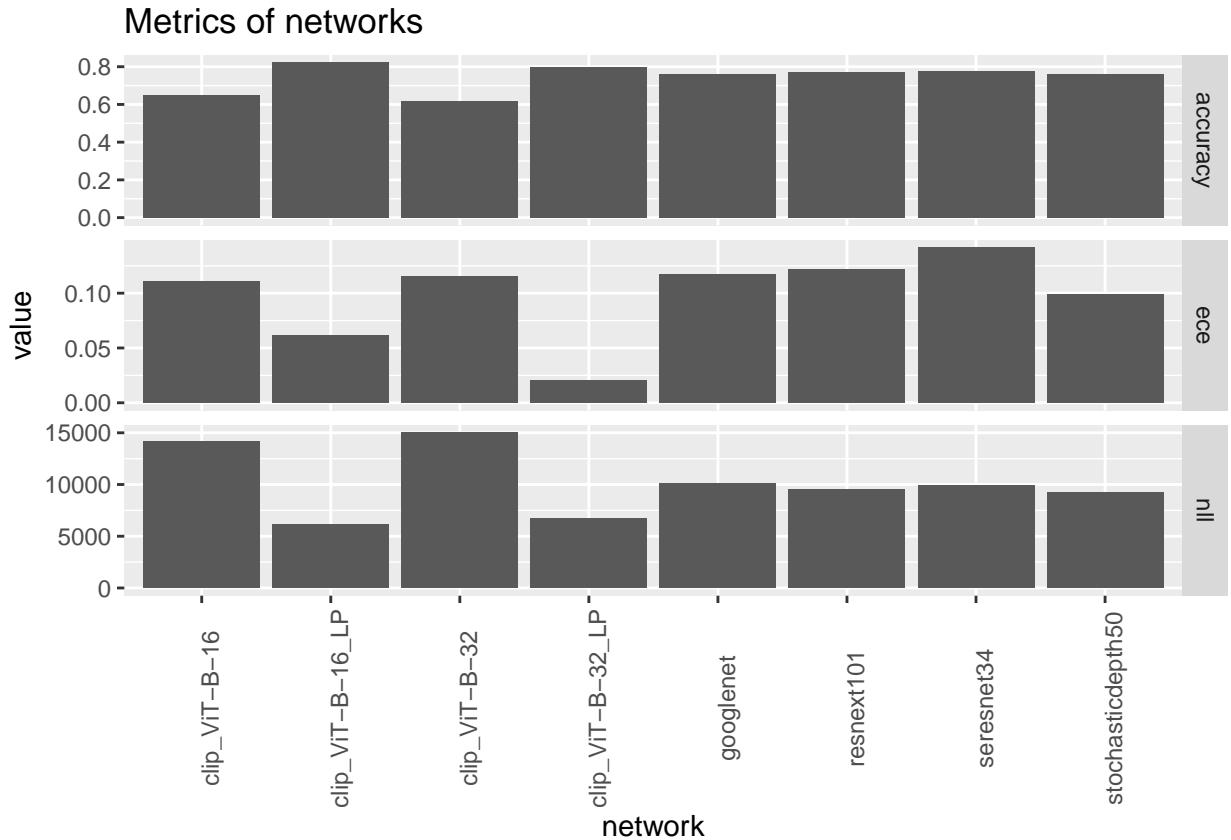
ens_df_cal <- merge(ens_df_cal, comb_stats_df)
ens_df_cal$acc_imp_avg <- ens_df_cal$accuracy - ens_df_cal$acc_avg
ens_df_cal$acc_imp_max <- ens_df_cal$accuracy - ens_df_cal$acc_max

ens_df_pwc <- merge(ens_df_pwc, comb_stats_df)
ens_df_pwc$acc_imp_avg <- ens_df_pwc$accuracy - ens_df_pwc$acc_avg
ens_df_pwc$acc_imp_max <- ens_df_pwc$accuracy - ens_df_pwc$acc_max

nets_plot <- ggplot(data = net_long) +
  geom_col(mapping = aes(x = network, y = value)) +
  facet_grid(rows = vars(metric), scales = "free") +
  theme(axis.text.x = element_text(angle = 90)) +
  ggtitle("Metrics of networks")

nets_plot

```



```

ens_pwc_plt_df <- ens_df_pwc %>% filter(combining_method != "lda")
ens_cal_plt_df <- ens_df_cal

ens_pwc_plt_df$combining_method <- factor(ens_pwc_plt_df$combining_method,
  levels = c(
    "average", "prob_average",

```

```

    "cal_average", "cal_prob_average",
    "logreg", "logreg_sweep_C",
    "logreg_no_interc", "logreg_no_interc_sweep_C"
)
)

for (sss in unique(ens_cal_plt_df$combination_size))
{
  for (ssi in unique(ens_cal_plt_df %>%
    filter(combination_size == sss) %>%
    pull(combination_id)))
  {
    cur_ens_cal <- ens_cal_plt_df %>% filter(combination_size == sss &
      combination_id == ssi)
    cur_ens_pwc <- ens_pwc_plt_df %>% filter(combination_size == sss &
      combination_id == ssi)
    cur_nets_vec <- to_vec(
      for (net in networks) {
        if (str_replace_all(net, "-", ".") %in% colnames(cur_ens_cal) &&
          cur_ens_cal[[str_replace_all(net, "-", ".")]][1] == "True") {
          net
        }
      }
    )
    cur_nets <- net_df %>% filter(network %in% cur_nets_vec)

    acc_plot <- ggplot() +
      (
        geom_hline(
          data = cur_nets,
          mapping = aes(yintercept = accuracy, colour1 = network),
          linetype = "dashed"
        ) %>%
          rename_geom_aes(new_aes = c("colour" = "colour1"))
      ) +
      geom_hline(
        data = cur_ens_cal,
        mapping = aes(yintercept = accuracy, color = "cal ensemble")
      ) +
      (
        geom_hpline(
          data = cur_ens_pwc,
          mapping = aes(
            x = combining_method, y = accuracy,
            colour2 = coupling_method
          ),
          size = 0.8, width = 0.11,
          position = position_dodge(width = 0.65)
        ) %>%
          rename_geom_aes(new_aes = c("colour" = "colour2"))
      ) +
      scale_colour_brewer(
        aesthetics = "colour1", palette = 1,

```

```

        name = "network", type = "qual"
    ) +
    scale_colour_brewer(
        aesthetics = "colour2", palette = 2,
        name = "coupling method", type = "qual"
    ) +
    scale_color_manual(values = c("black"), name = "averaging ensemble") +
    theme(
        axis.text.x = element_blank(),
        axis.title.x = element_blank()
    )

y_limits <- layer_scales(acc_plot)$y$get_limits()
x_limits <- layer_scales(acc_plot)$x$get_limits()
all_y_lim <- c(y_limits[1], cur_ens_cal$all_cor)

acc_plot <- acc_plot +
    geom_rect(
        data = cur_ens_cal,
        mapping = aes(
            xmin = 0.5,
            xmax = length(x_limits) + 0.5,
            ymin = max(all_cor, y_limits[1]), ymax = all_cor + err_incons
        ),
        fill = "orange",
        alpha = 0.3, color = NA
    )

if (all_y_lim[1] < all_y_lim[2]) {
    acc_plot <- acc_plot +
        geom_rect(
            data = cur_ens_cal,
            mapping = aes(
                xmin = 0.5,
                xmax = length(x_limits) + 0.5,
                ymin = all_y_lim[1], ymax = all_y_lim[2]
            ), fill = "green",
            alpha = 0.3, color = NA
        )
}

nll_plot <- ggplot() +
    (
        geom_hline(
            data = cur_nets,
            mapping = aes(yintercept = nll, colour1 = network),
            linetype = "dashed"
        ) %>%
            rename_geom_aes(new_aes = c("colour" = "colour1"))
    ) +
    geom_hline(
        data = cur_ens_cal,
        mapping = aes(yintercept = nll, color = "cal ensemble")
    )

```

```

) +
(
  geom_hline(
    data = cur_ens_pwc,
    mapping = aes(
      x = combining_method, y = nll,
      colour2 = coupling_method
    ),
    size = 0.8, width = 0.11,
    position = position_dodge(width = 0.65)
  ) %>%
    rename_geom_aes(new_aes = c("colour" = "colour2"))
) +
  scale_colour_brewer(
    aesthetics = "colour1", palette = 1,
    name = "network", type = "qual"
  ) +
  scale_colour_brewer(
    aesthetics = "colour2", palette = 2,
    name = "coupling method", type = "qual"
  ) +
  scale_color_manual(values = c("black"), name = "averaging ensemble") +
  scale_y_reverse() +
  theme(
    axis.text.x = element_blank(),
    axis.title.x = element_blank()
  )

ece_plot <- ggplot() +
(
  geom_hline(
    data = cur_nets,
    mapping = aes(yintercept = ece, colour1 = network),
    linetype = "dashed"
  ) %>%
    rename_geom_aes(new_aes = c("colour" = "colour1"))
) +
  geom_hline(
    data = cur_ens_cal,
    mapping = aes(yintercept = ece, color = "cal ensemble")
  ) +
(
  geom_hline(
    data = cur_ens_pwc,
    mapping = aes(
      x = combining_method, y = ece,
      colour2 = coupling_method
    ),
    size = 0.8, width = 0.11,
    position = position_dodge(width = 0.65)
  ) %>%
    rename_geom_aes(new_aes = c("colour" = "colour2"))
) +

```

```

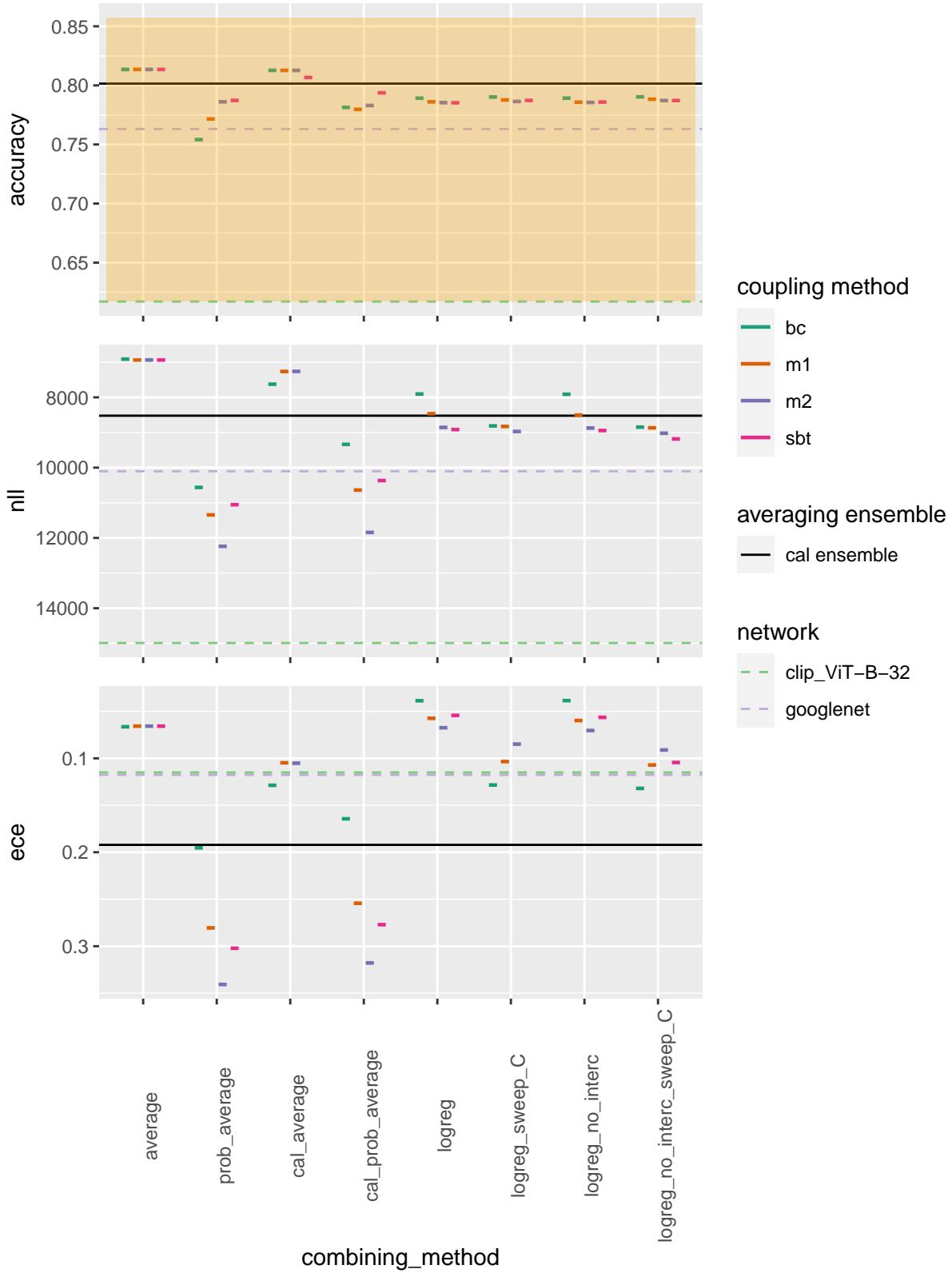
scale_colour_brewer(
  aesthetics = "colour1", palette = 1,
  name = "network", type = "qual"
) +
scale_colour_brewer(
  aesthetics = "colour2", palette = 2,
  name = "coupling method", type = "qual"
) +
scale_color_manual(values = c("black"), name = "averaging ensemble") +
scale_y_reverse() +
theme(axis.text.x = element_text(angle = 90))

print(acc_plot / nll_plot / ece_plot + plot_layout(guides = "collect") +
plot_annotation(title = paste(
  "Ensemble metrics",
  paste(
    c("Error inconsistency", cur_ens_cal$err_incons[[1]]),
    collapse = " "
  ),
  sep = "\n"
)))
}

}

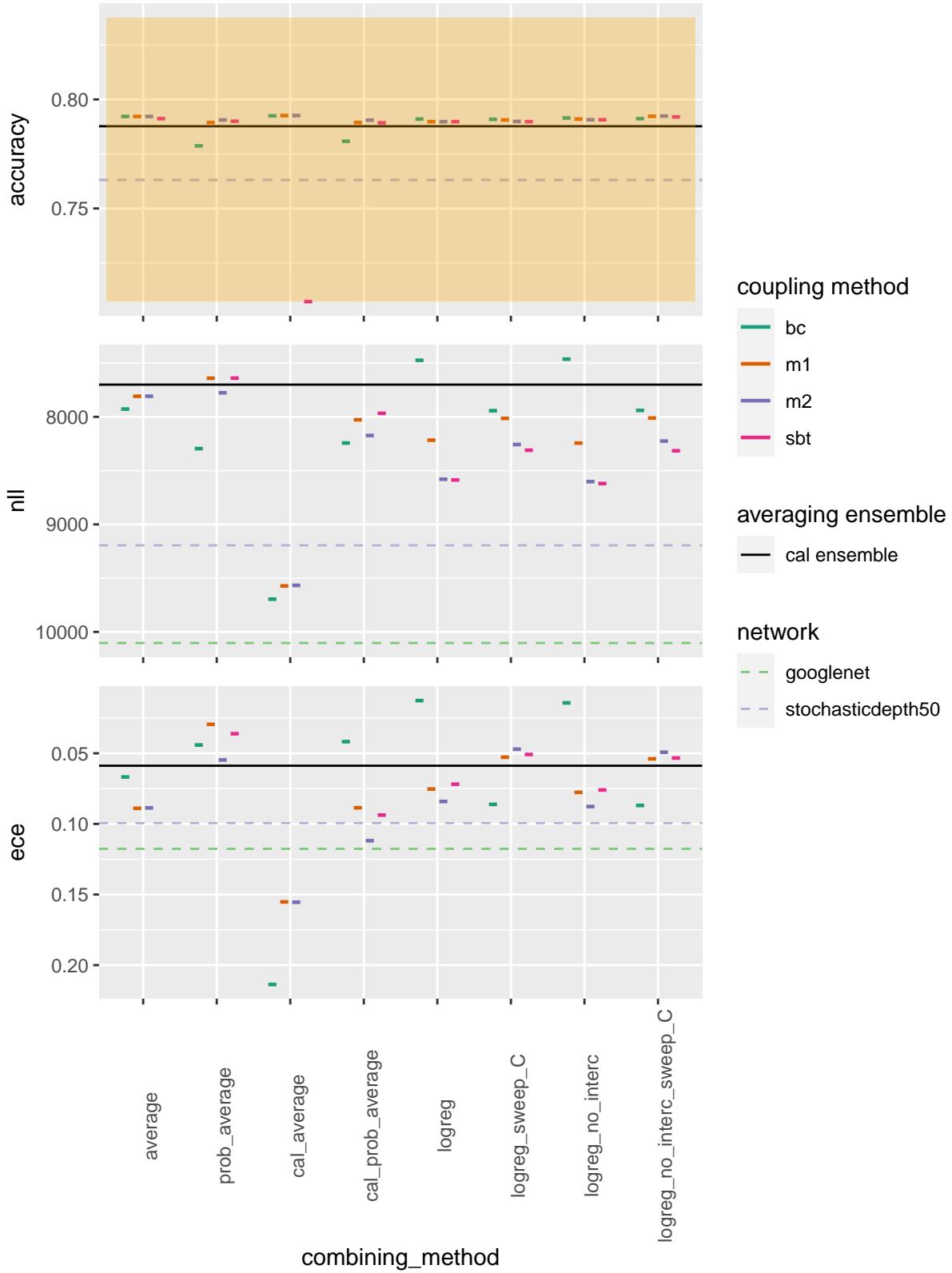
```

Ensemble metrics  
Error inconsistency 0.334399998188019

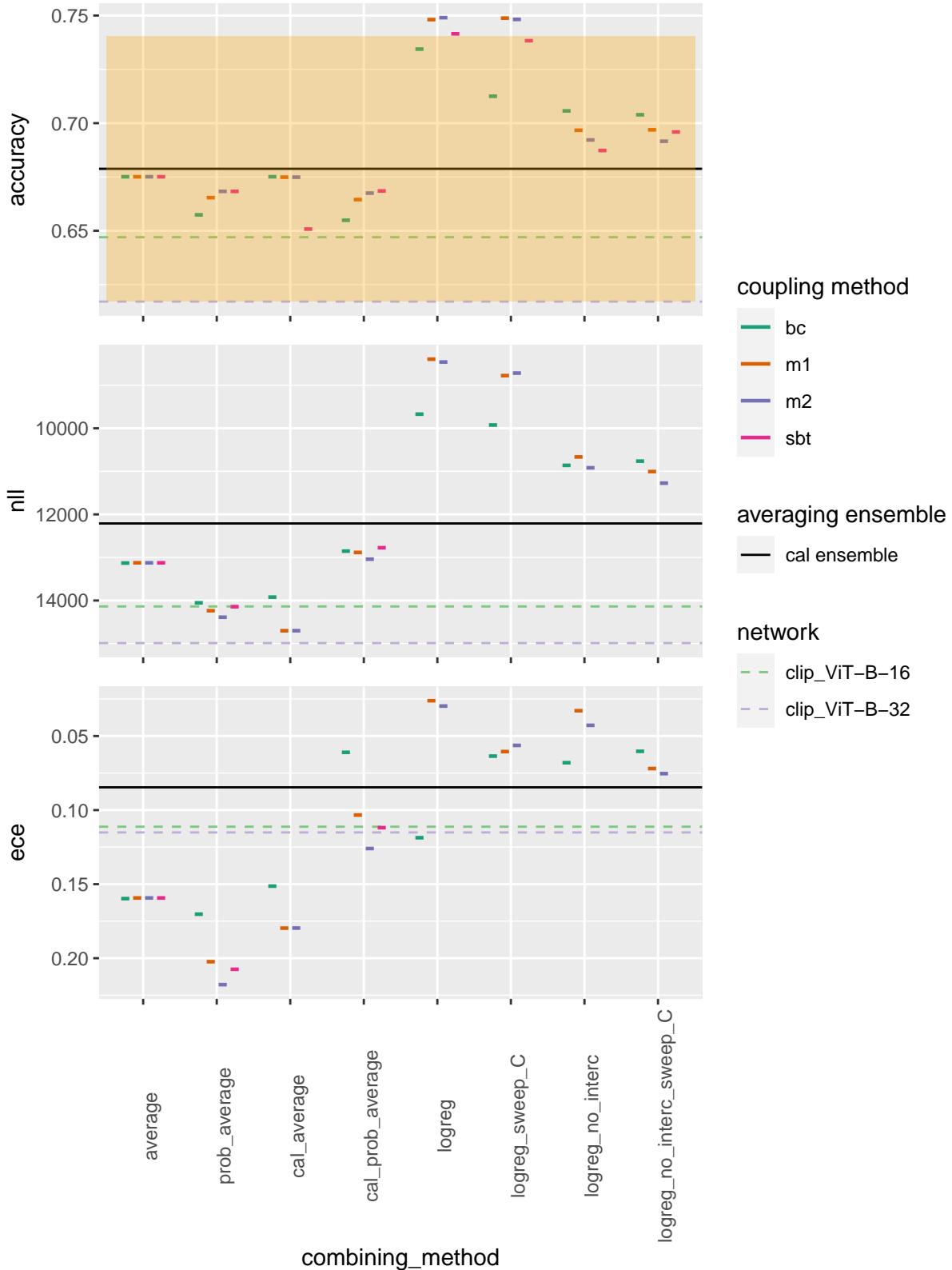


## Ensemble metrics

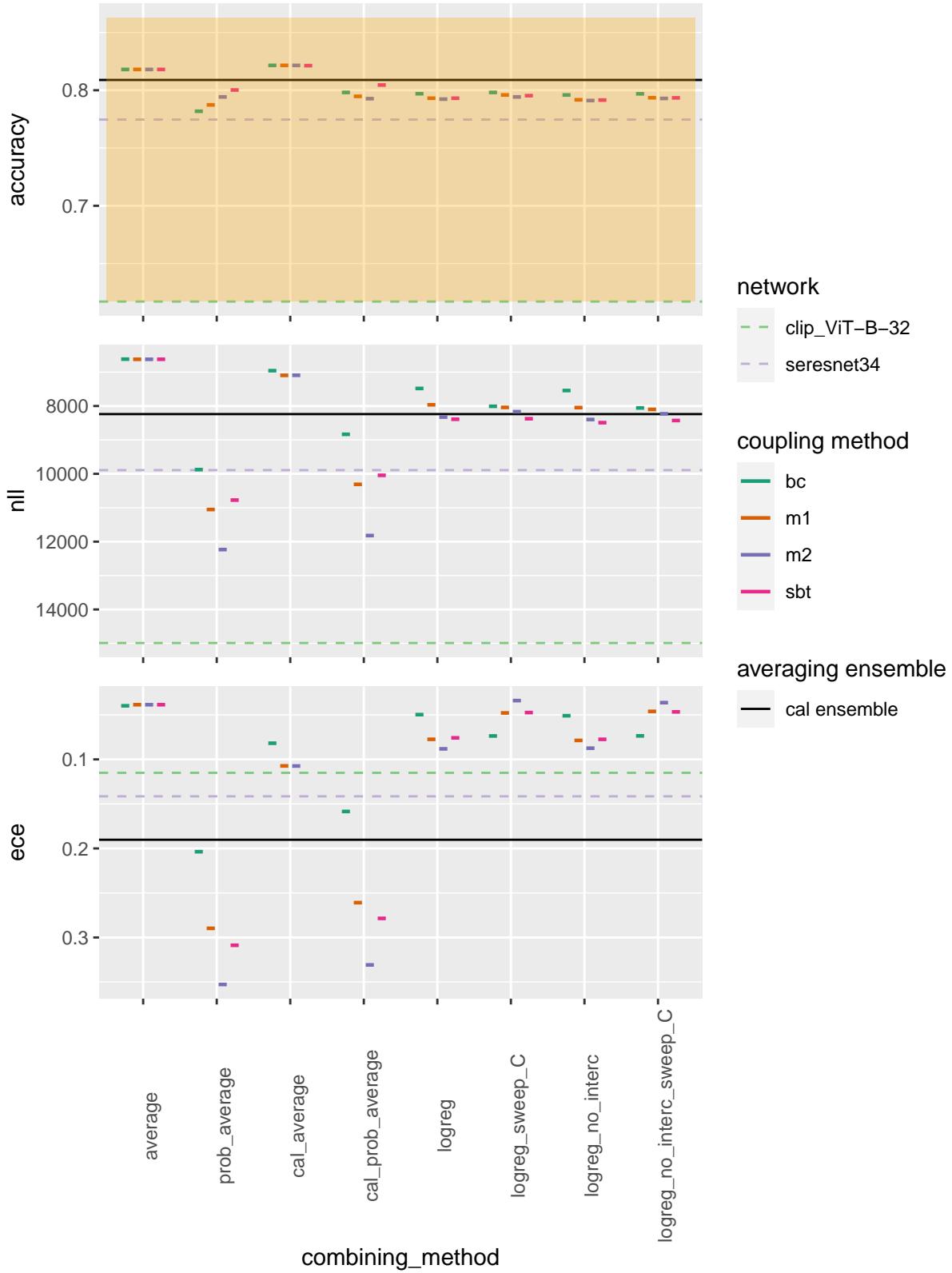
Error inconsistency 0.148800000548363



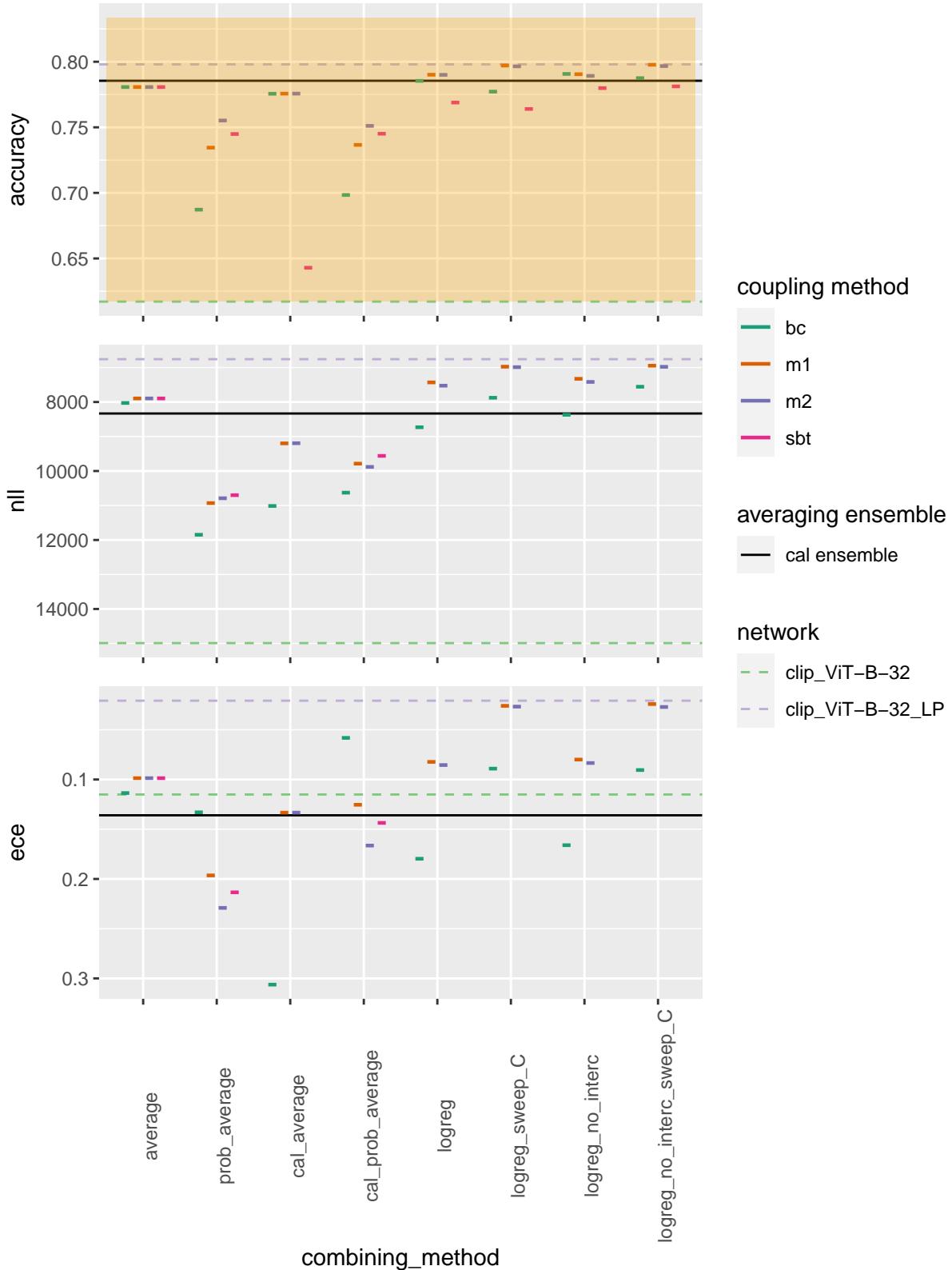
Ensemble metrics  
Error inconsistency 0.216699987649918



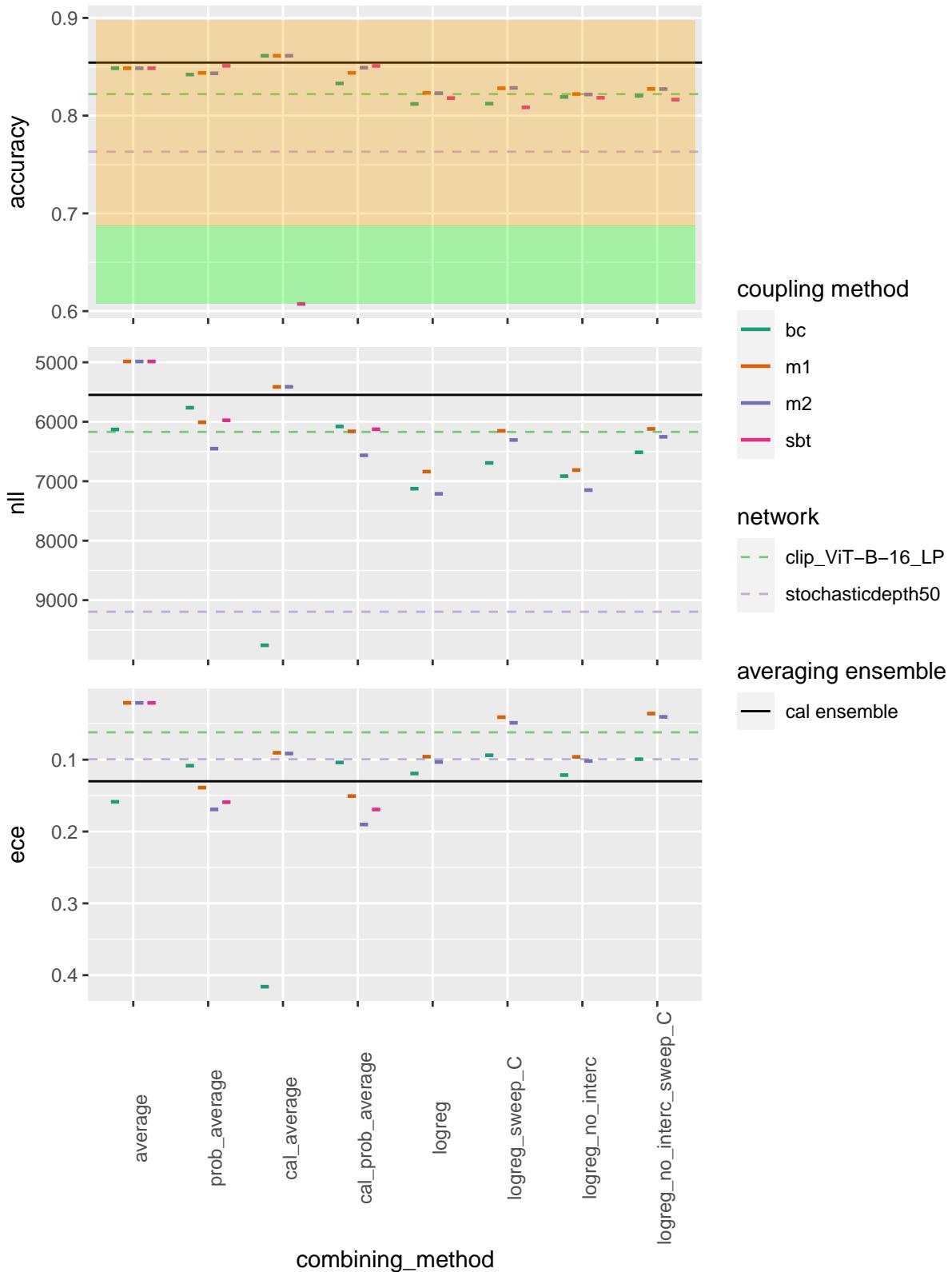
Ensemble metrics  
Error inconsistency 0.33390000462532



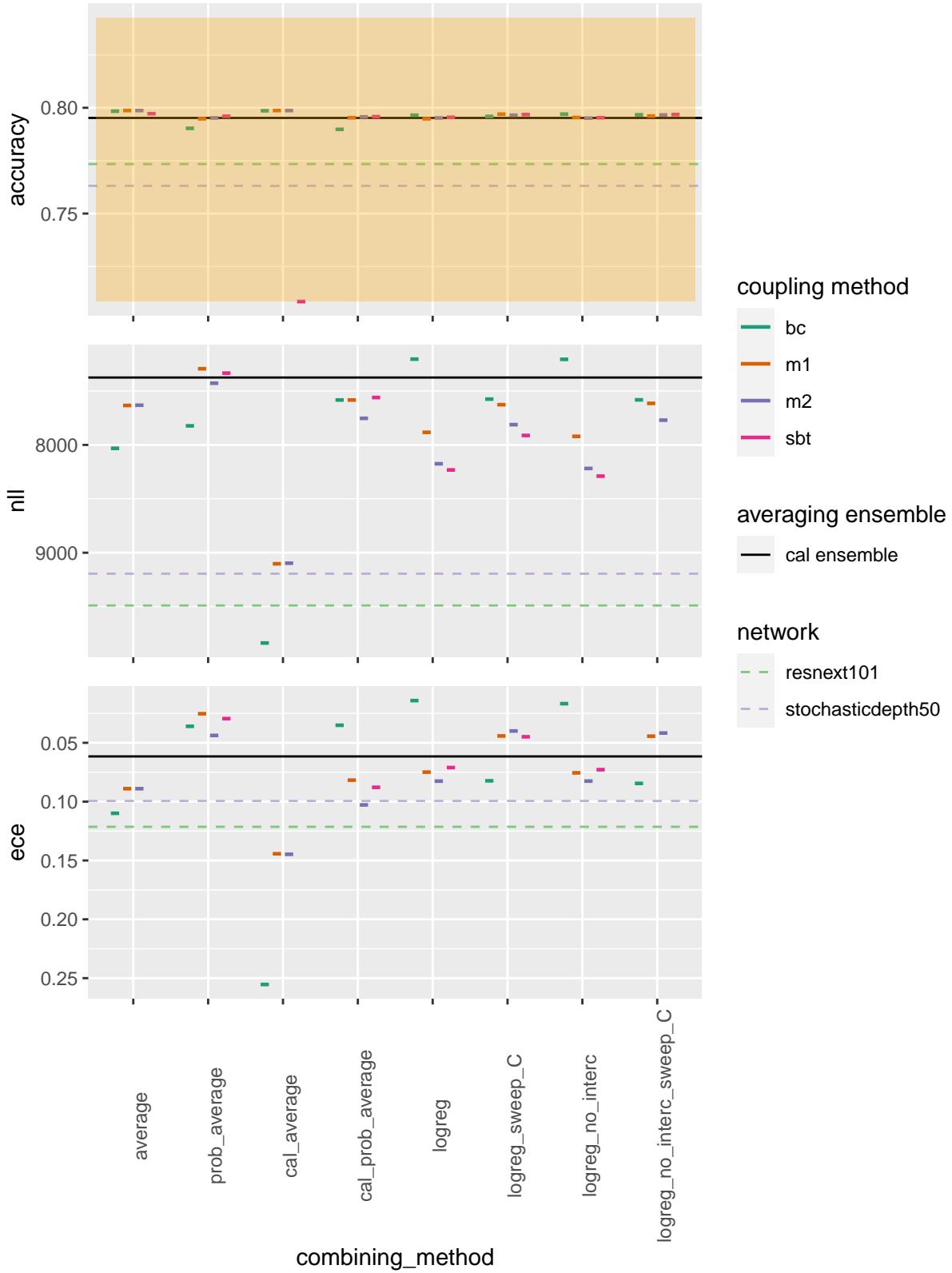
Ensemble metrics  
Error inconsistency 0.252099990844727



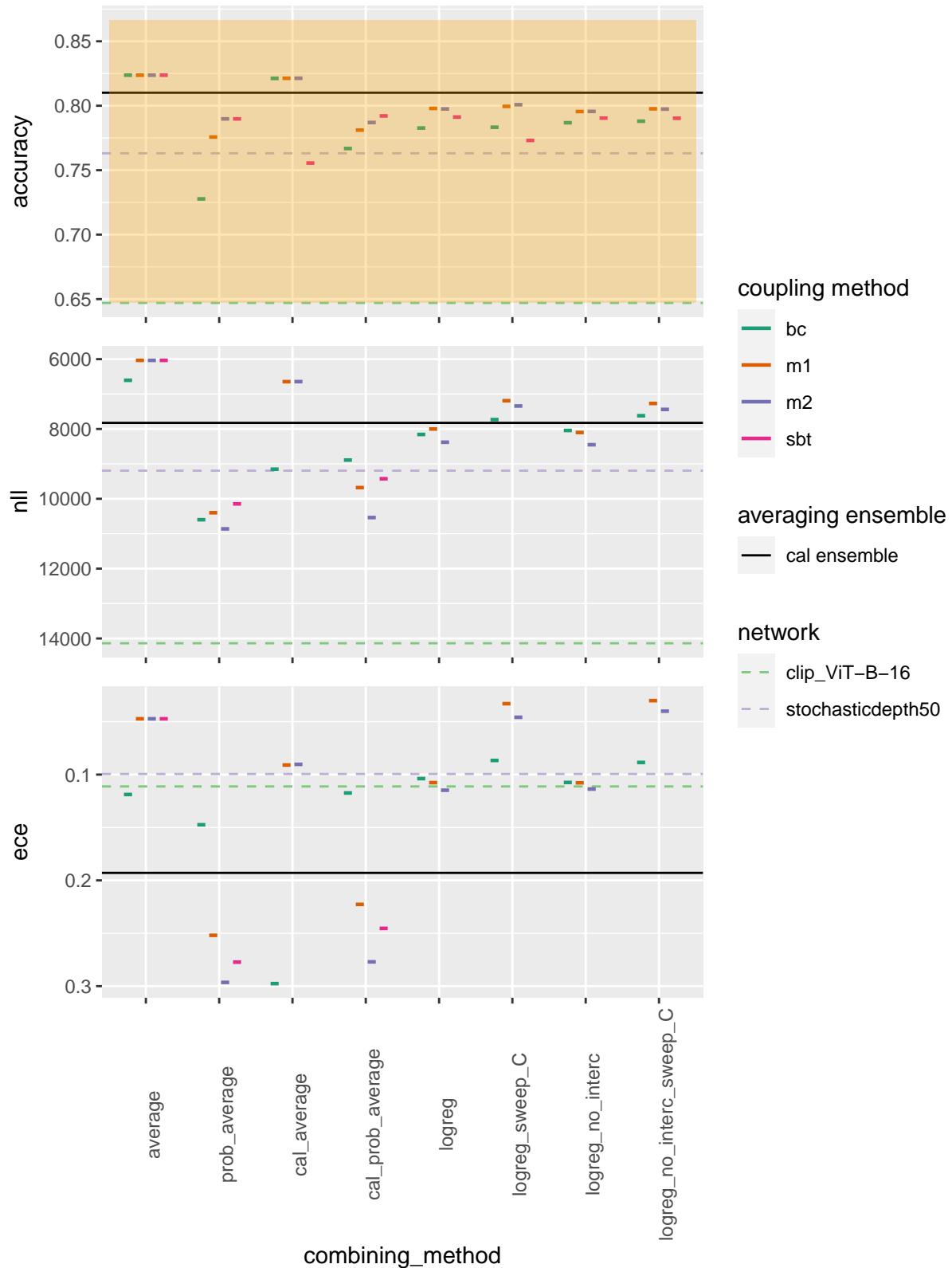
Ensemble metrics  
Error inconsistency 0.210400000214577



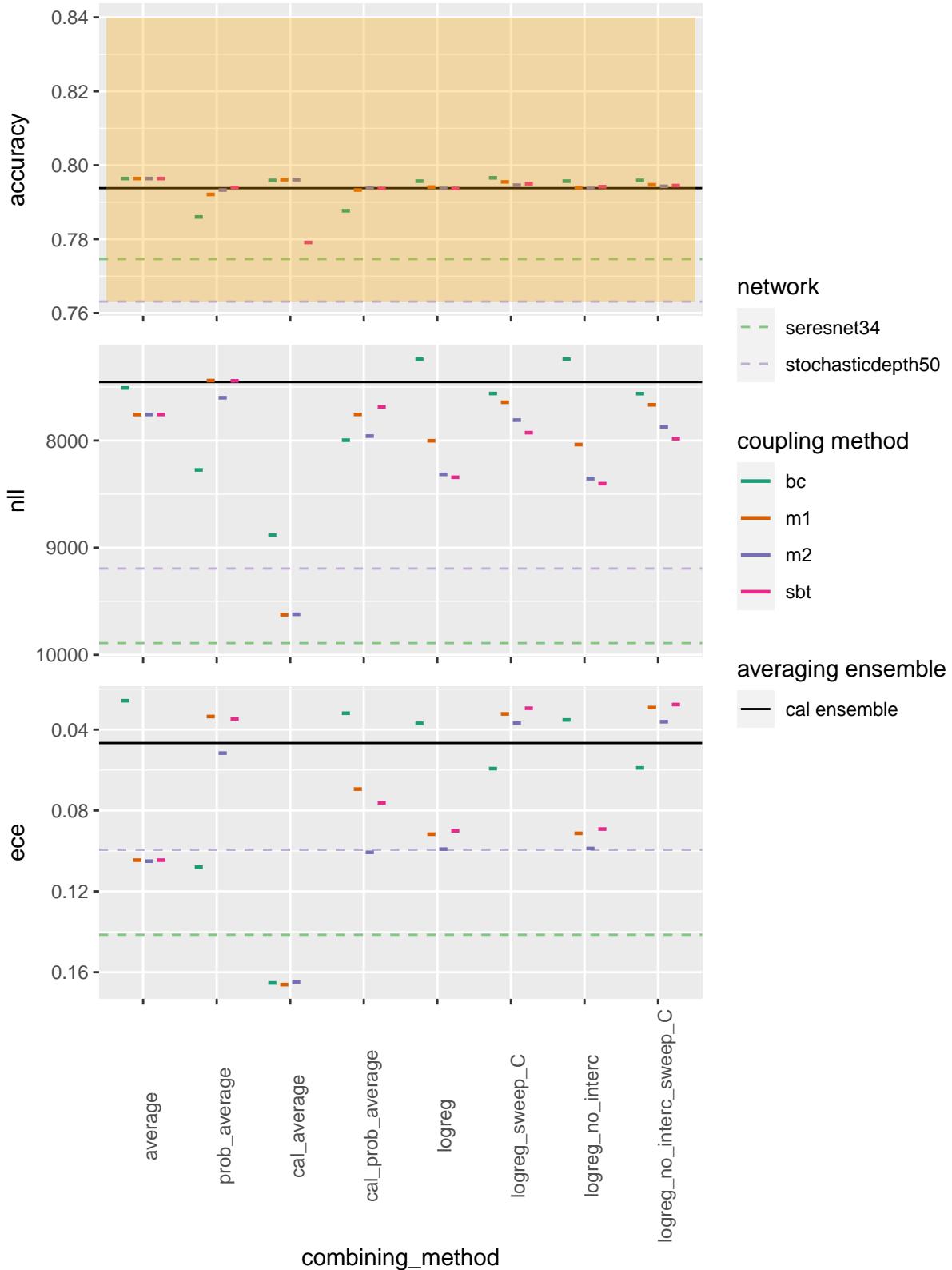
Ensemble metrics  
Error inconsistency 0.148699998855591



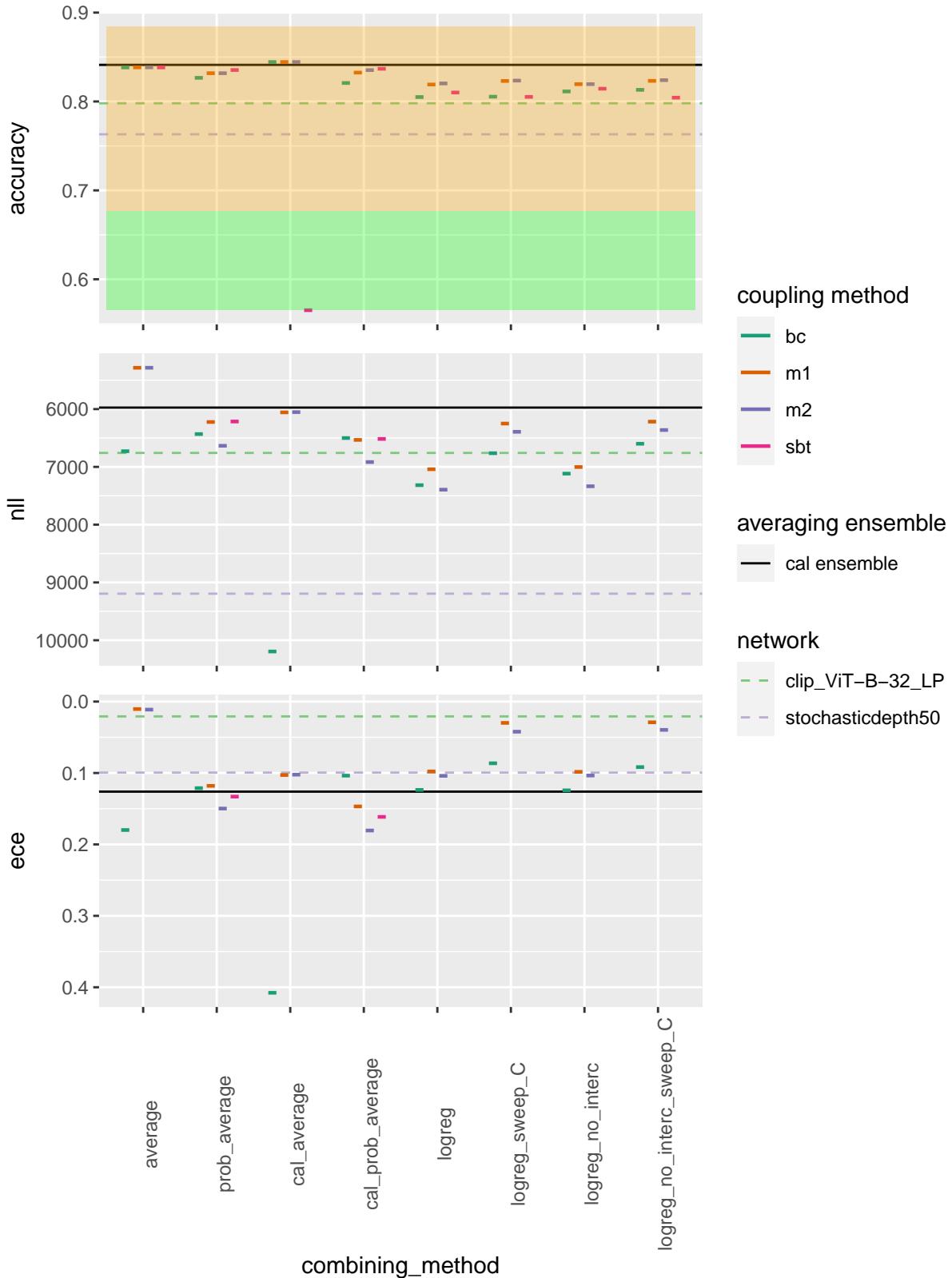
Ensemble metrics  
Error inconsistency 0.322899997234344



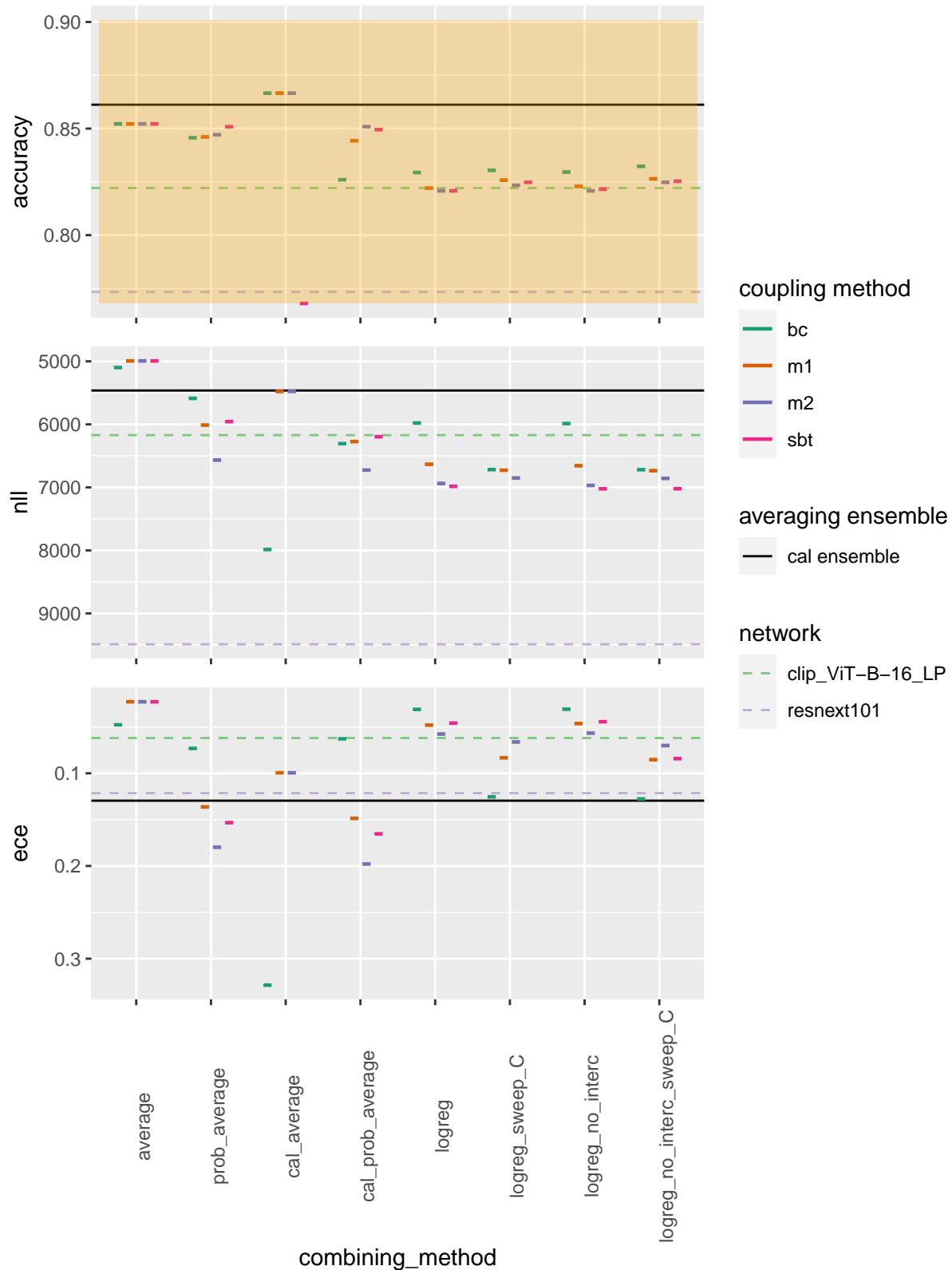
Ensemble metrics  
Error inconsistency 0.142099991440773



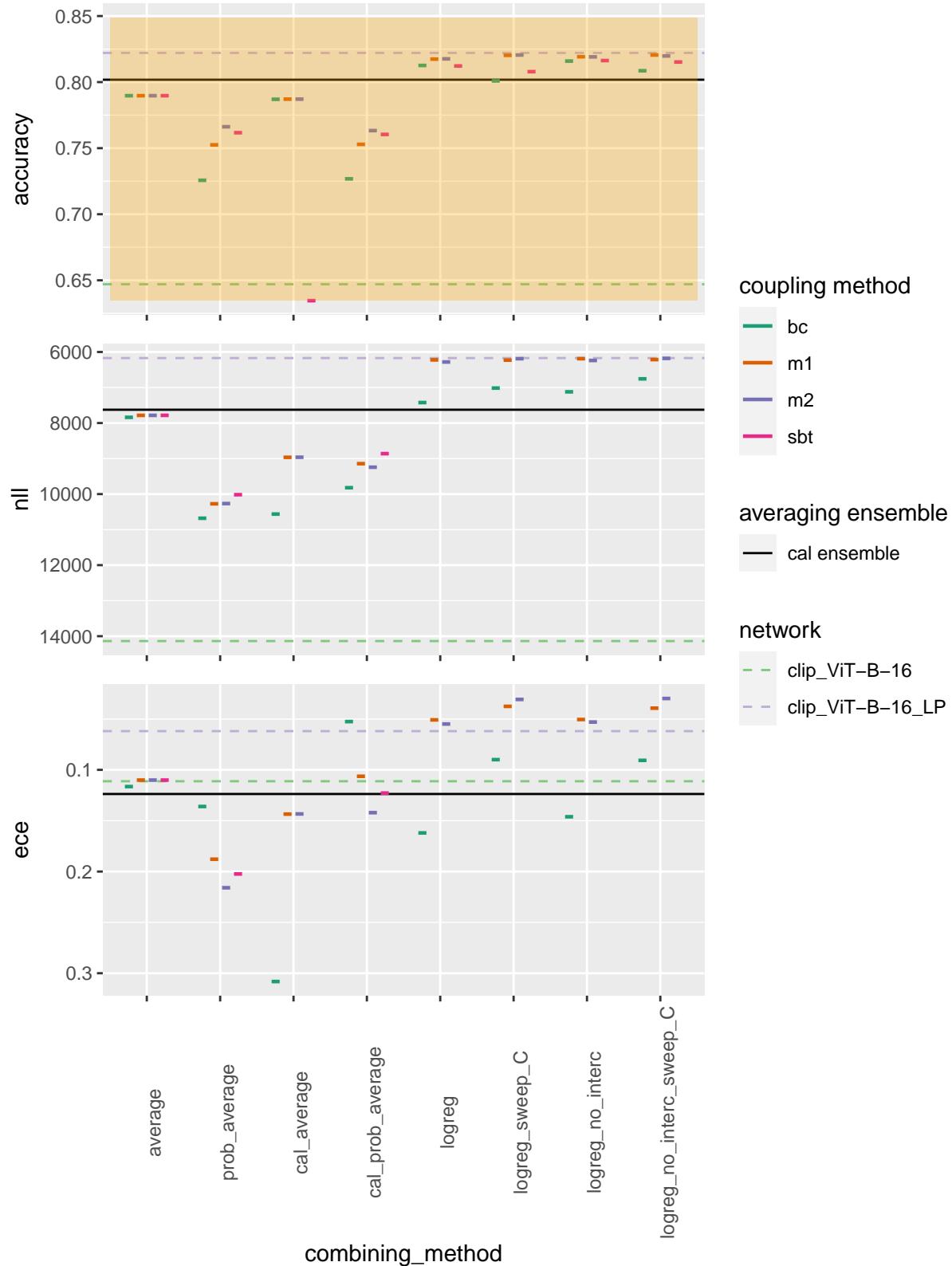
Ensemble metrics  
Error inconsistency 0.207699999213219



Ensemble metrics  
Error inconsistency 0.206299990415573

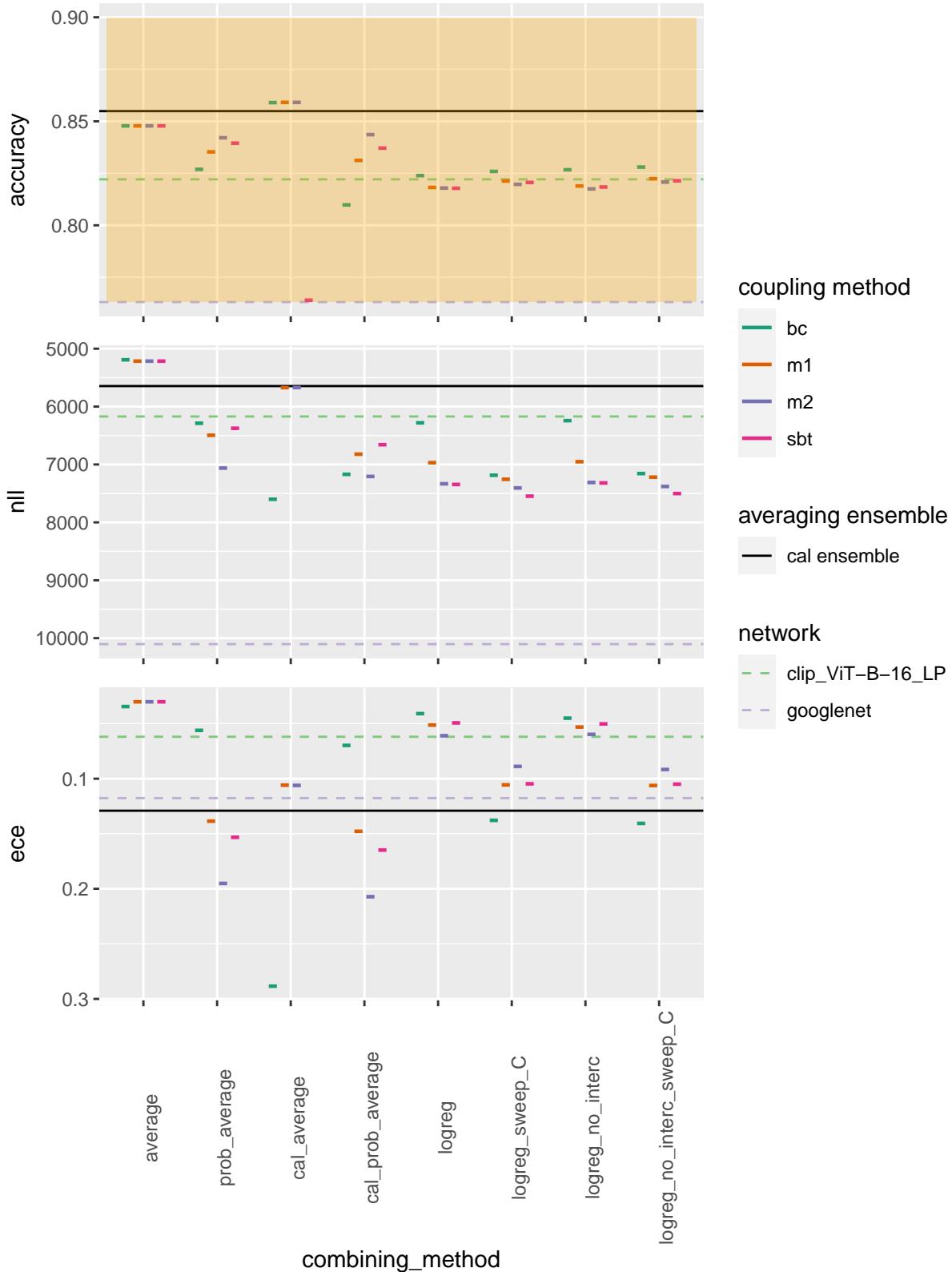


Ensemble metrics  
Error inconsistency 0.228499993681908

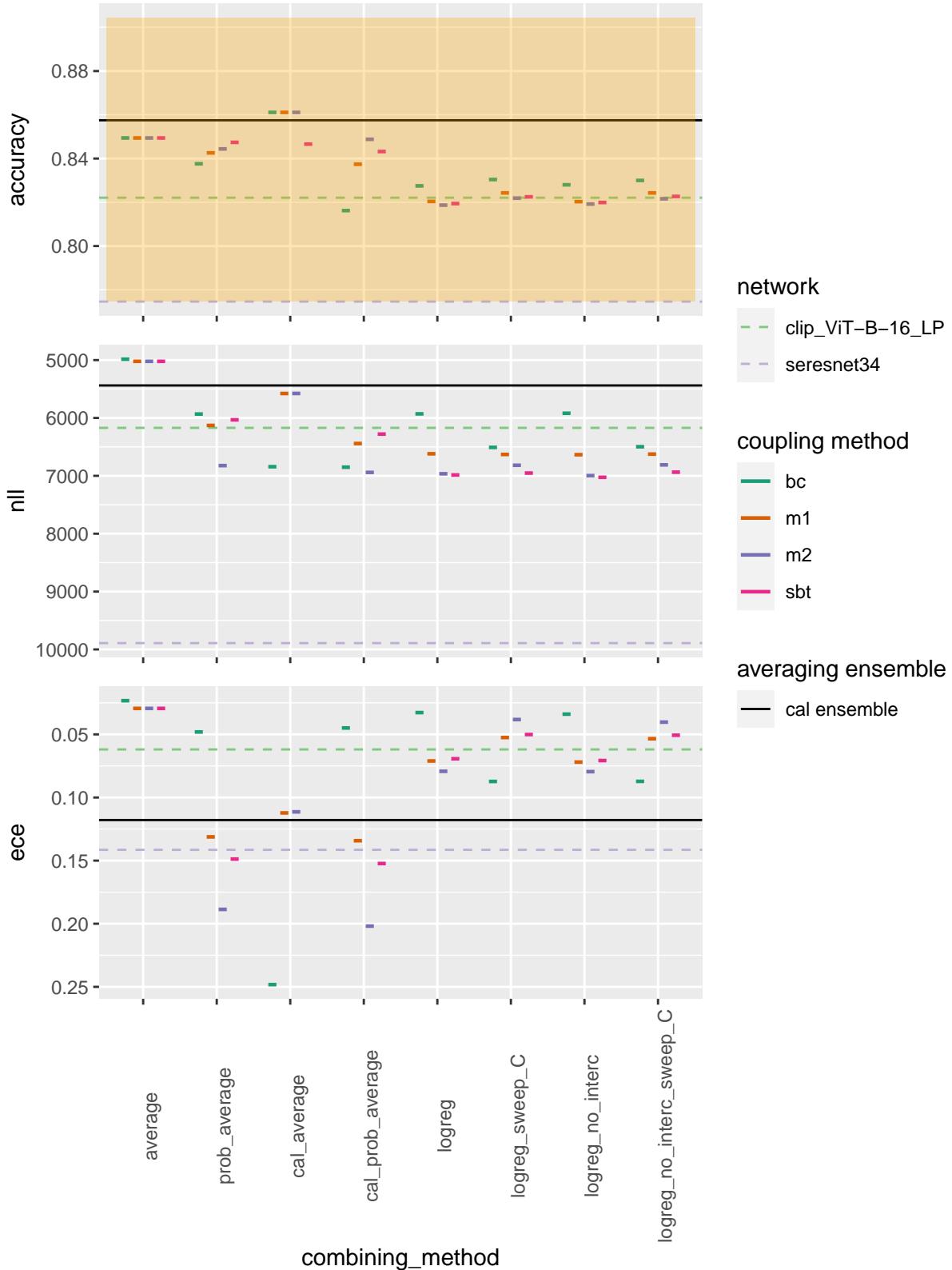


## Ensemble metrics

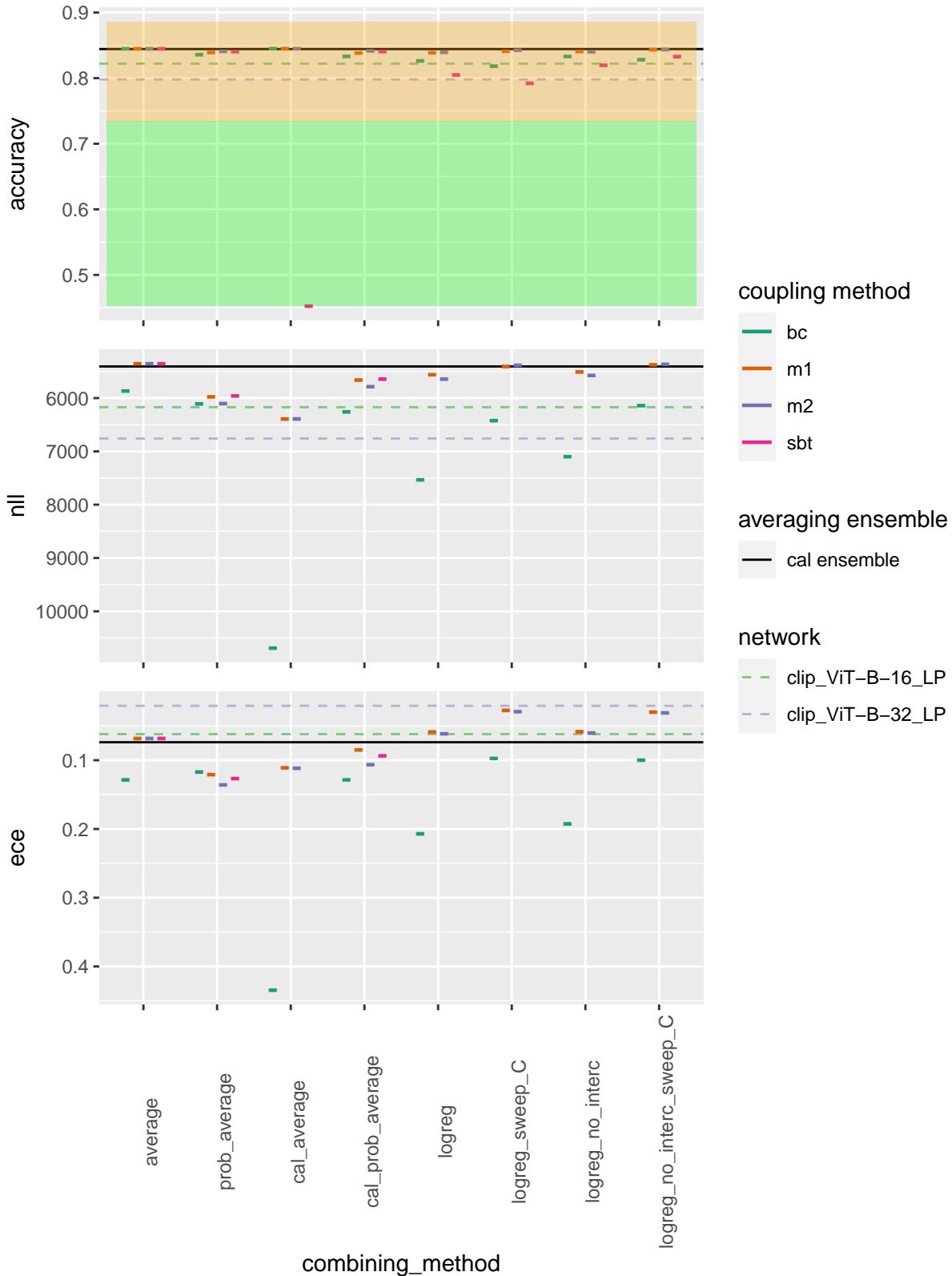
Error inconsistency 0.214399993419647



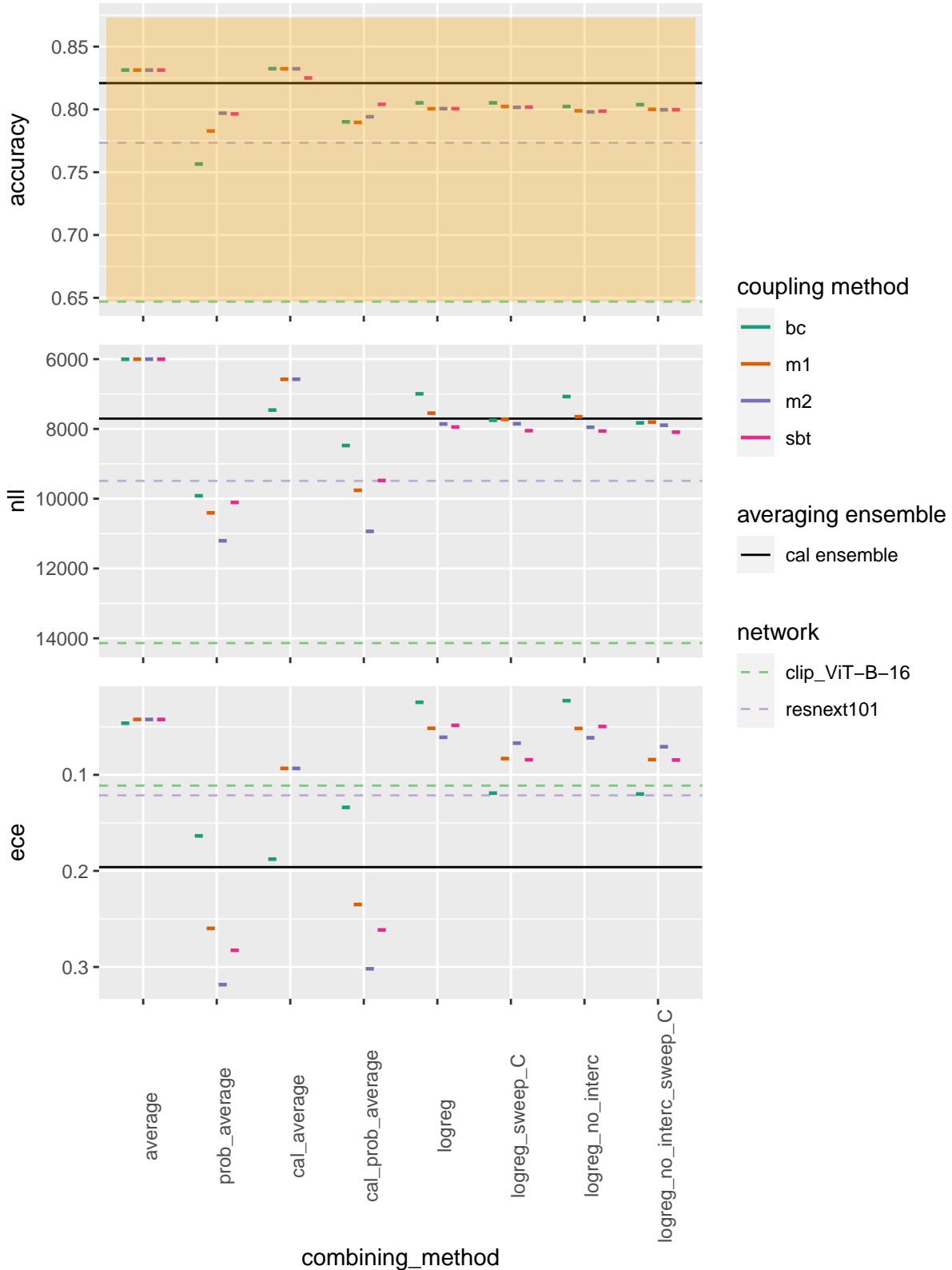
Ensemble metrics  
Error inconsistency 0.212099999189377



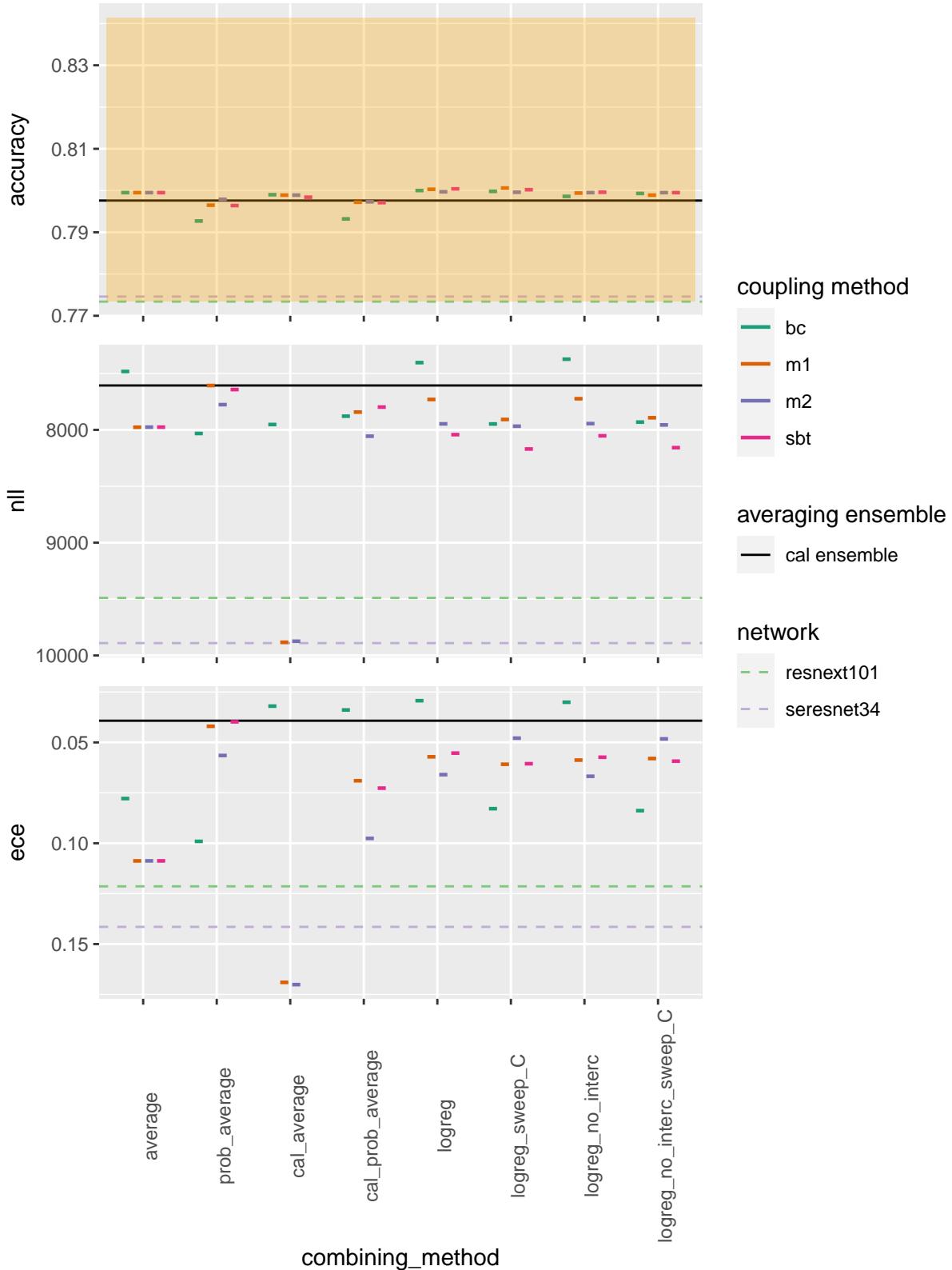
Ensemble metrics  
Error inconsistency 0.151899993419647



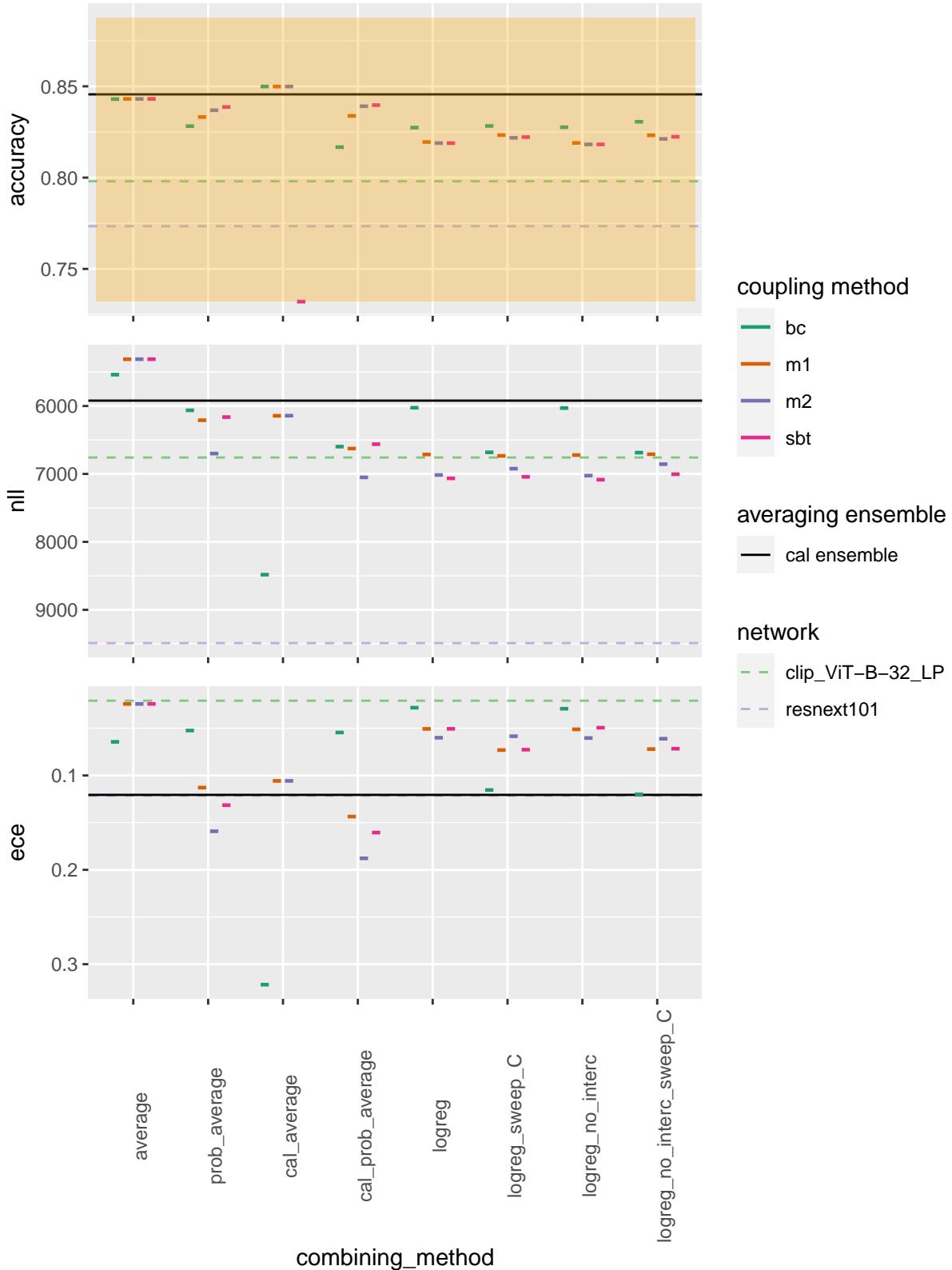
Ensemble metrics  
Error inconsistency 0.325599998235703



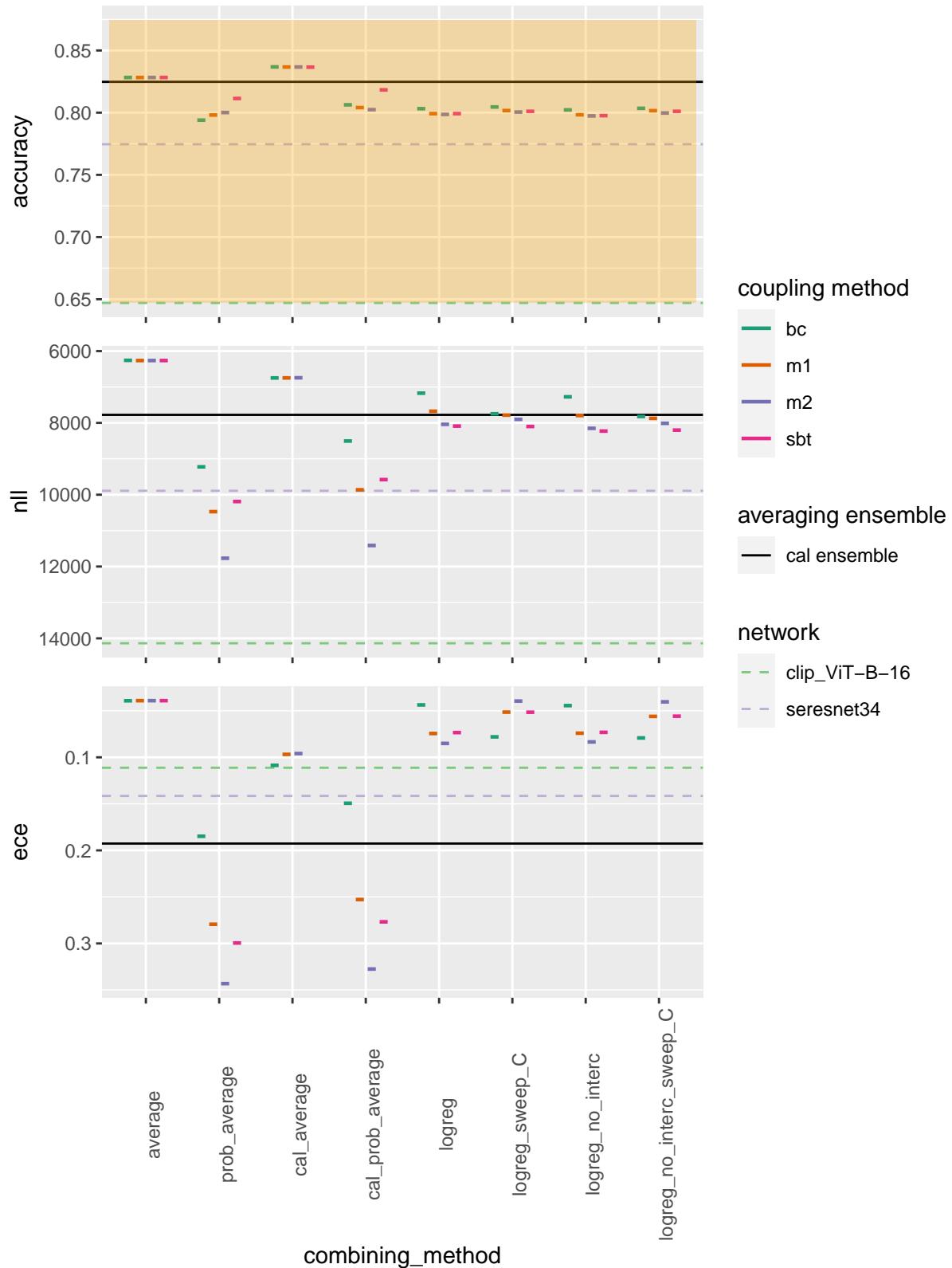
Ensemble metrics  
Error inconsistency 0.134800001978874



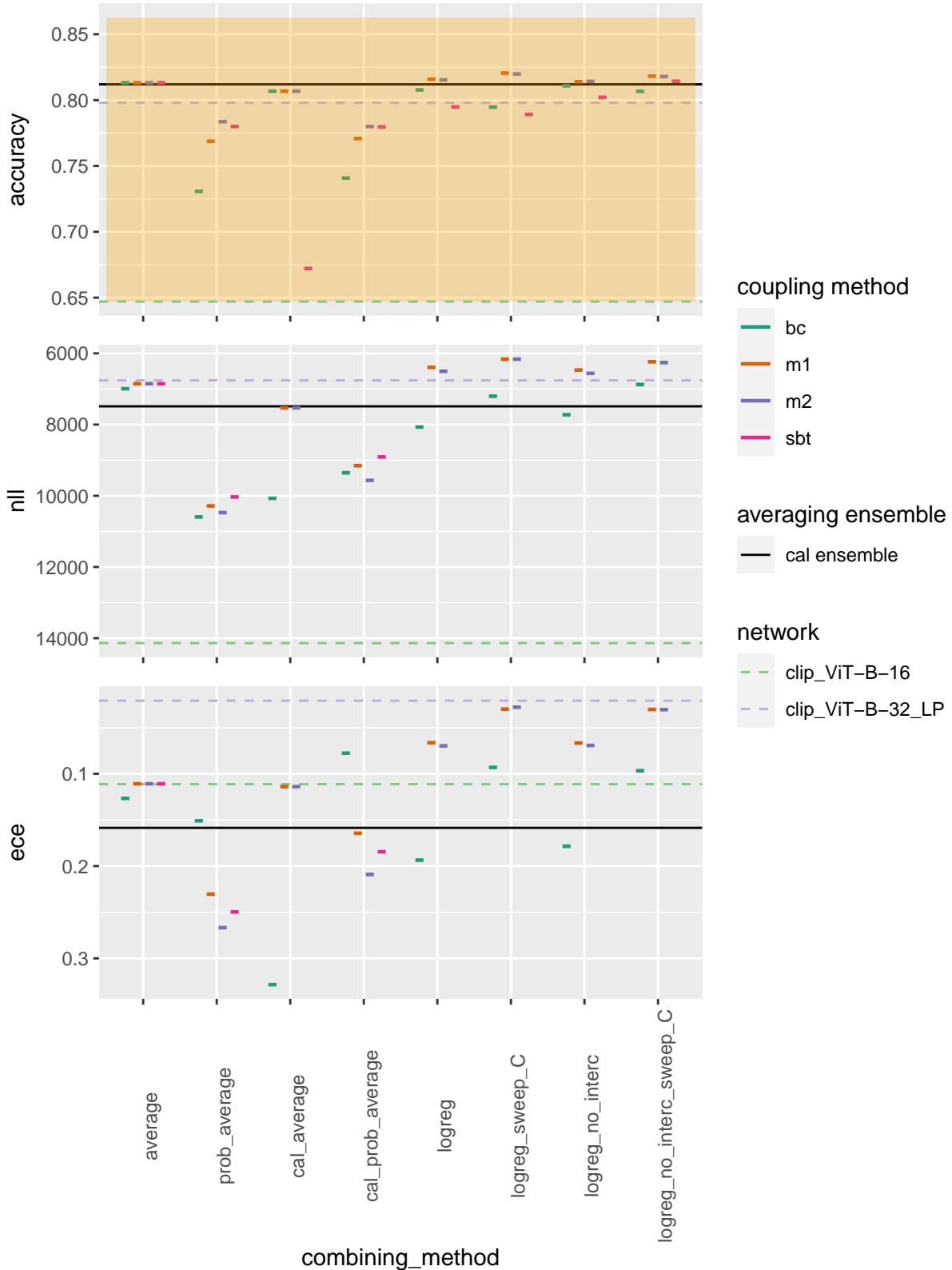
Ensemble metrics  
Error inconsistency 0.203799992799759



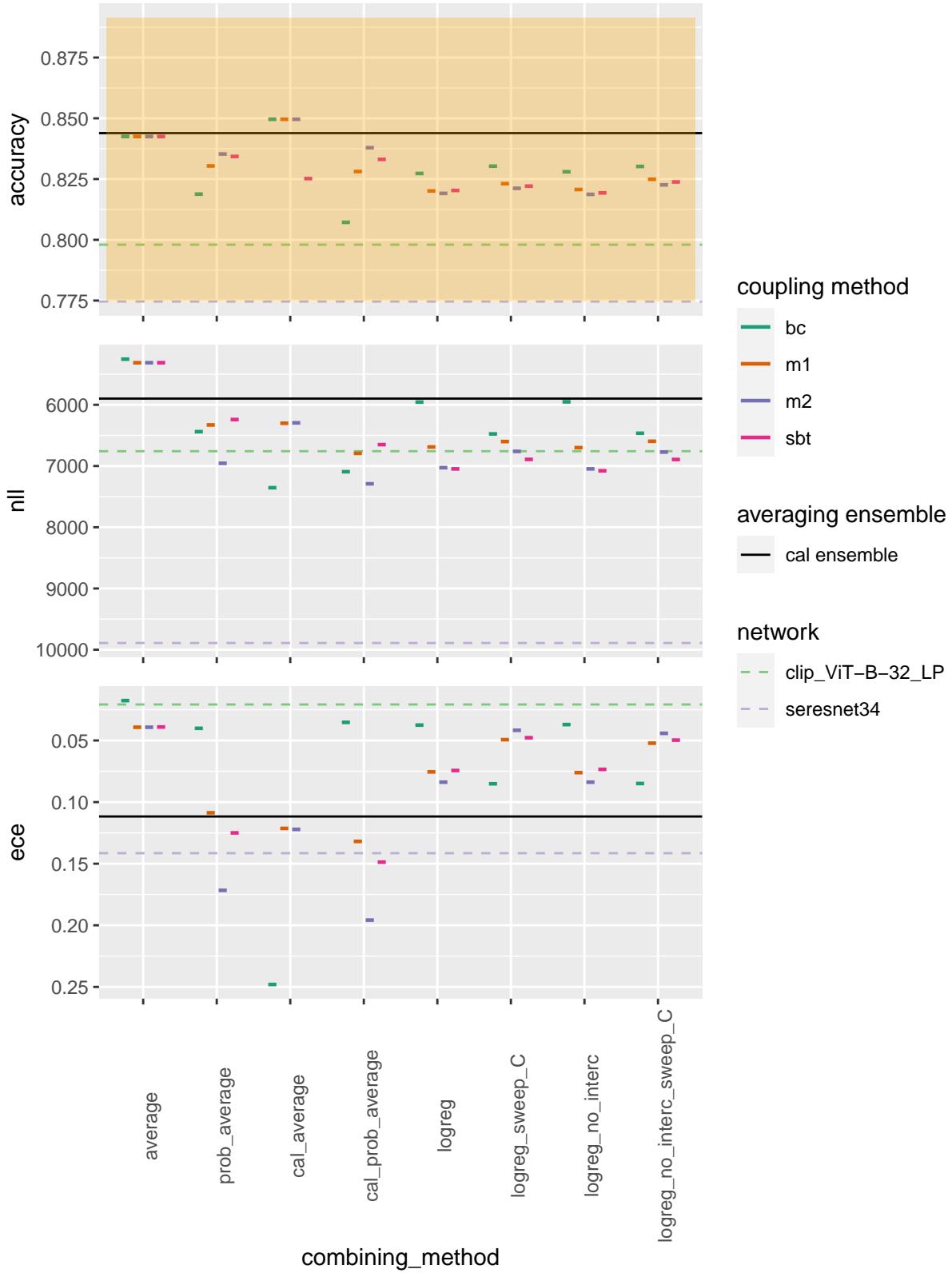
Ensemble metrics  
Error inconsistency 0.327399998903274



Ensemble metrics  
Error inconsistency 0.280200004577637

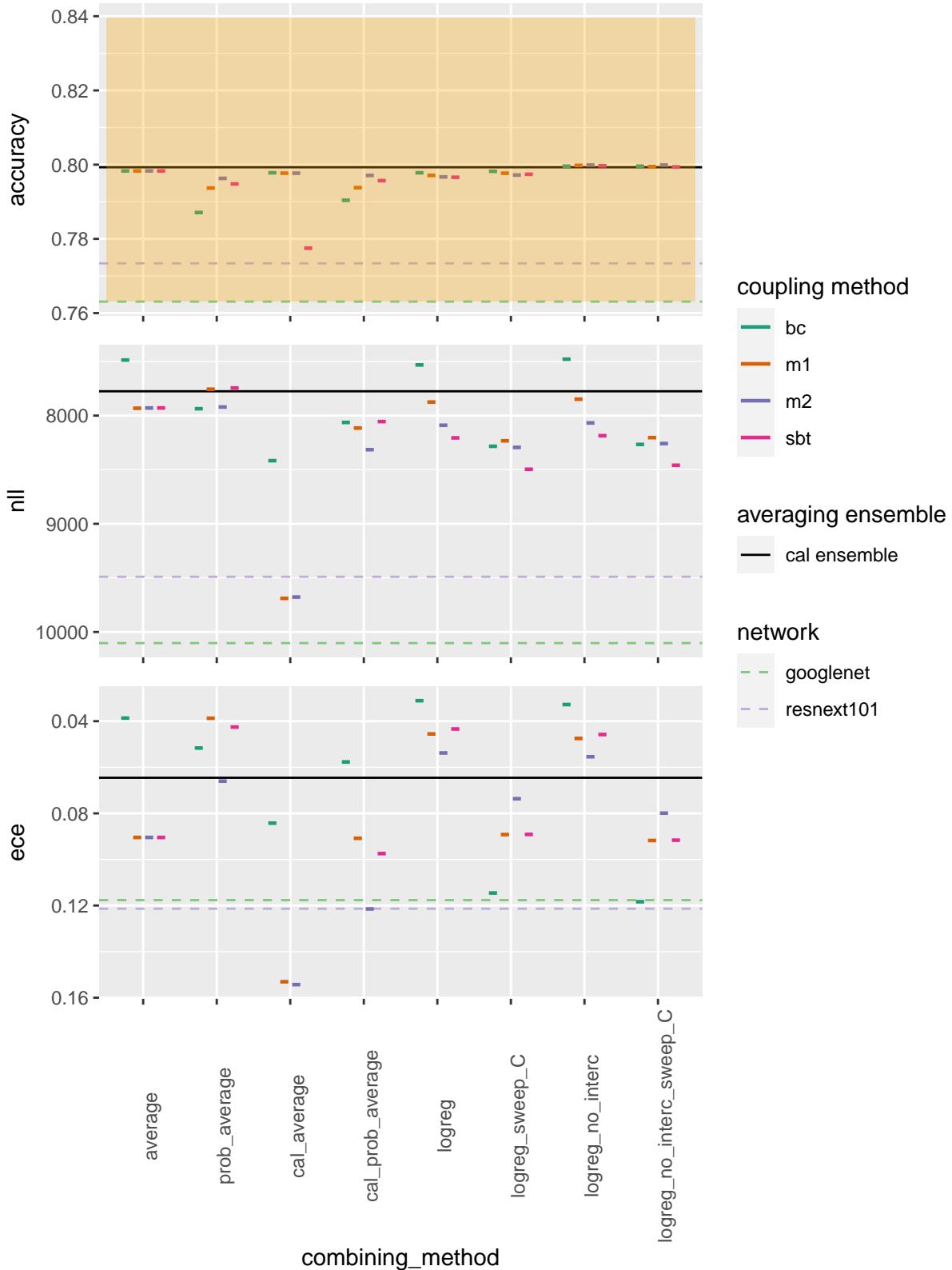


Ensemble metrics  
Error inconsistency 0.210199996829033

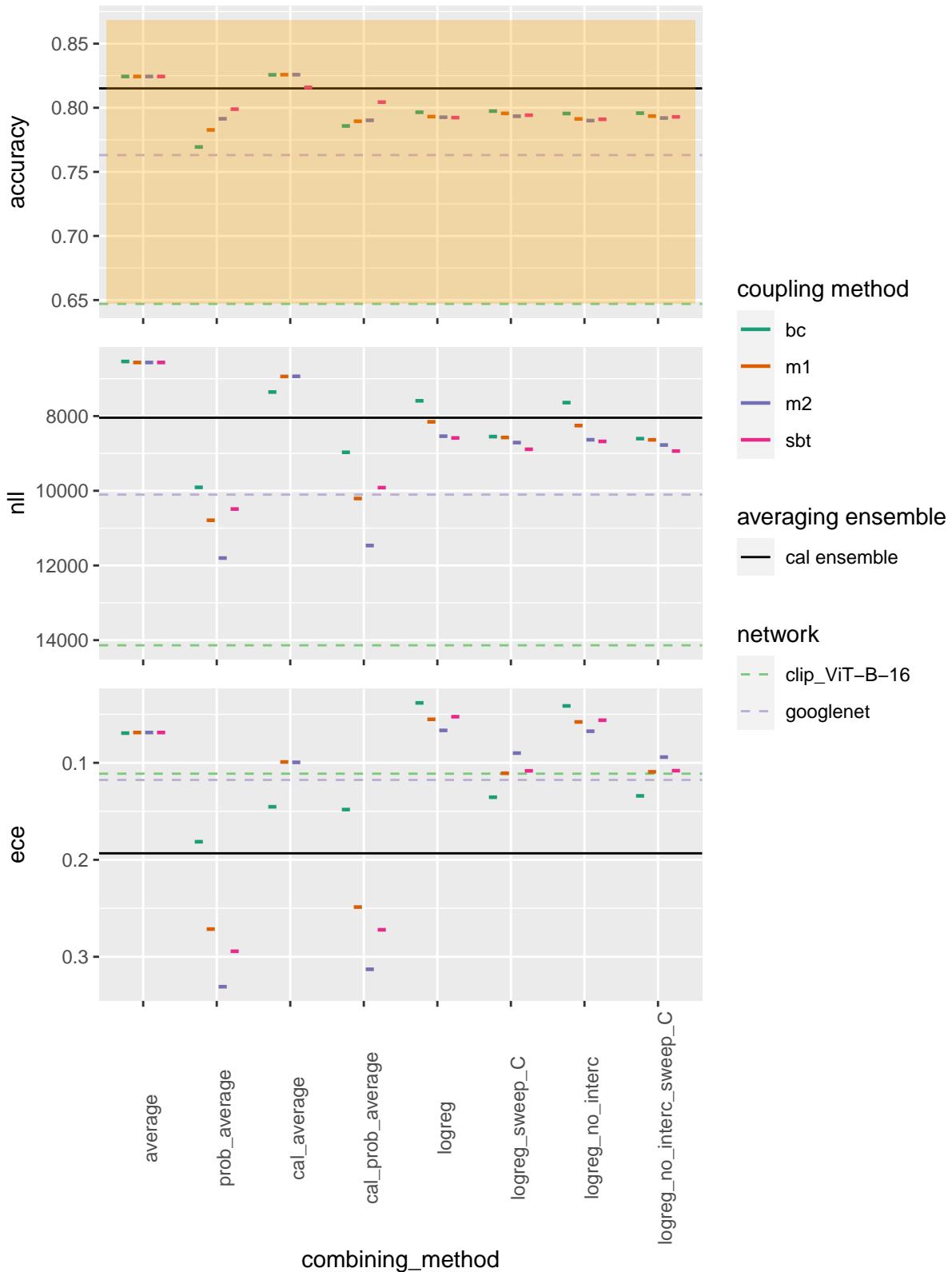


## Ensemble metrics

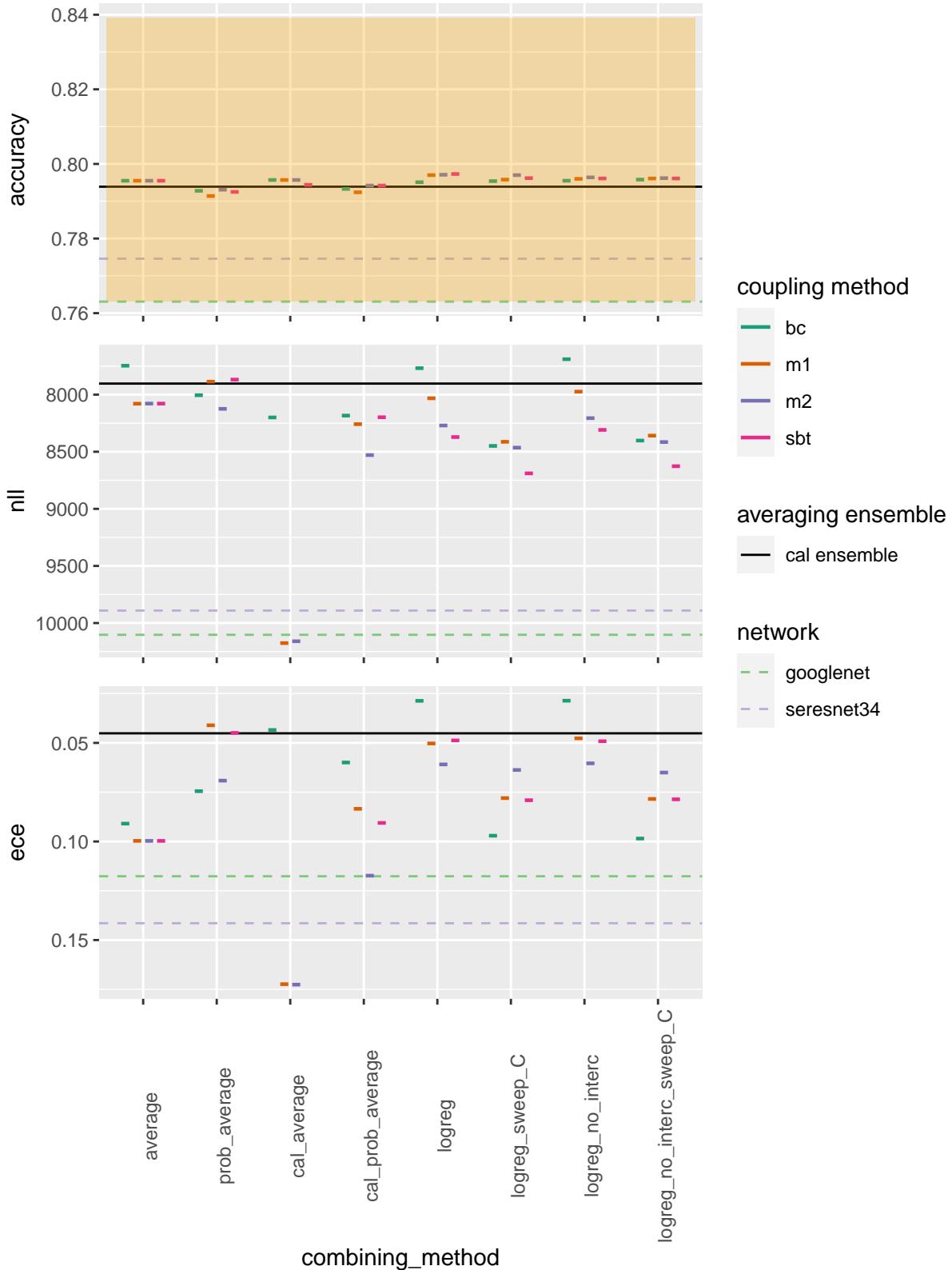
Error inconsistency 0.142700001597404



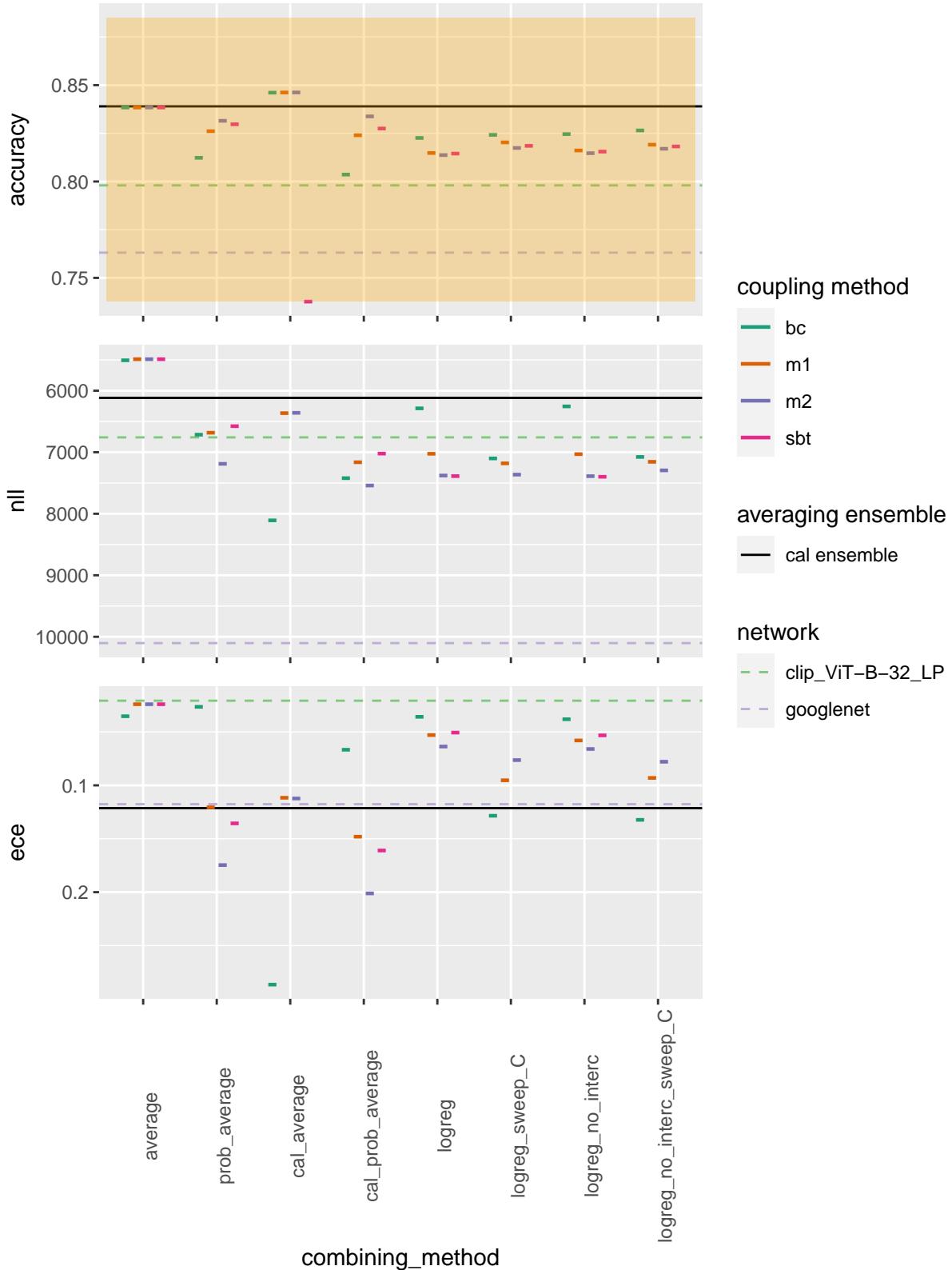
Ensemble metrics  
Error inconsistency 0.326700001955032



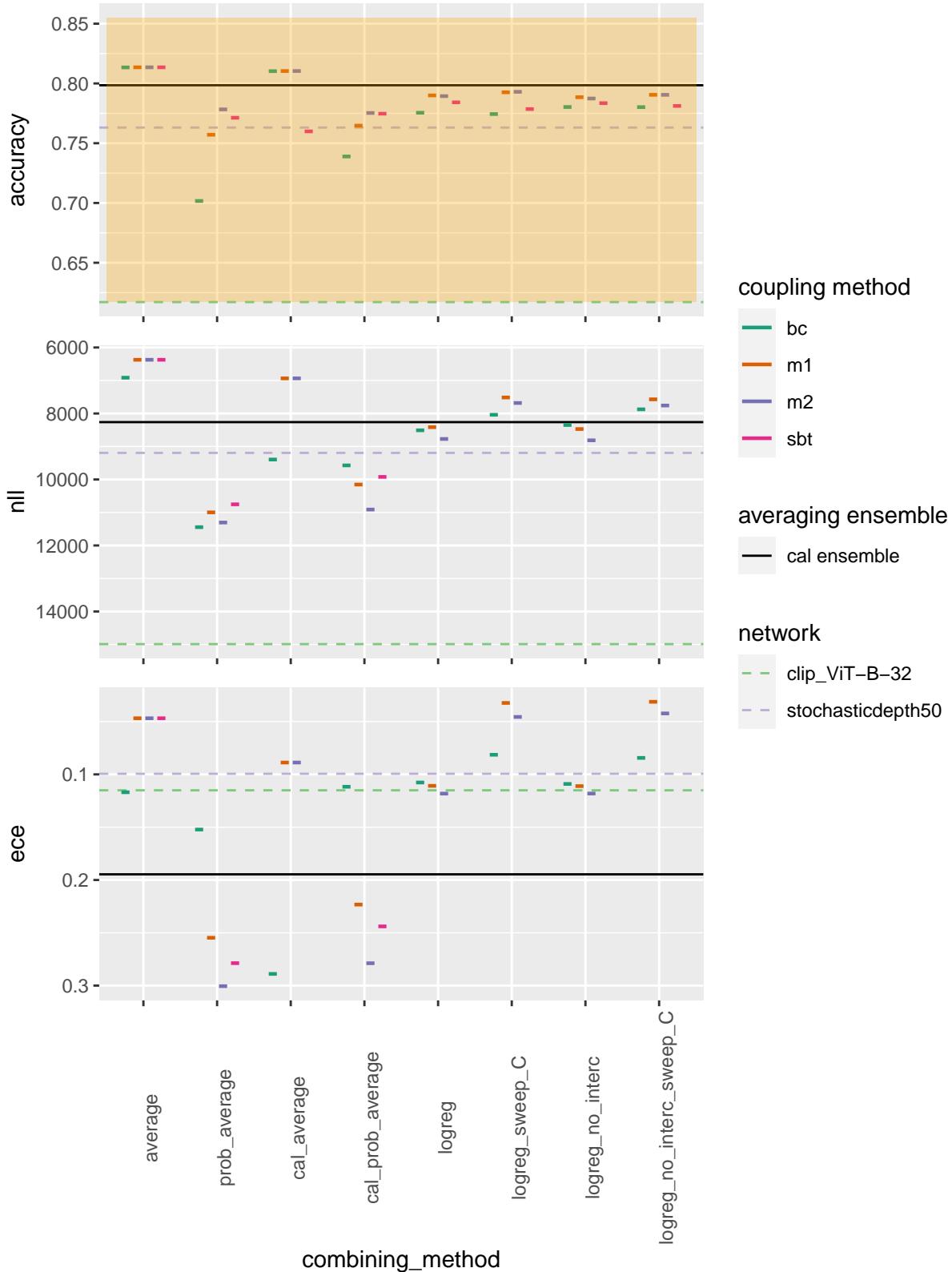
Ensemble metrics  
Error inconsistency 0.140699997544289



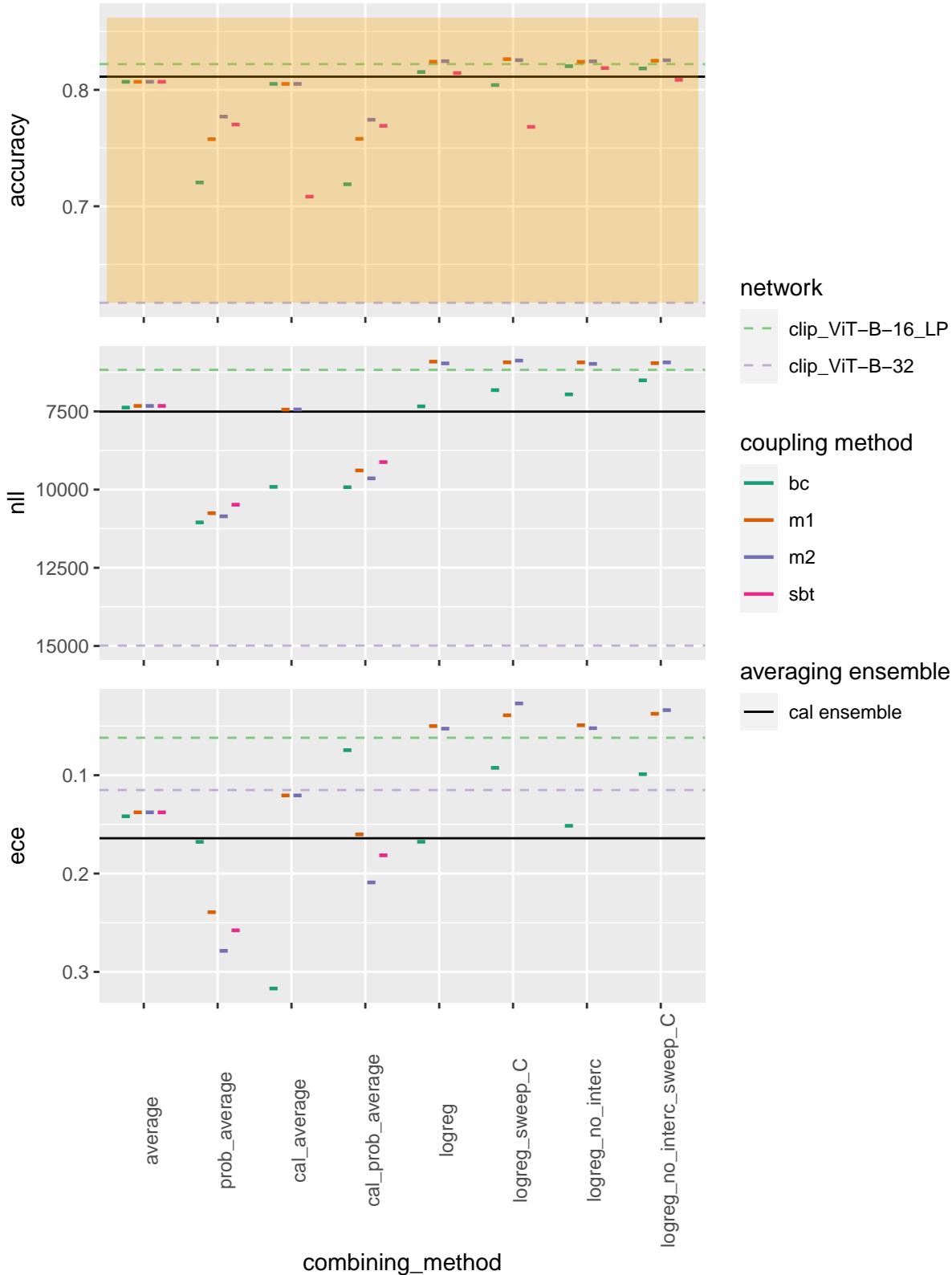
Ensemble metrics  
Error inconsistency 0.208899989724159



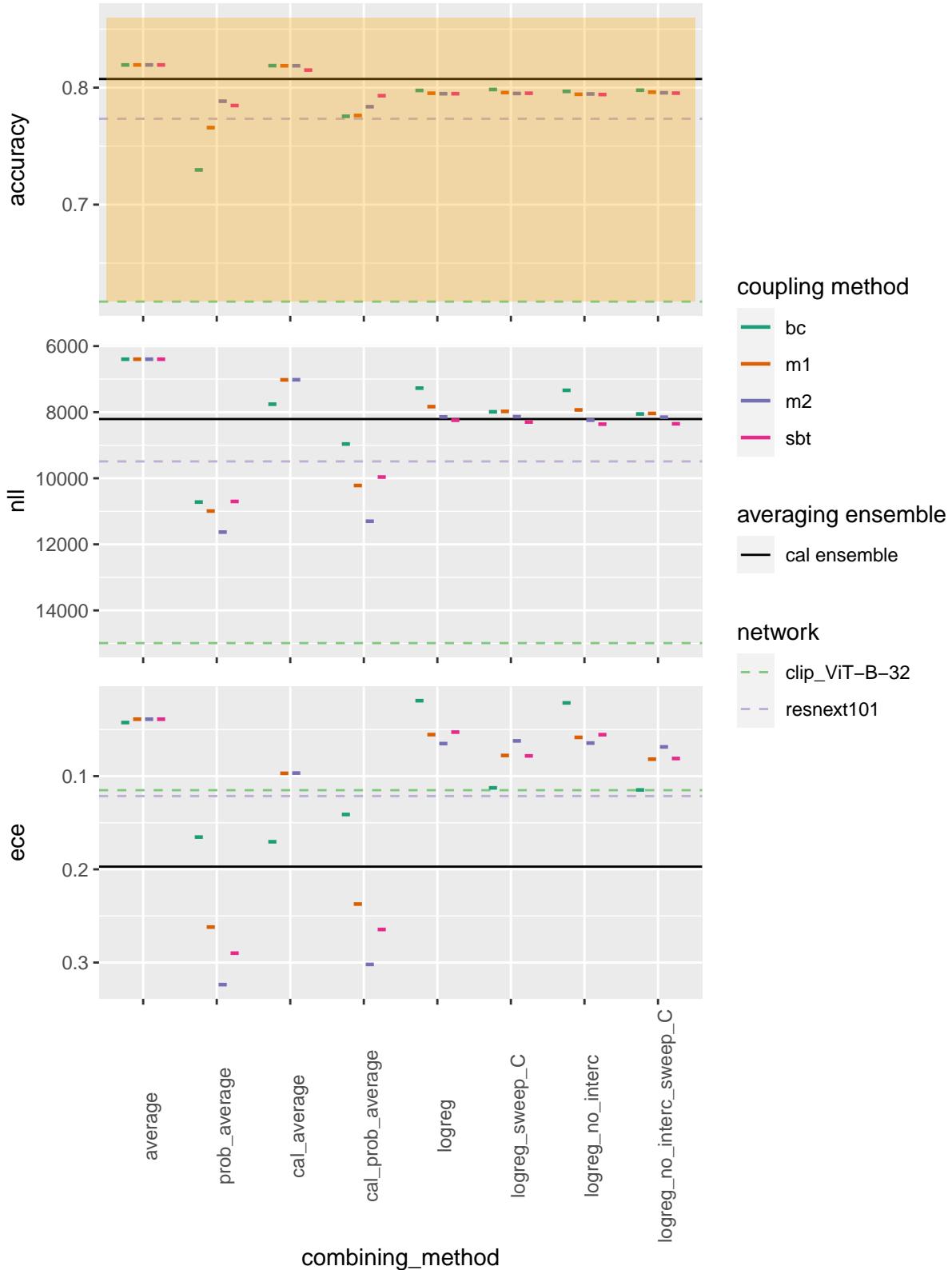
Ensemble metrics  
Error inconsistency 0.329799979925156



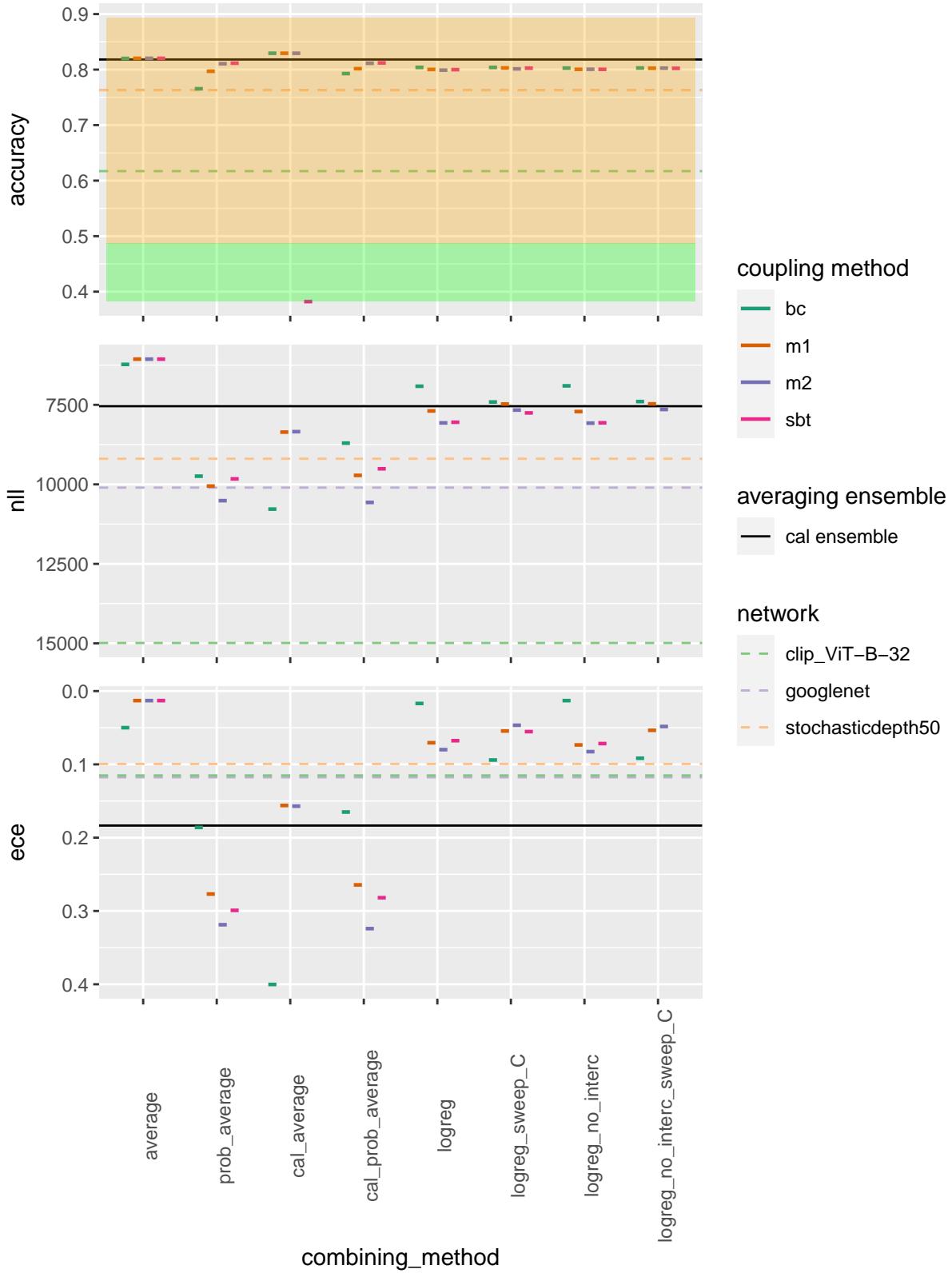
Ensemble metrics  
Error inconsistency 0.284599989652634



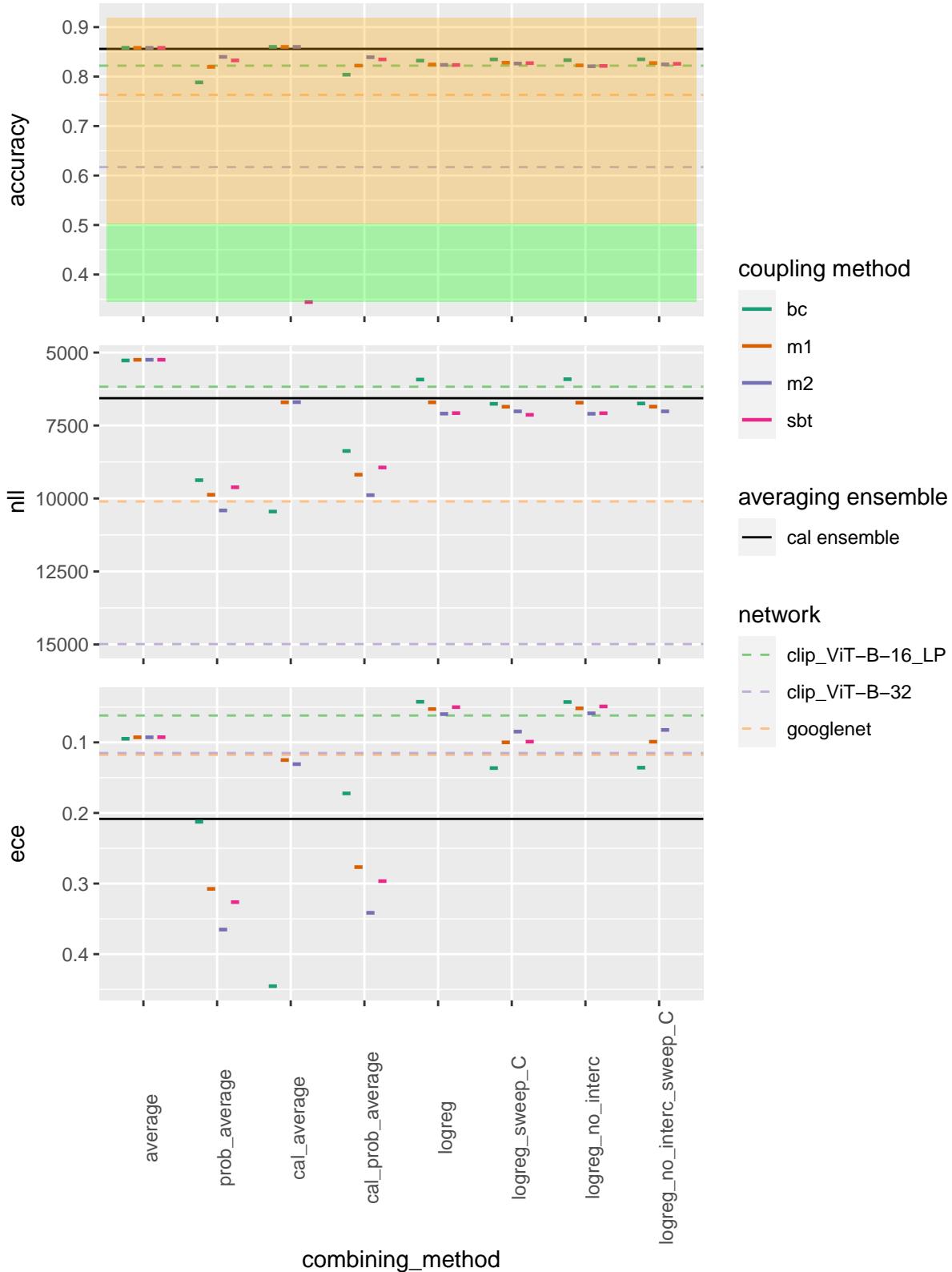
Ensemble metrics  
Error inconsistency 0.329099982976913



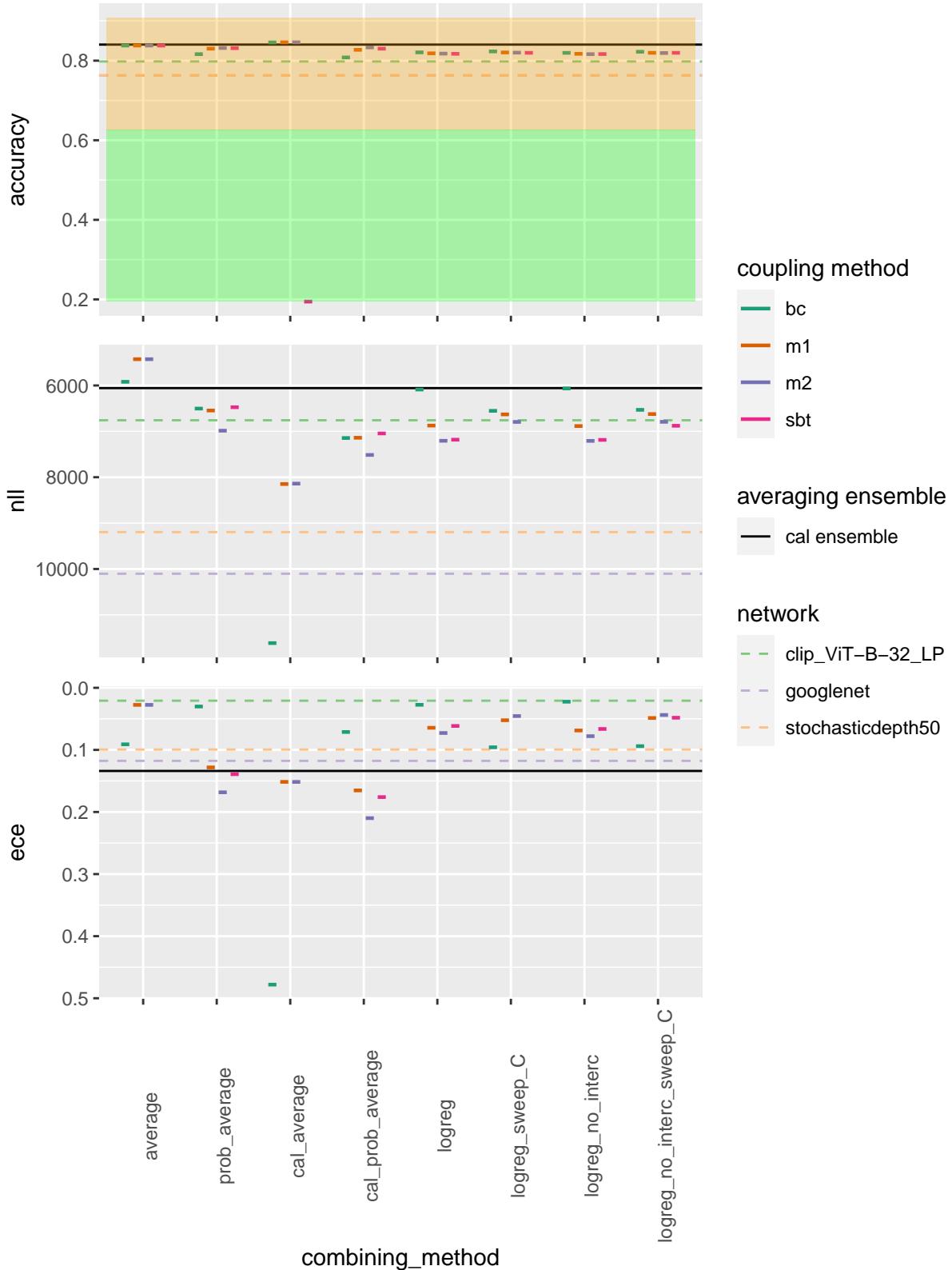
Ensemble metrics  
Error inconsistency 0.406499981880188



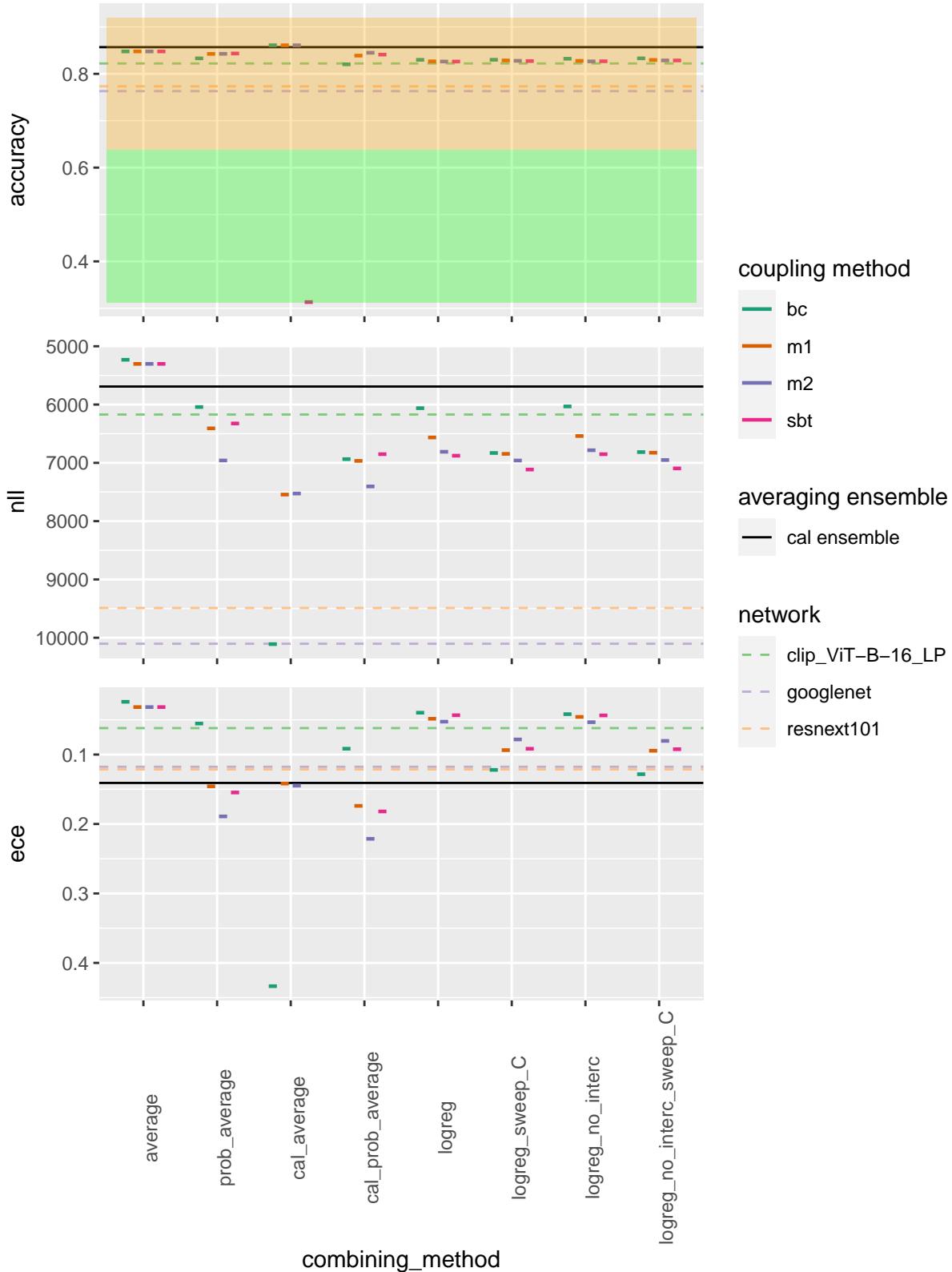
Ensemble metrics  
Error inconsistency 0.416699975728989



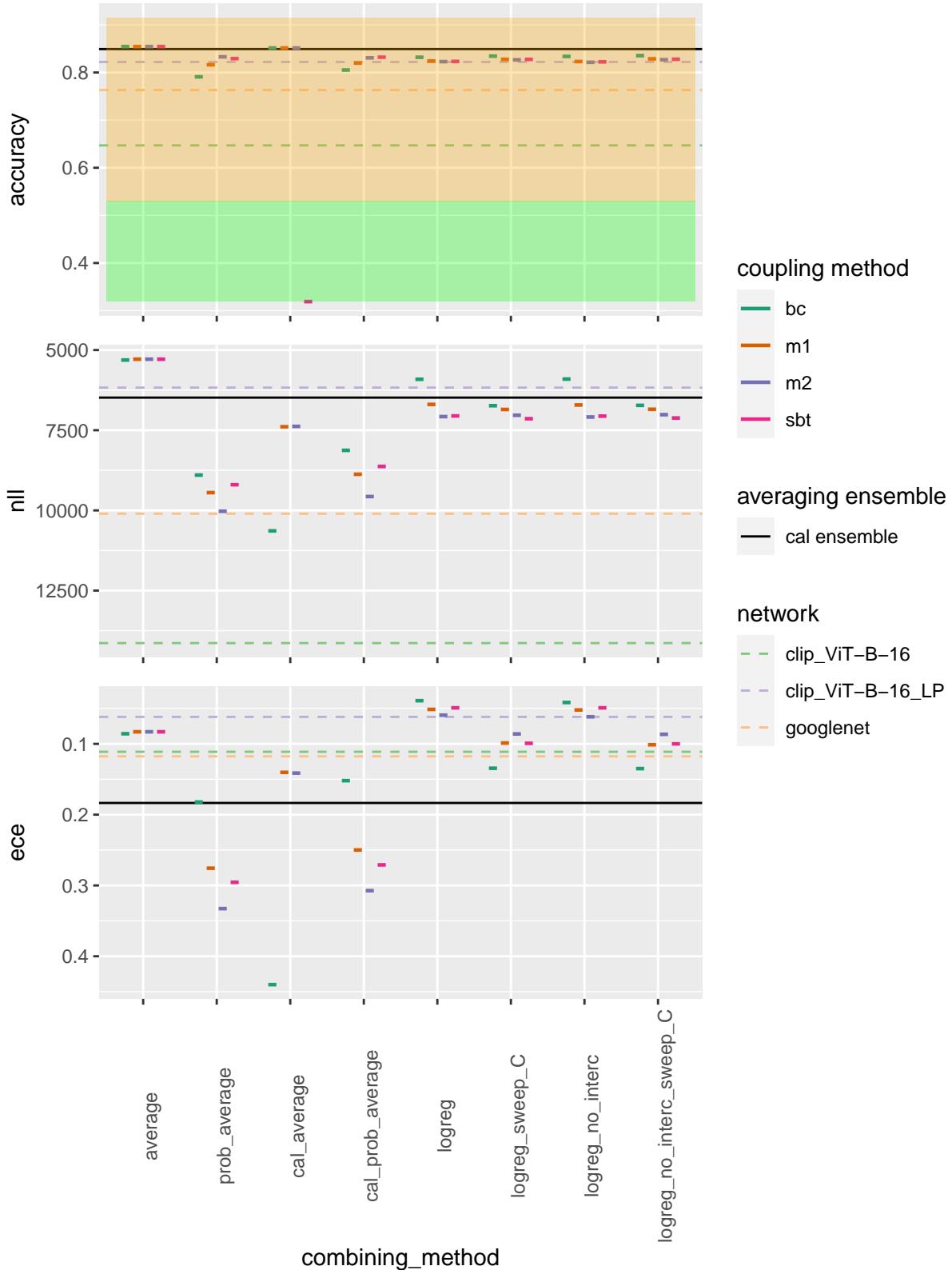
Ensemble metrics  
Error inconsistency 0.282700002193451



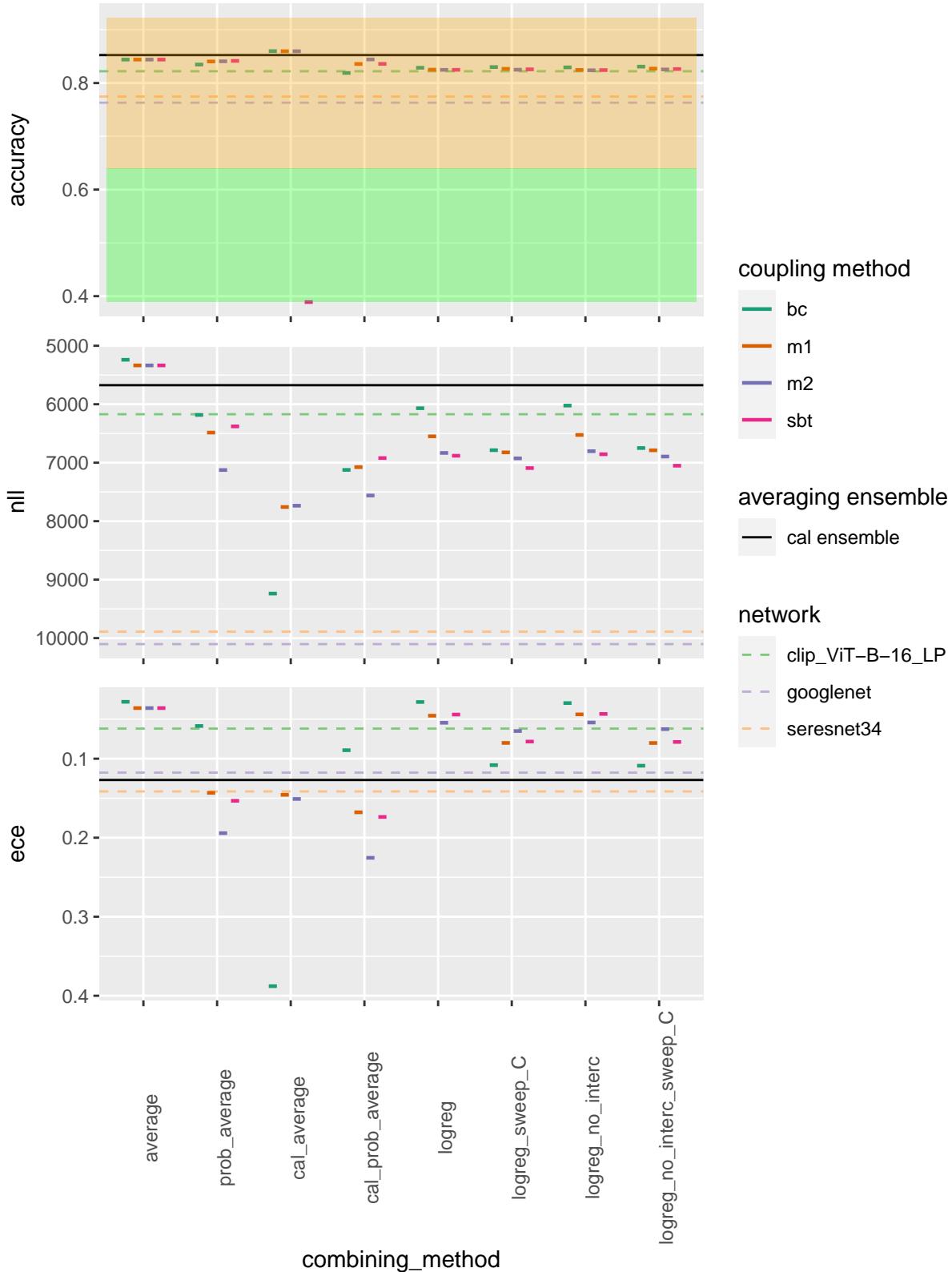
Ensemble metrics  
Error inconsistency 0.281699985265732



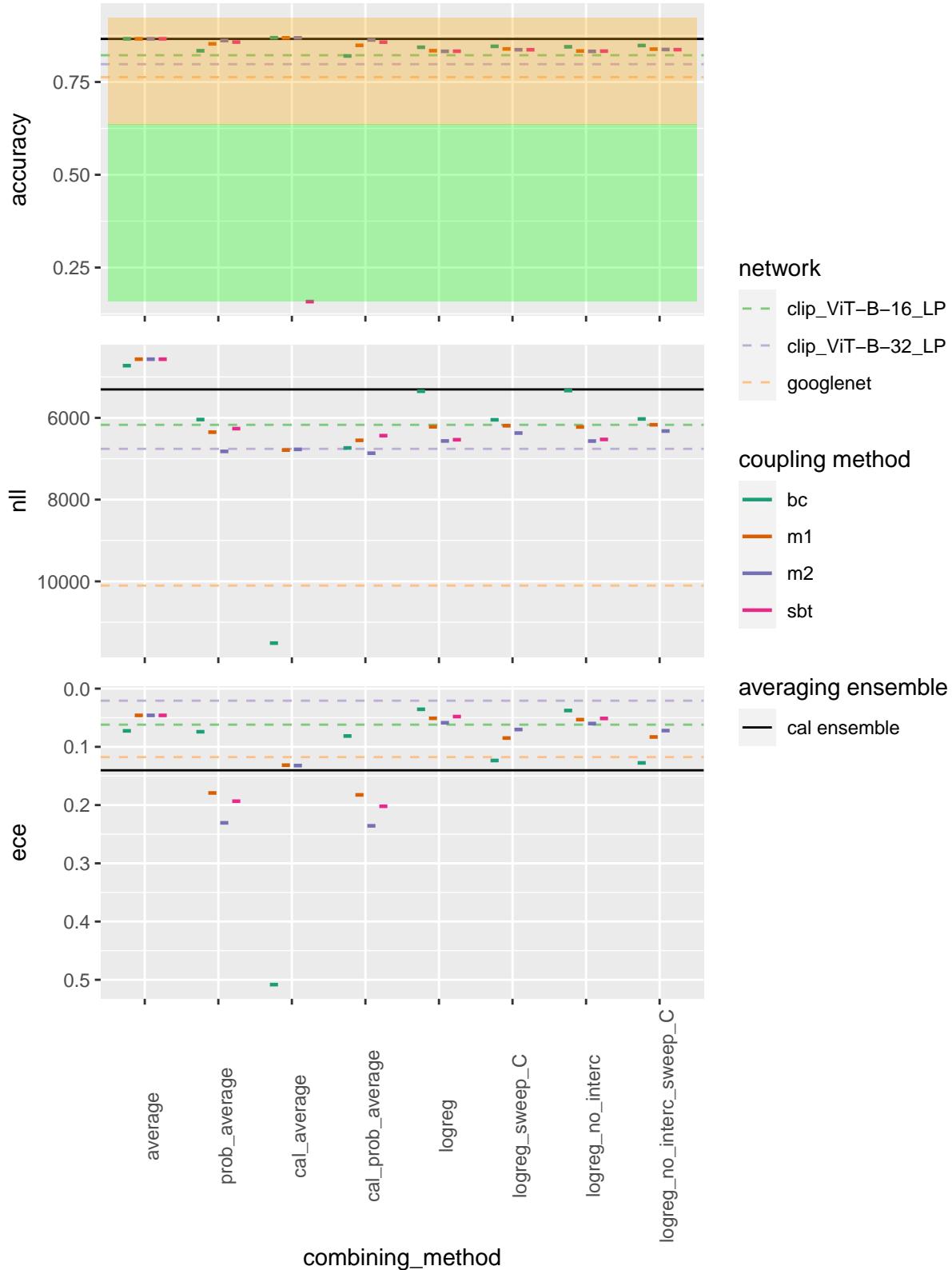
Ensemble metrics  
Error inconsistency 0.384799987077713



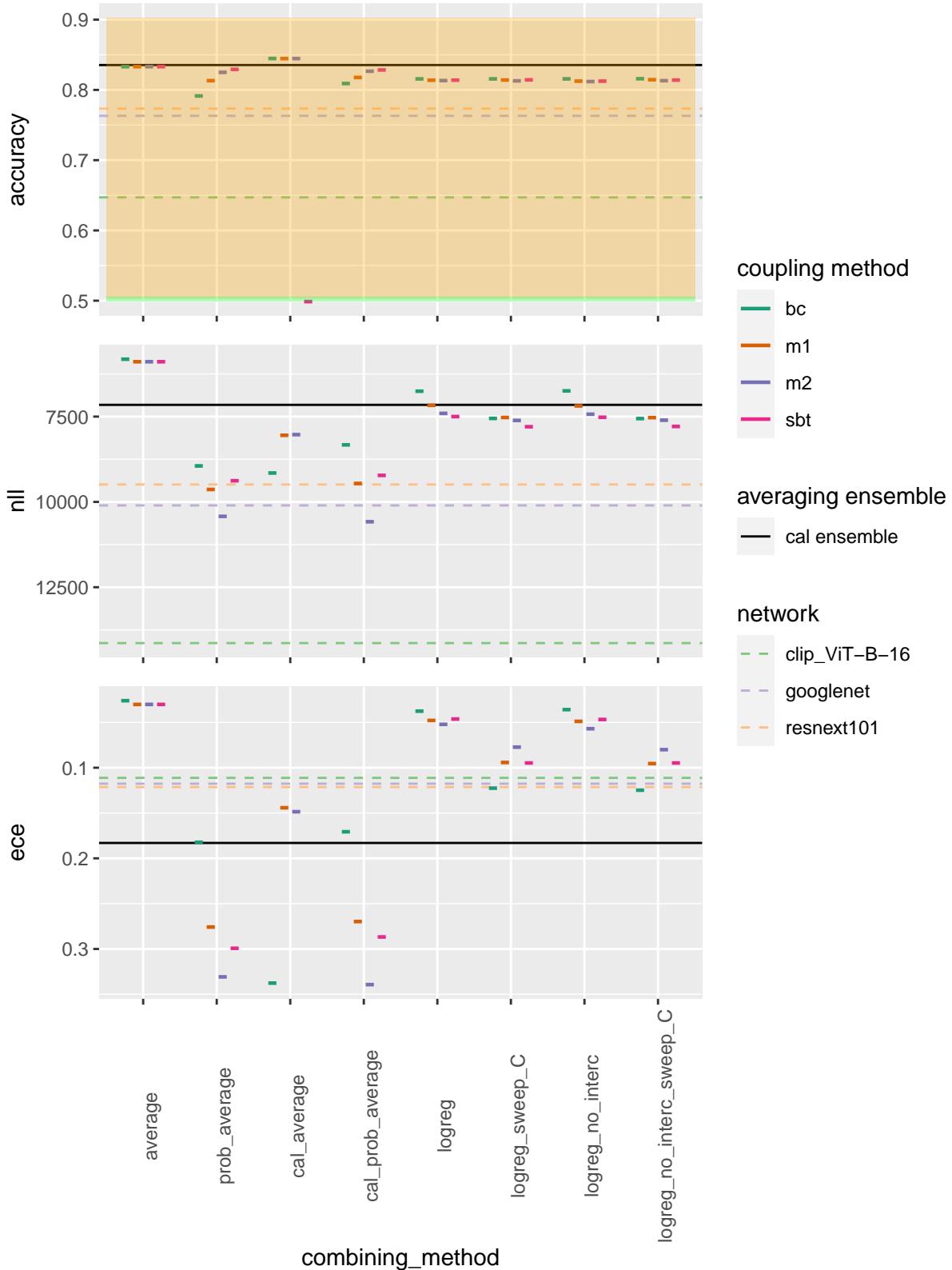
Ensemble metrics  
Error inconsistency 0.283600002527237



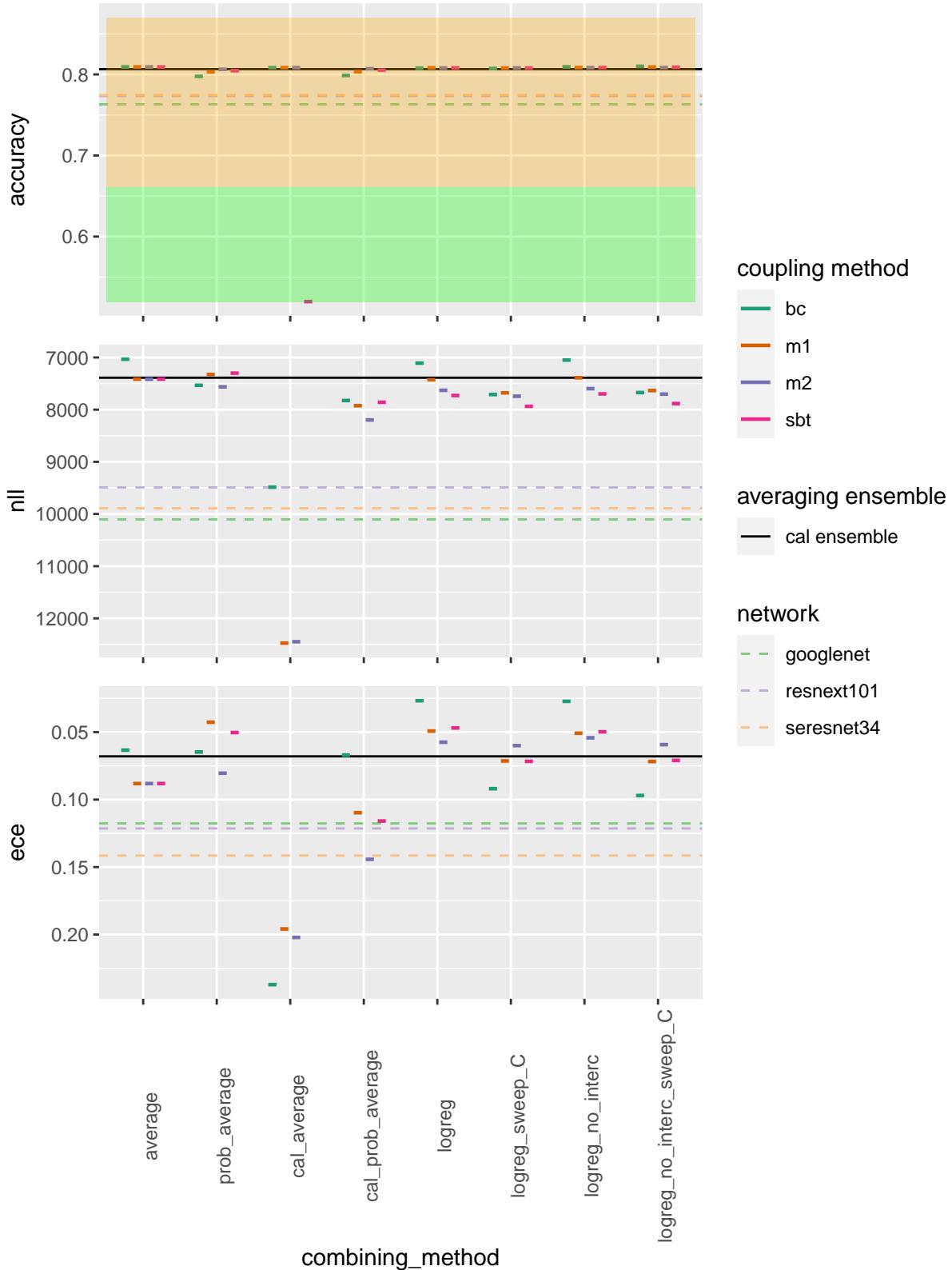
Ensemble metrics  
Error inconsistency 0.287599980831146



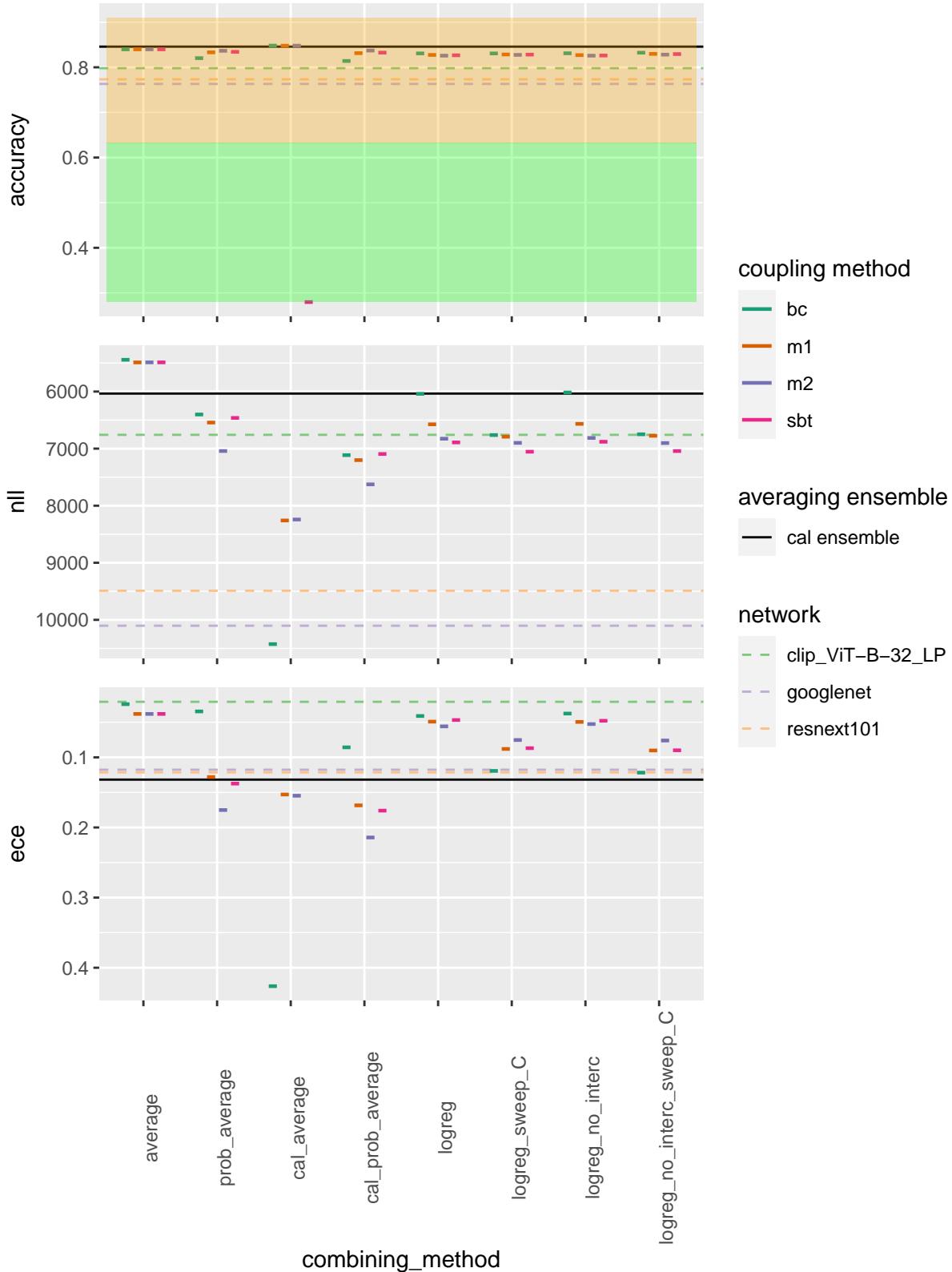
Ensemble metrics  
Error inconsistency 0.397499978542328



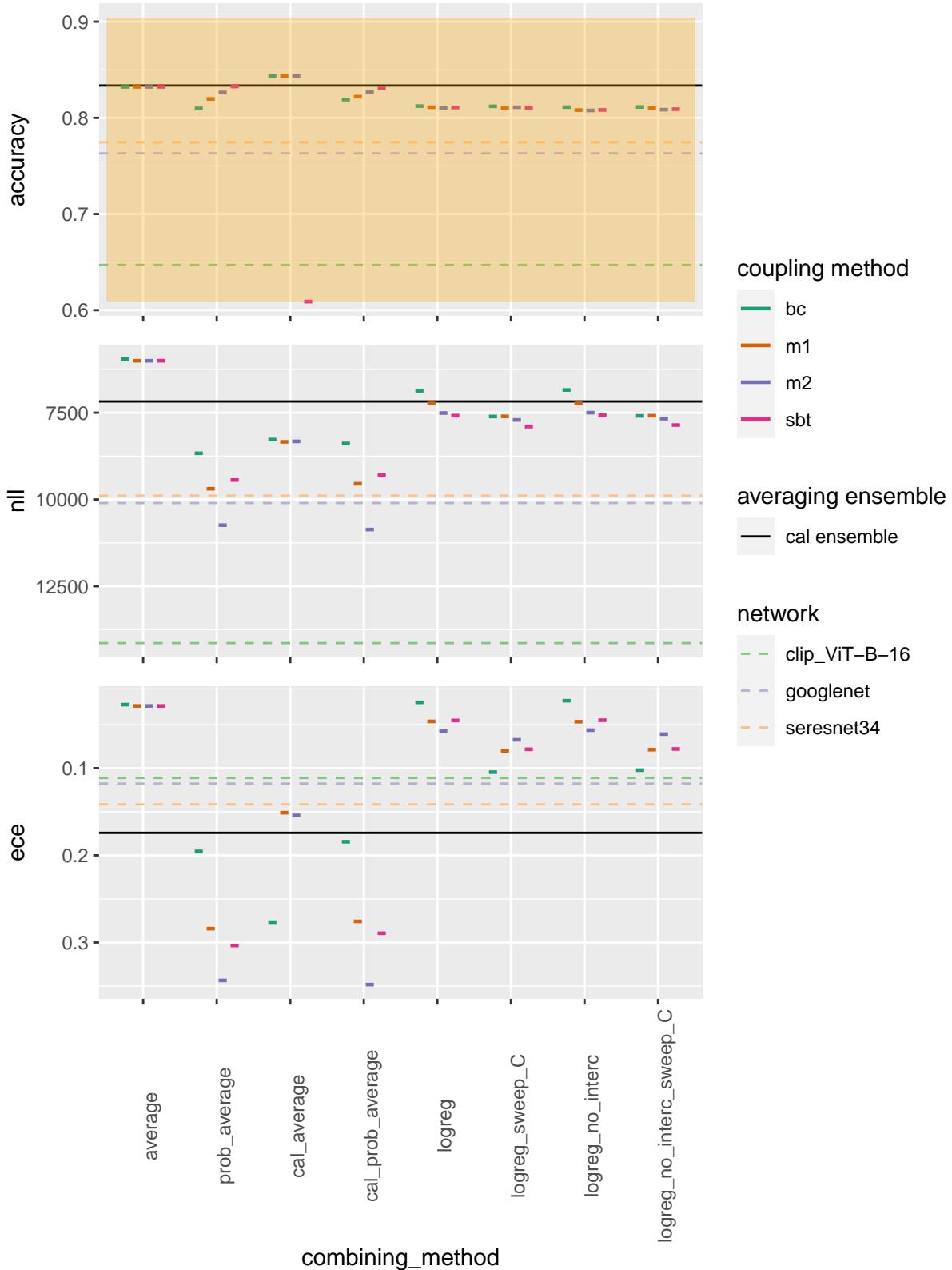
Ensemble metrics  
Error inconsistency 0.209099993109703



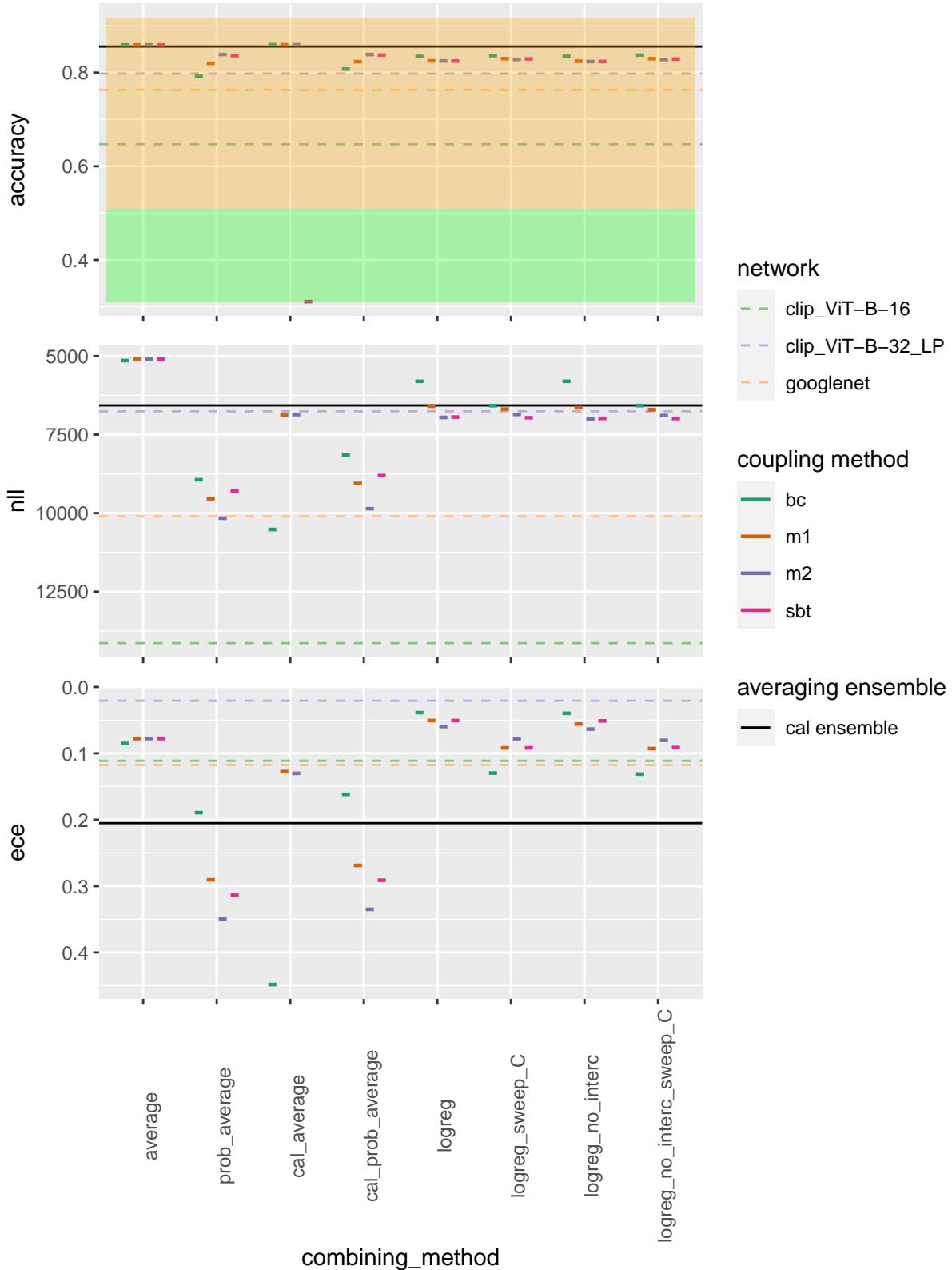
Ensemble metrics  
Error inconsistency 0.277700006961823



Ensemble metrics  
Error inconsistency 0.397399991750717

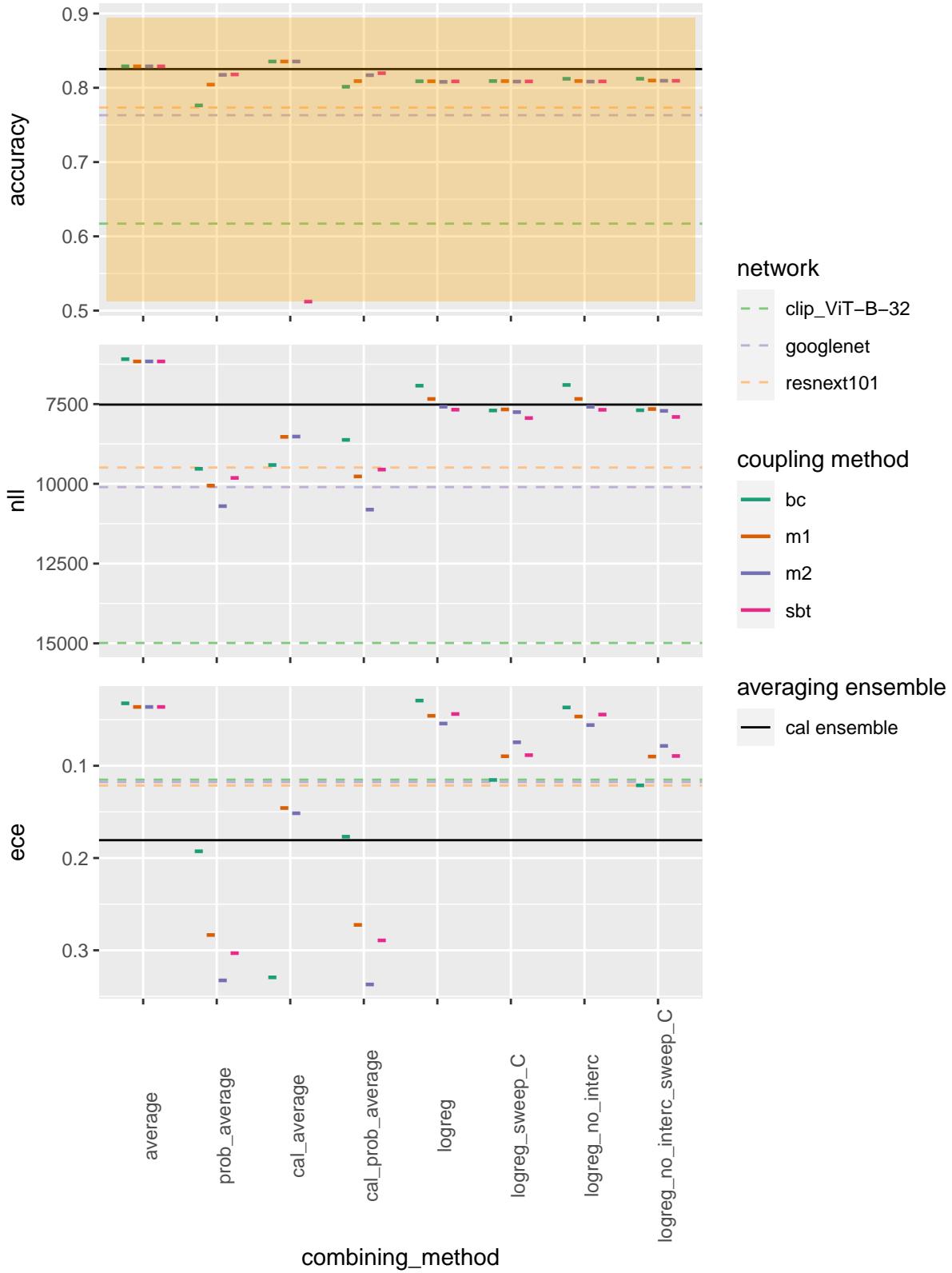


Ensemble metrics  
Error inconsistency 0.407899975776672

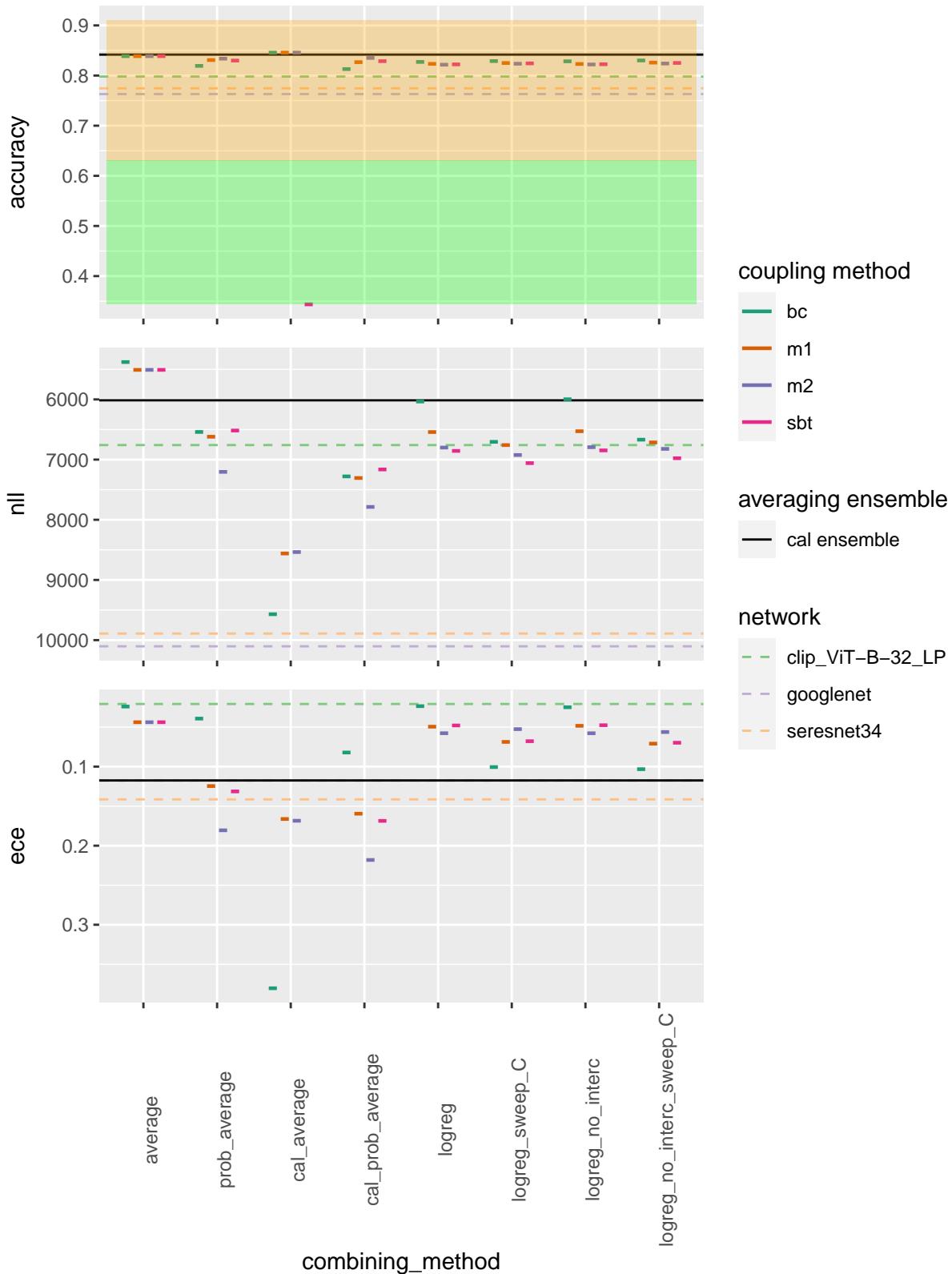


## Ensemble metrics

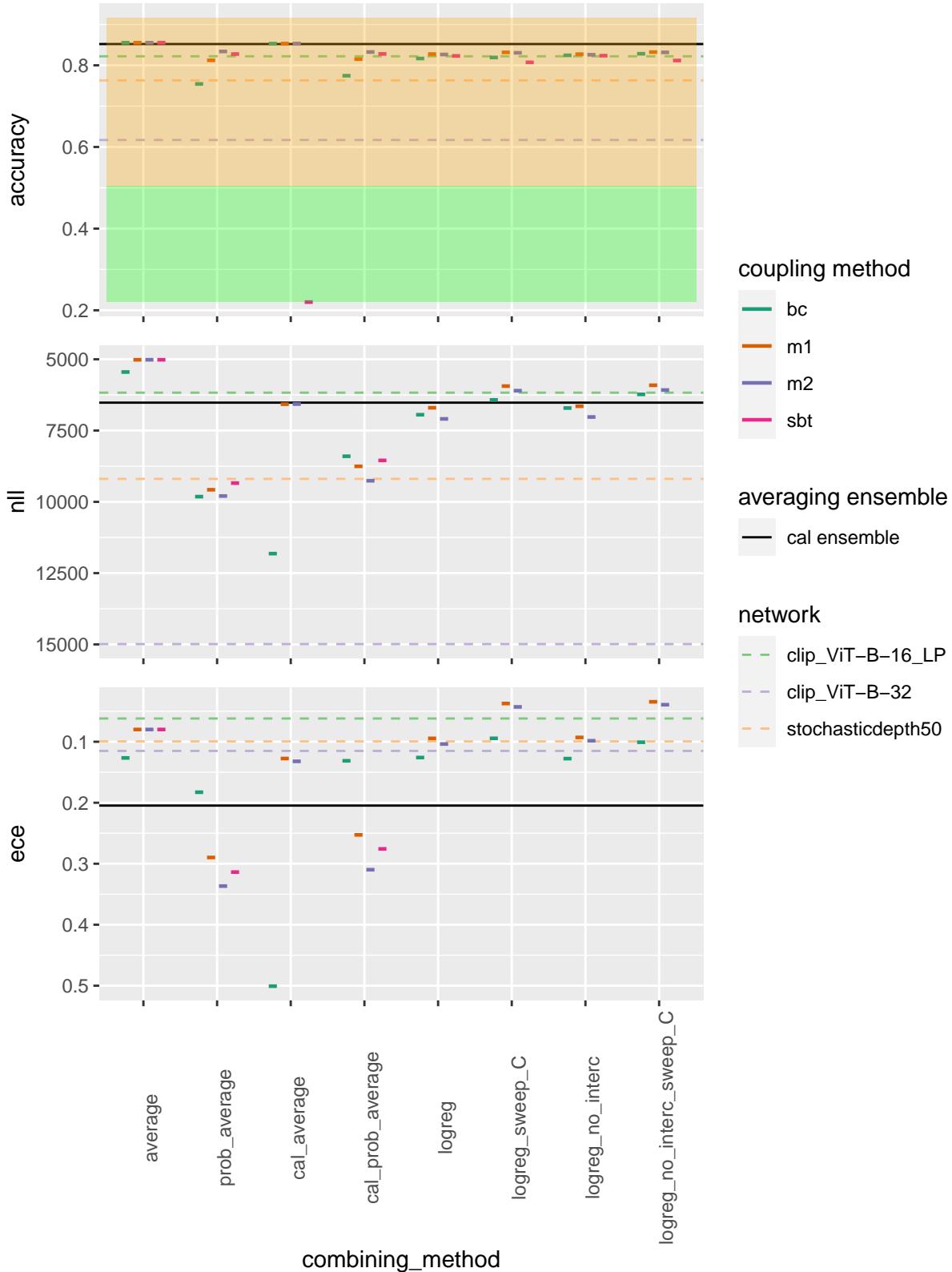
Error inconsistency 0.403099983930588



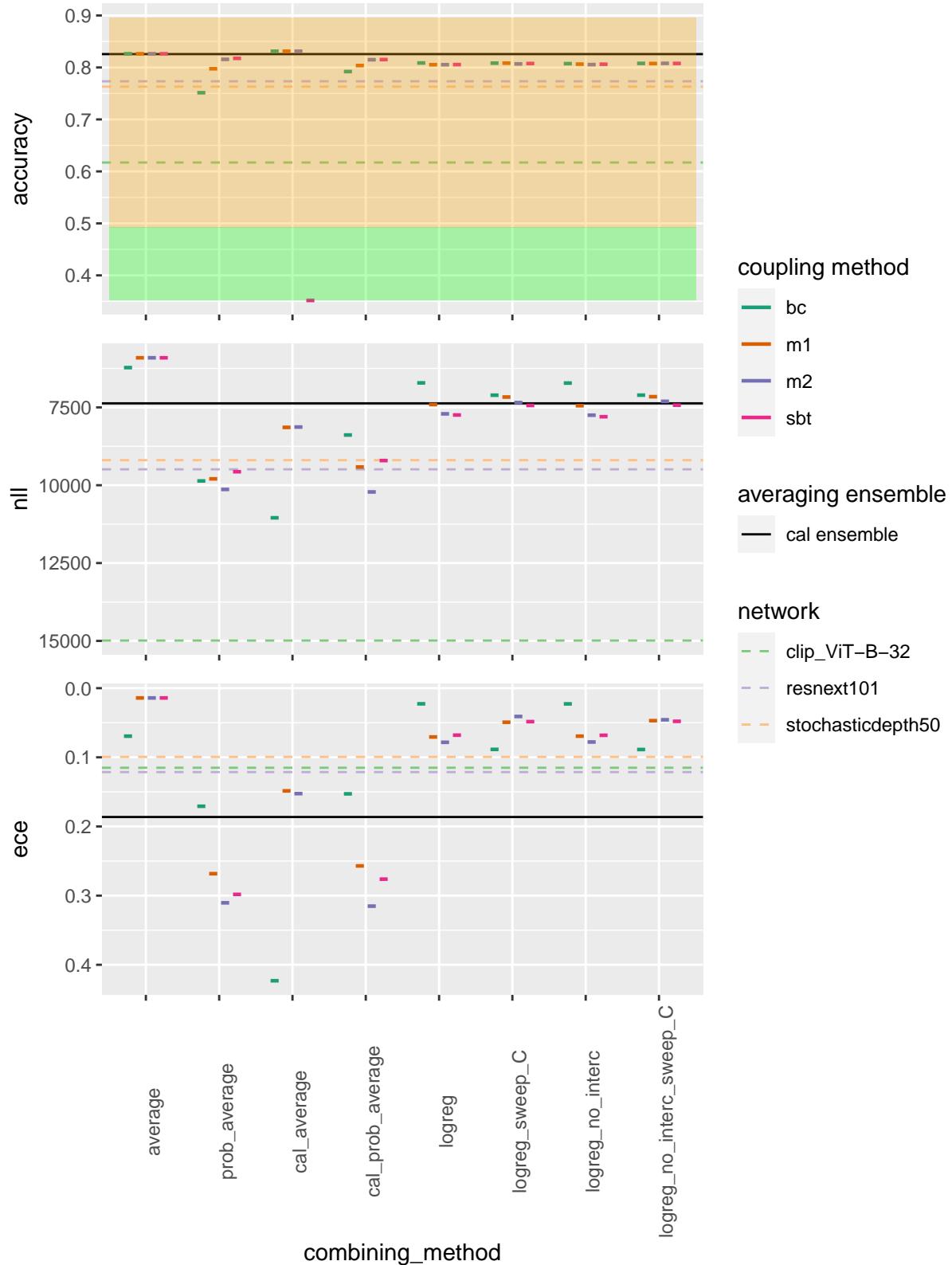
Ensemble metrics  
Error inconsistency 0.27989998459816



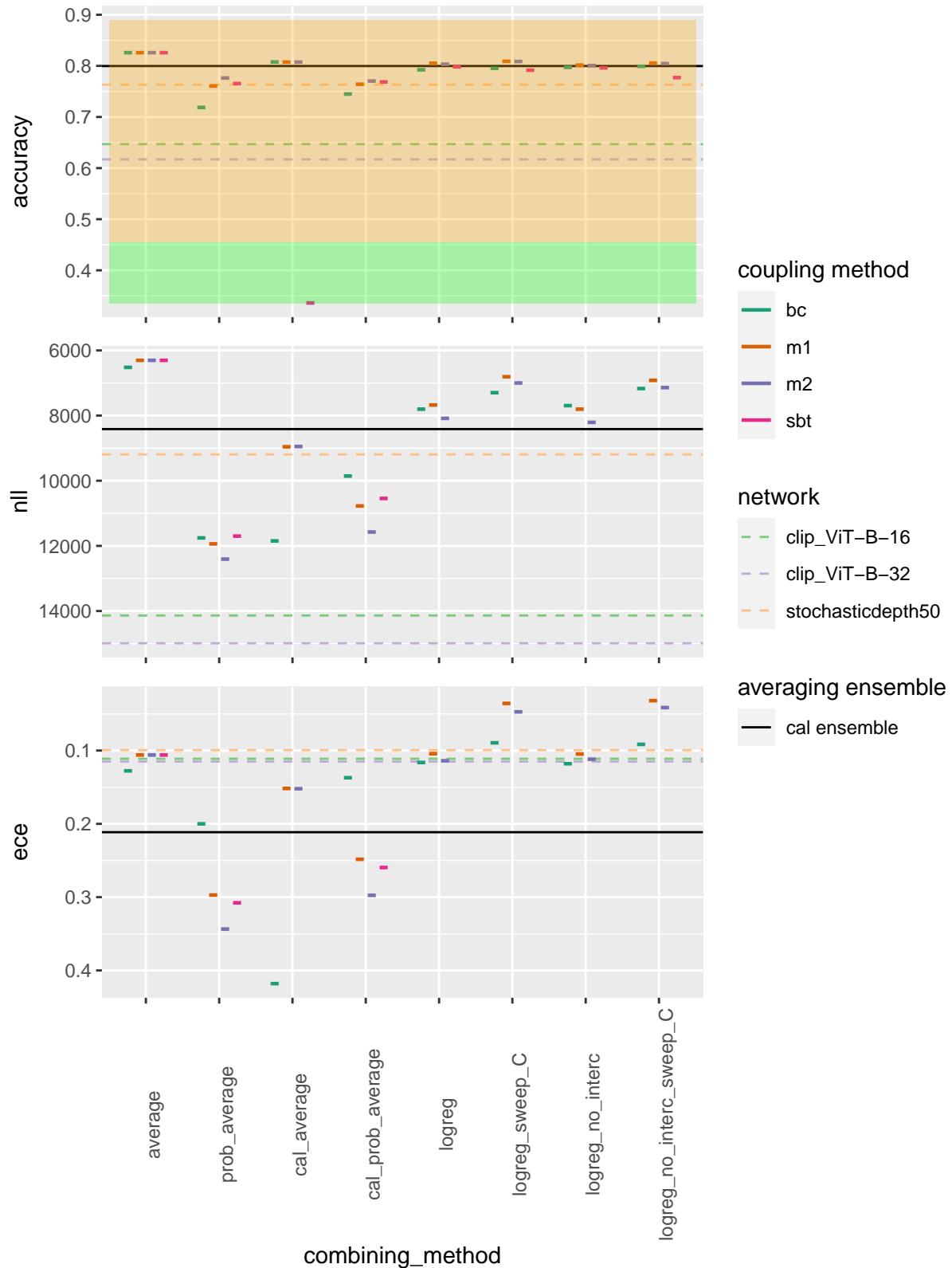
Ensemble metrics  
Error inconsistency 0.412399977445602



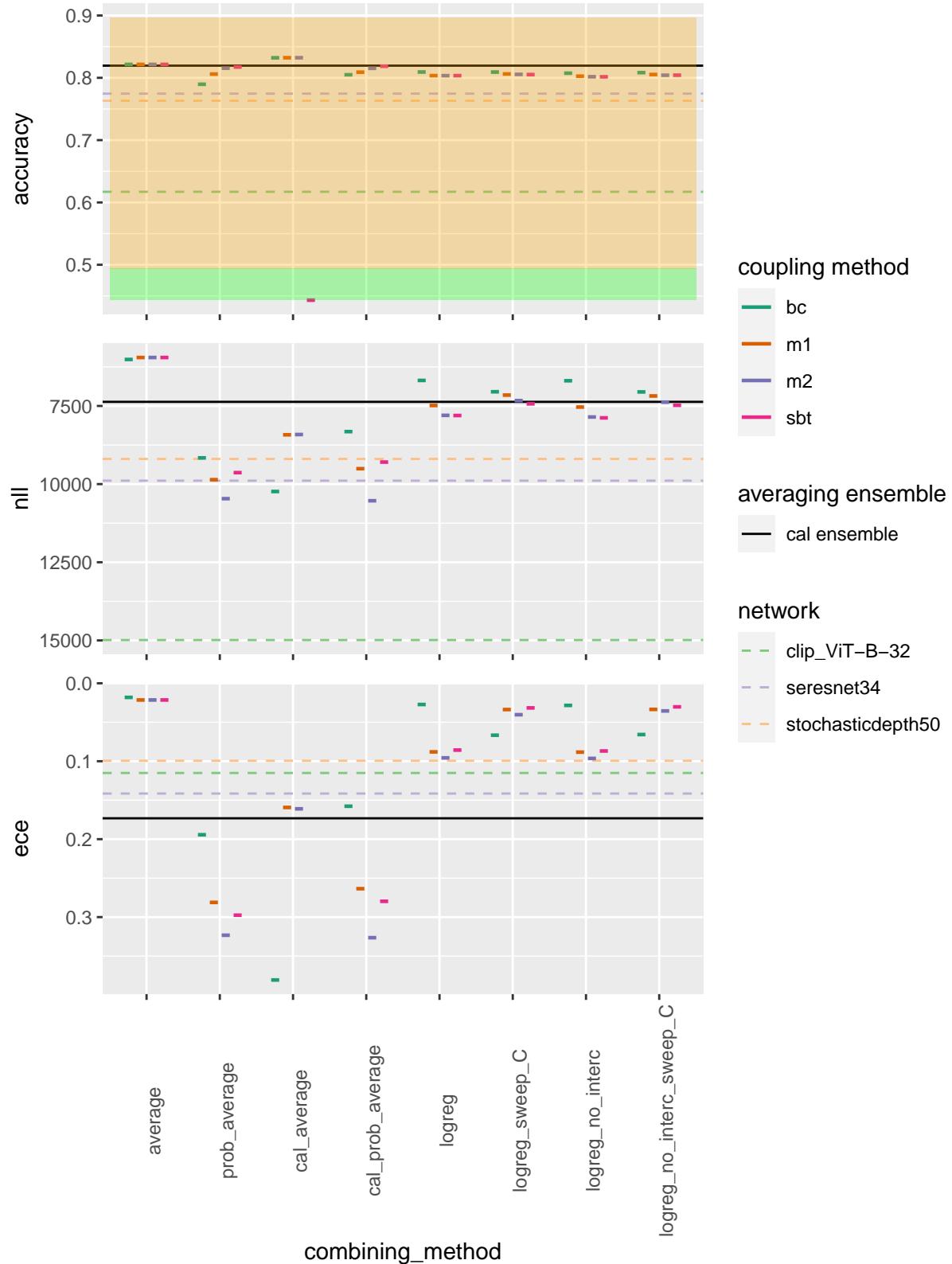
Ensemble metrics  
Error inconsistency 0.40379998087883



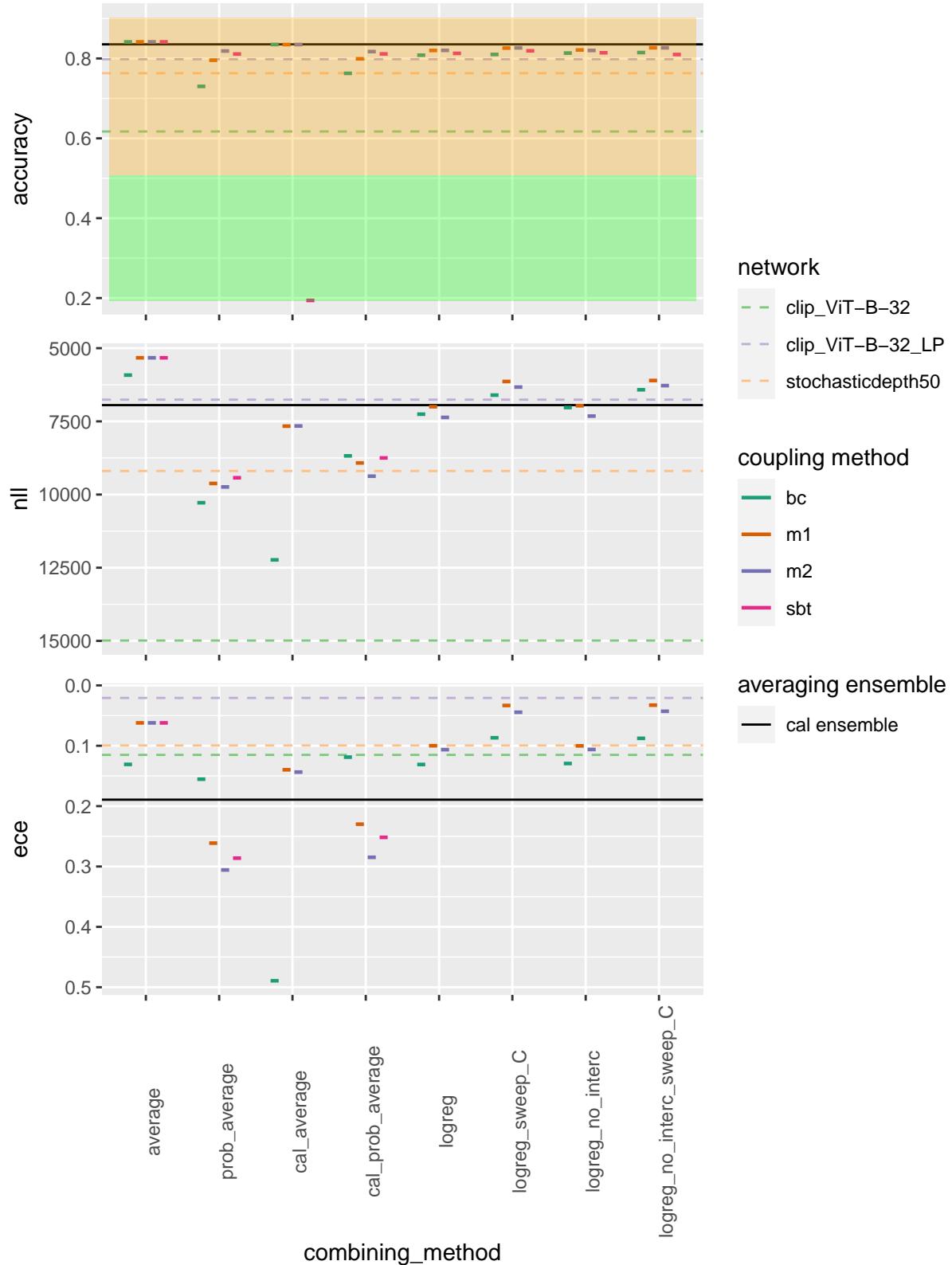
Ensemble metrics  
Error inconsistency 0.434699982404709



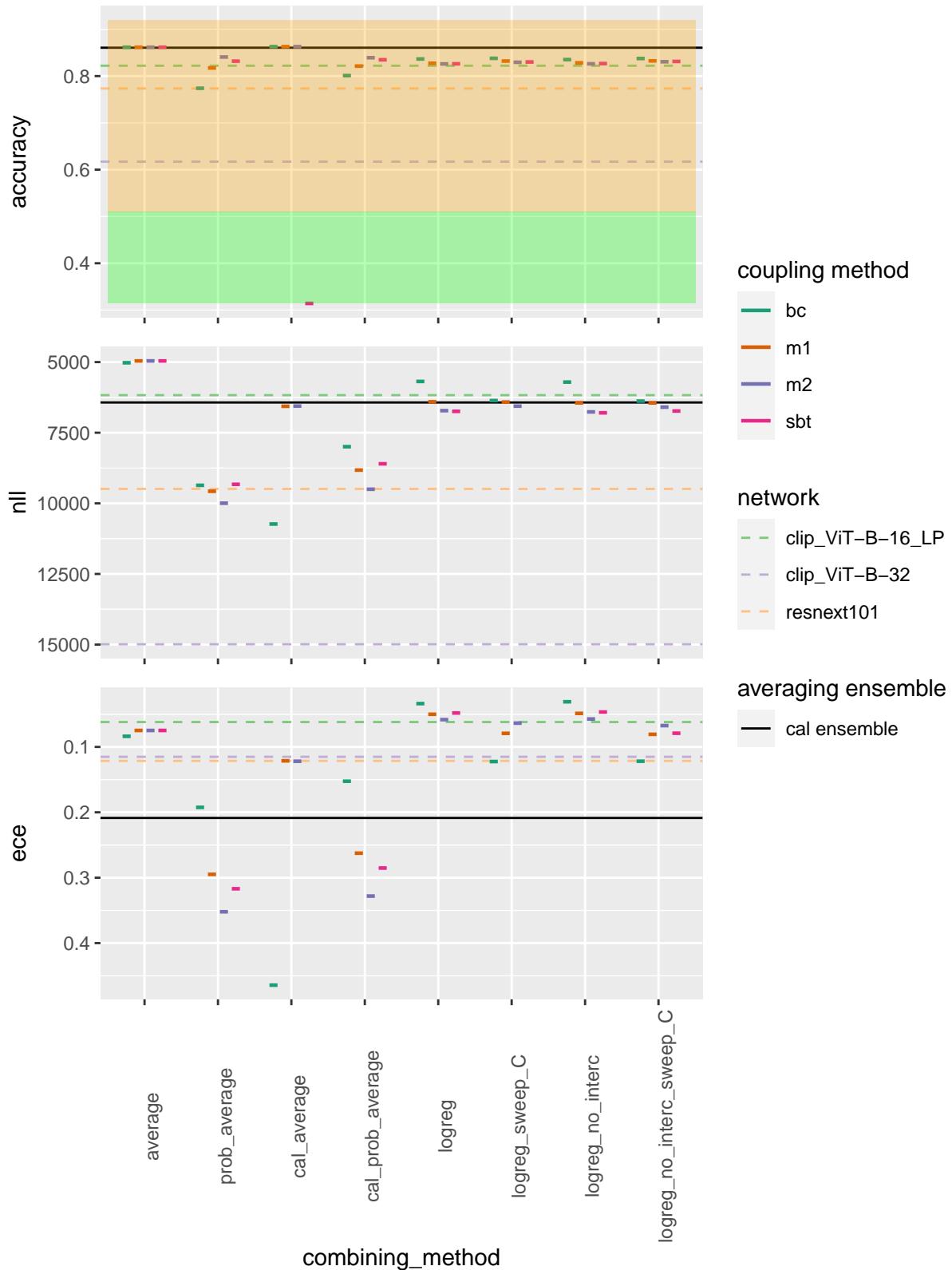
Ensemble metrics  
Error inconsistency 0.402899980545044



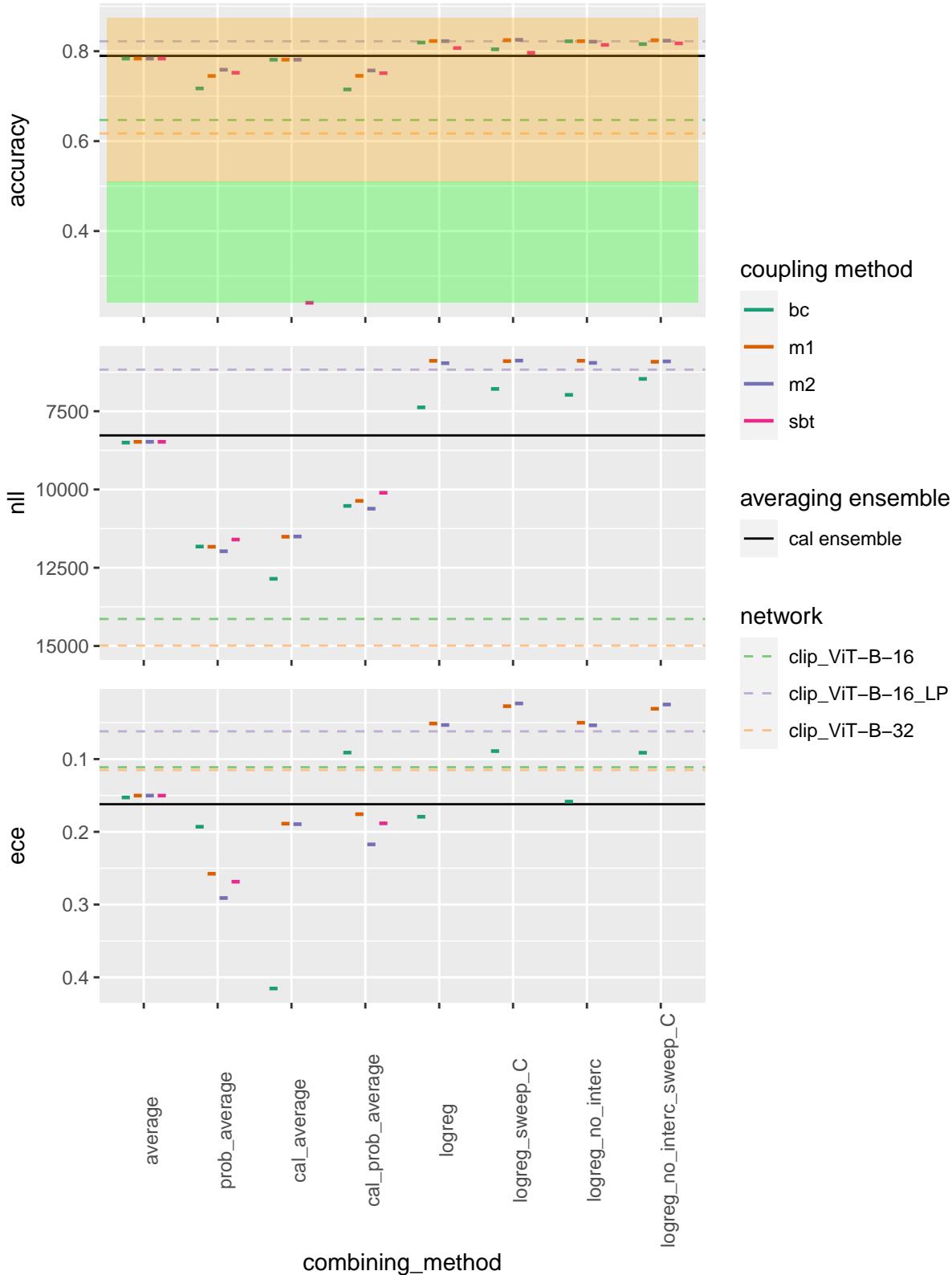
Ensemble metrics  
Error inconsistency 0.39479997754097



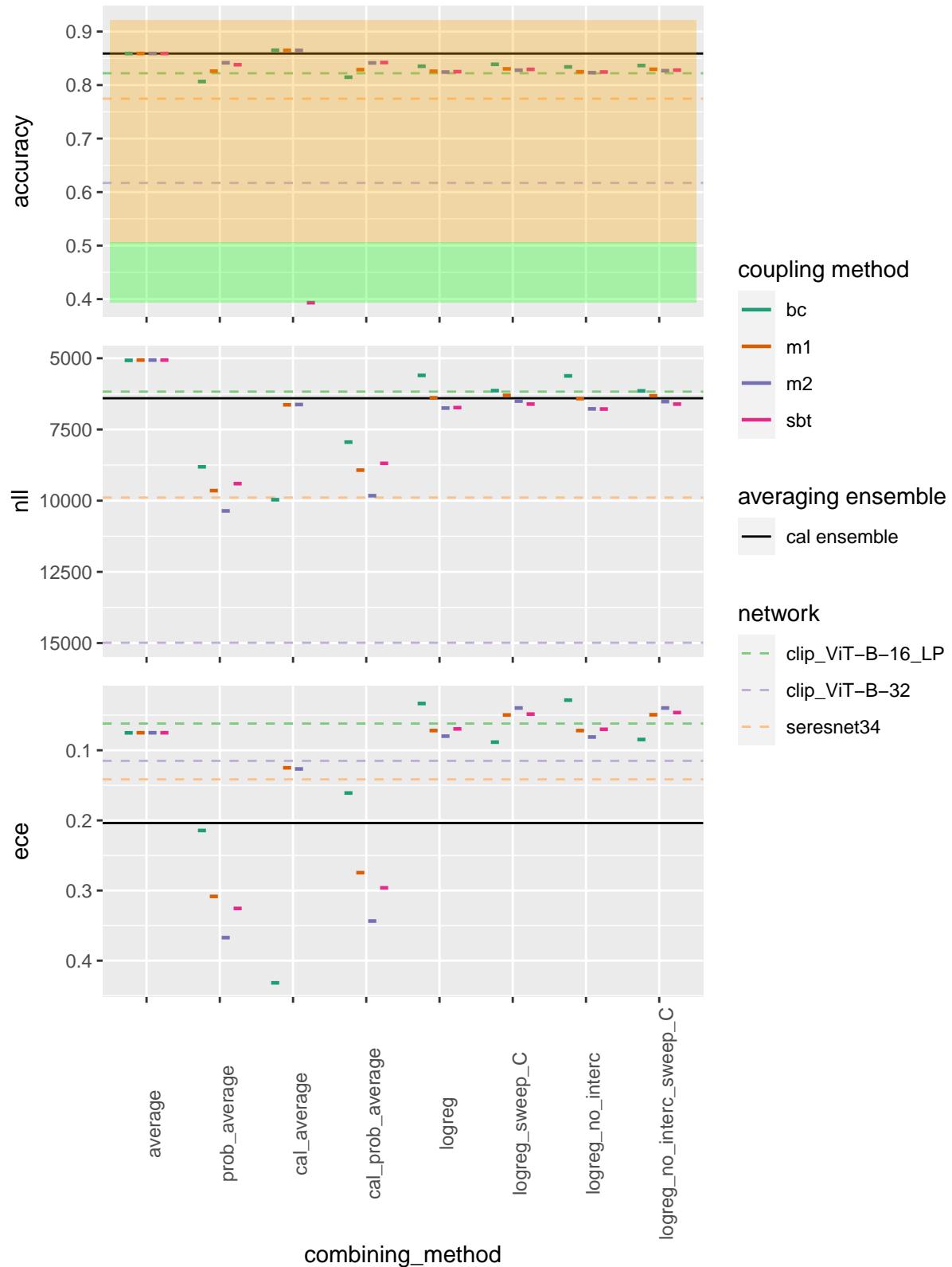
Ensemble metrics  
Error inconsistency 0.409999996423721



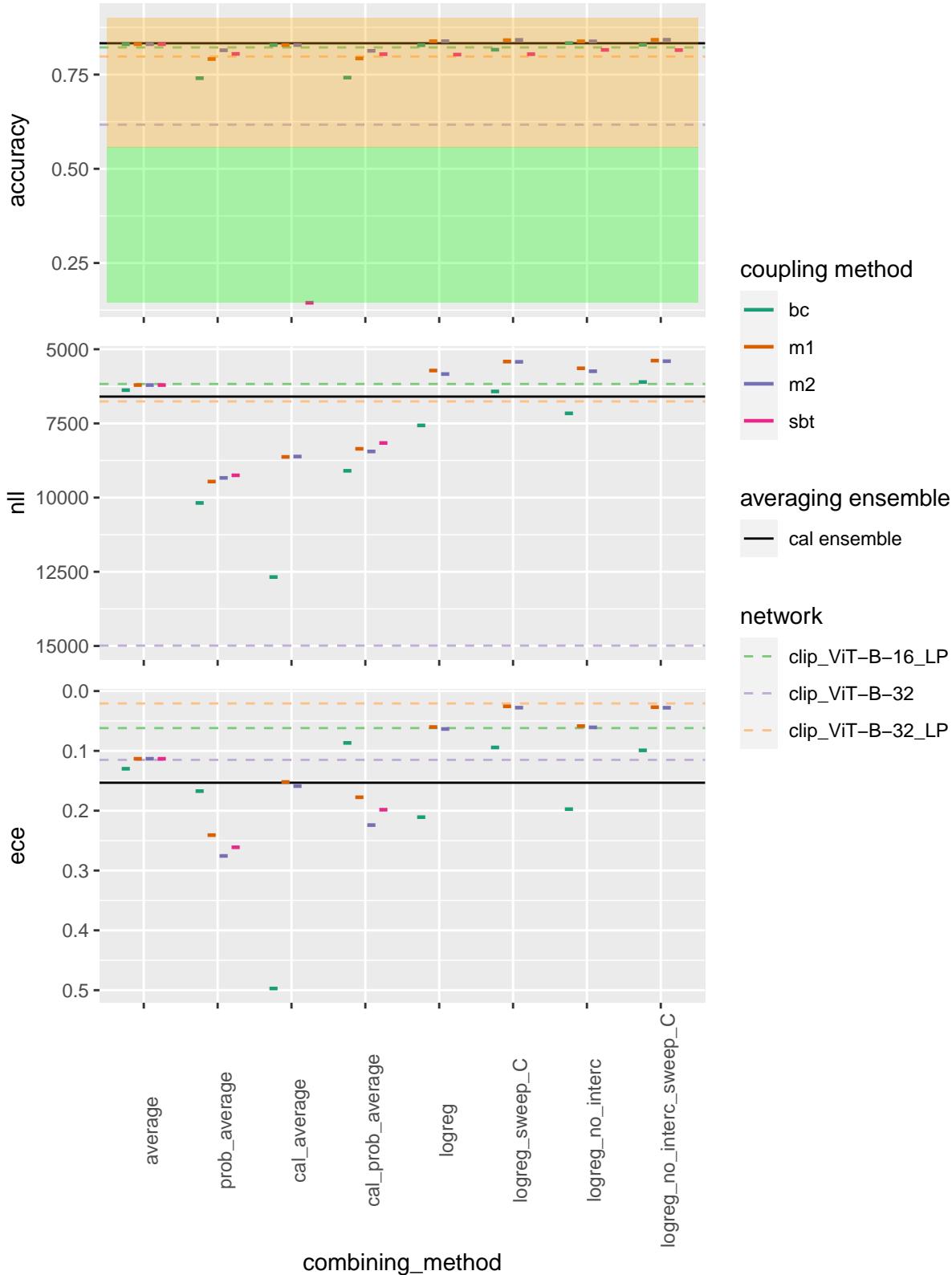
Ensemble metrics  
Error inconsistency 0.36489999294281



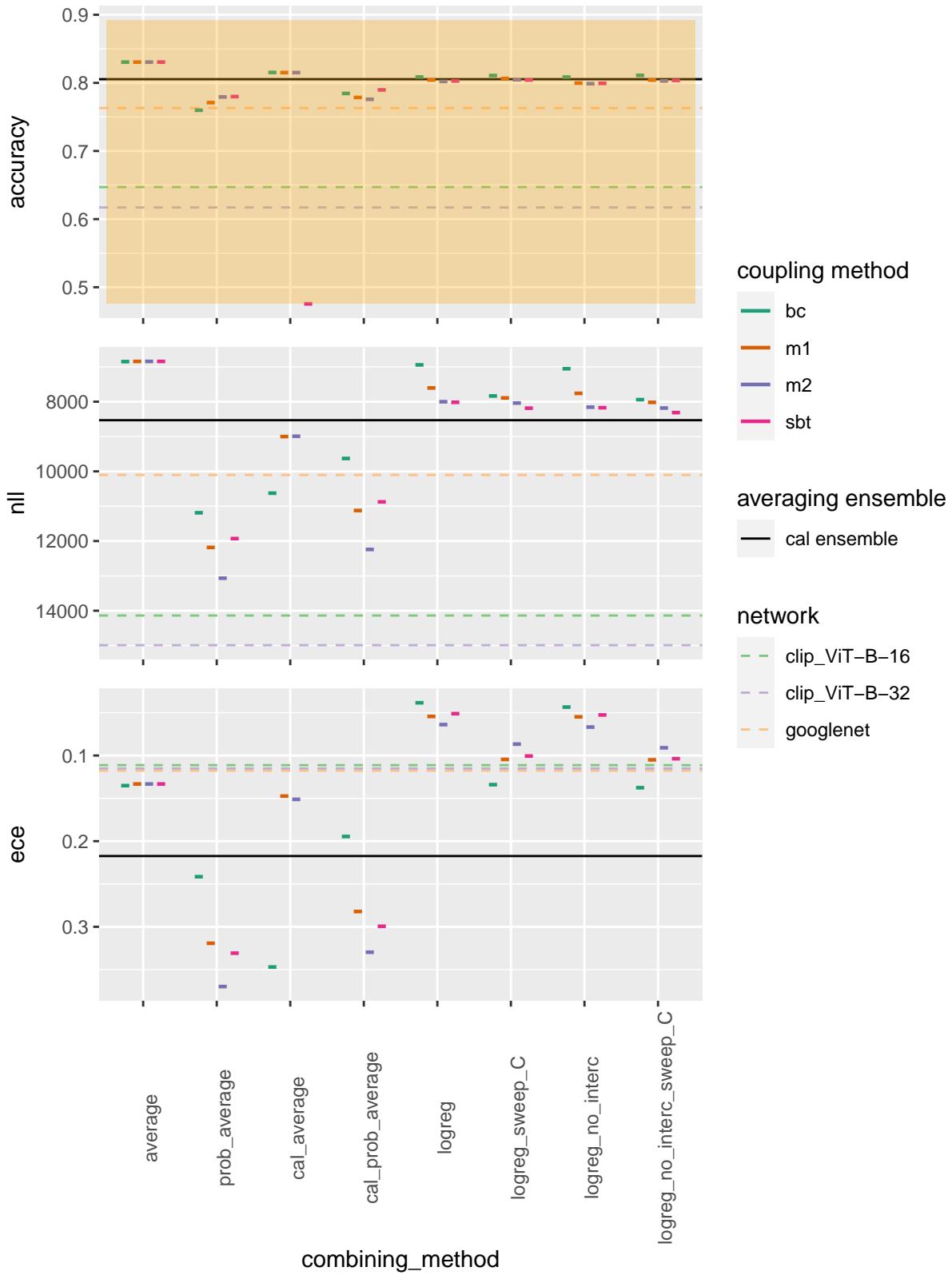
Ensemble metrics  
Error inconsistency 0.415299981832504



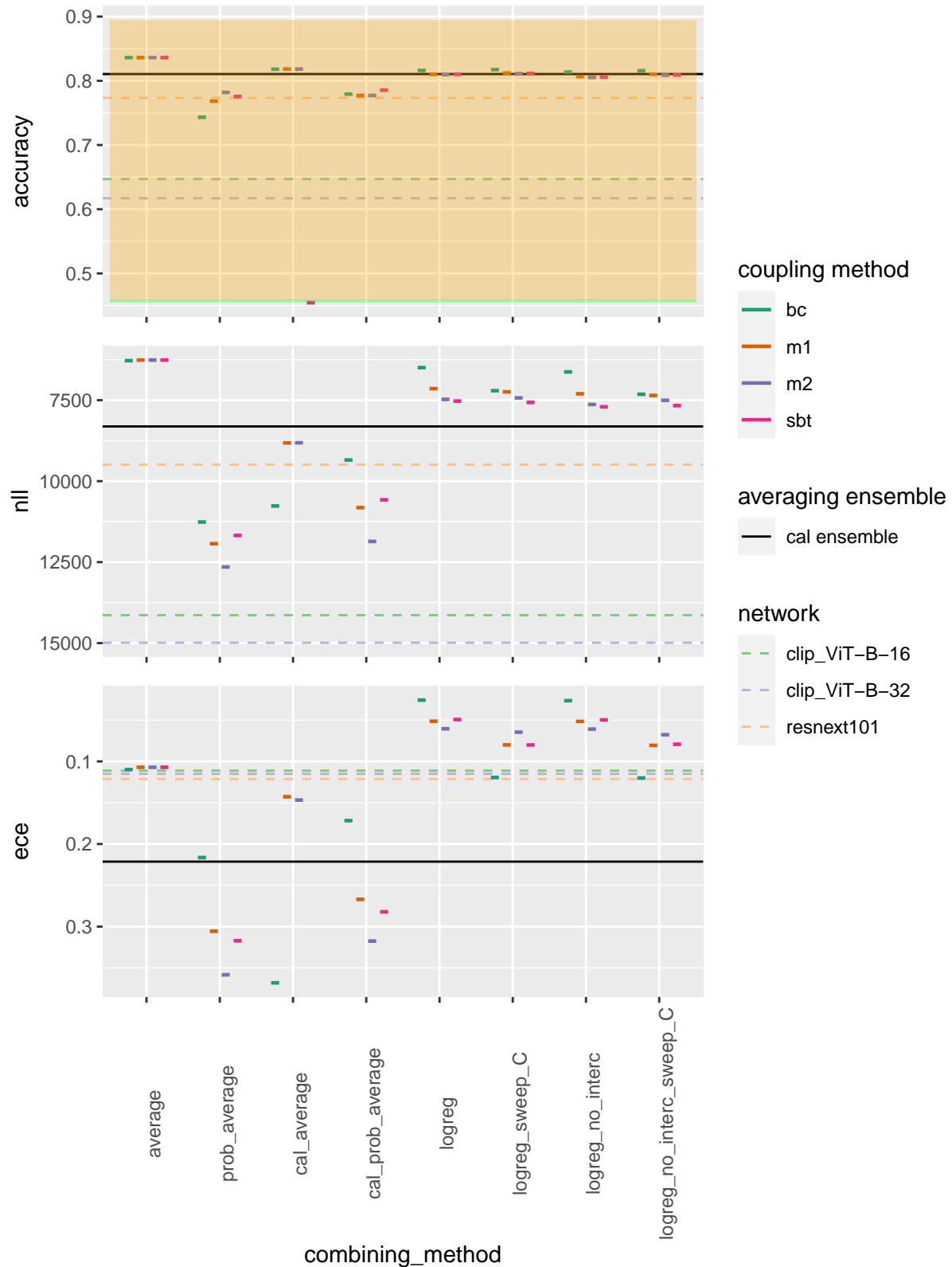
Ensemble metrics  
Error inconsistency 0.344300001859665



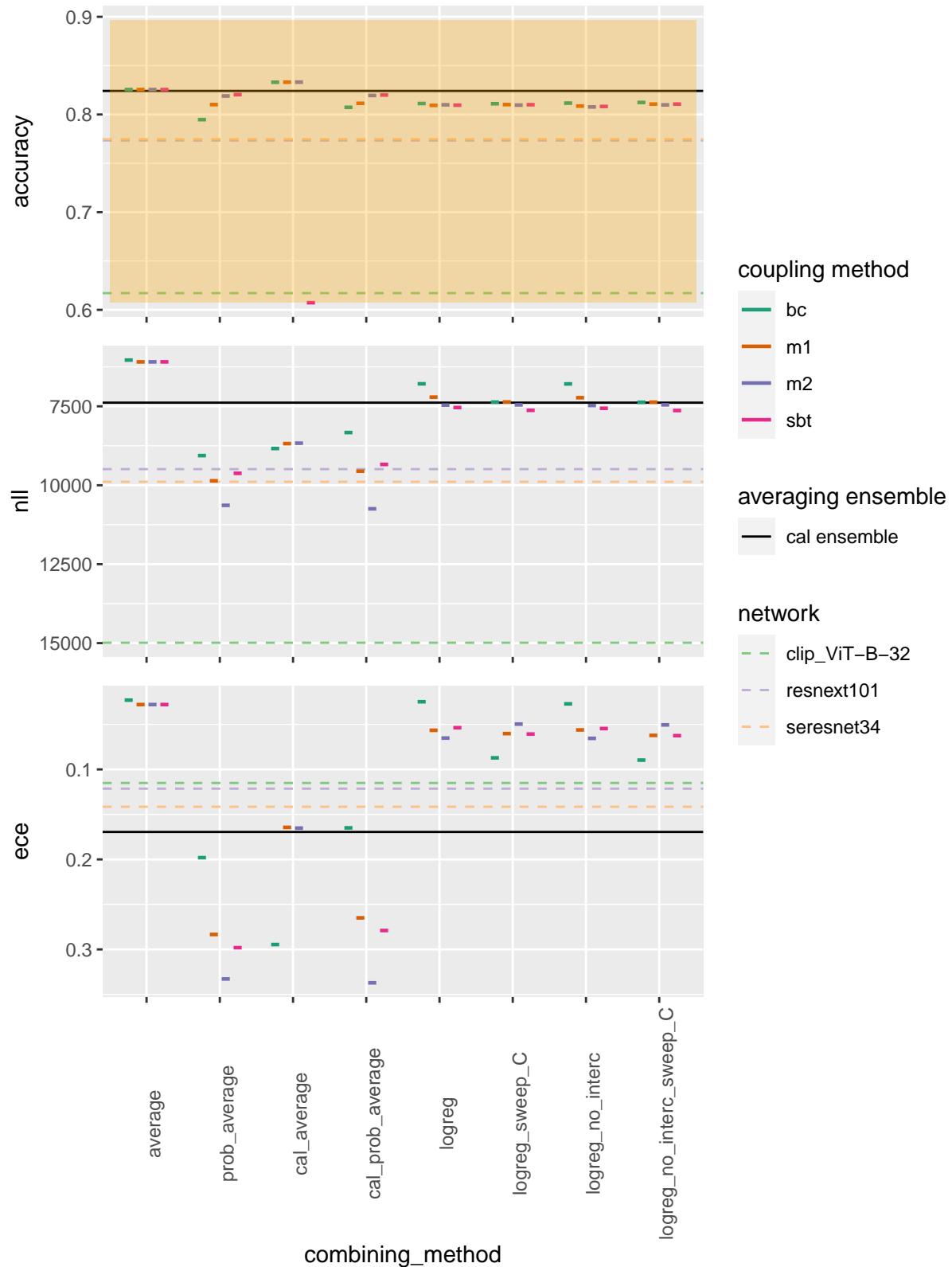
Ensemble metrics  
Error inconsistency 0.438899993896484



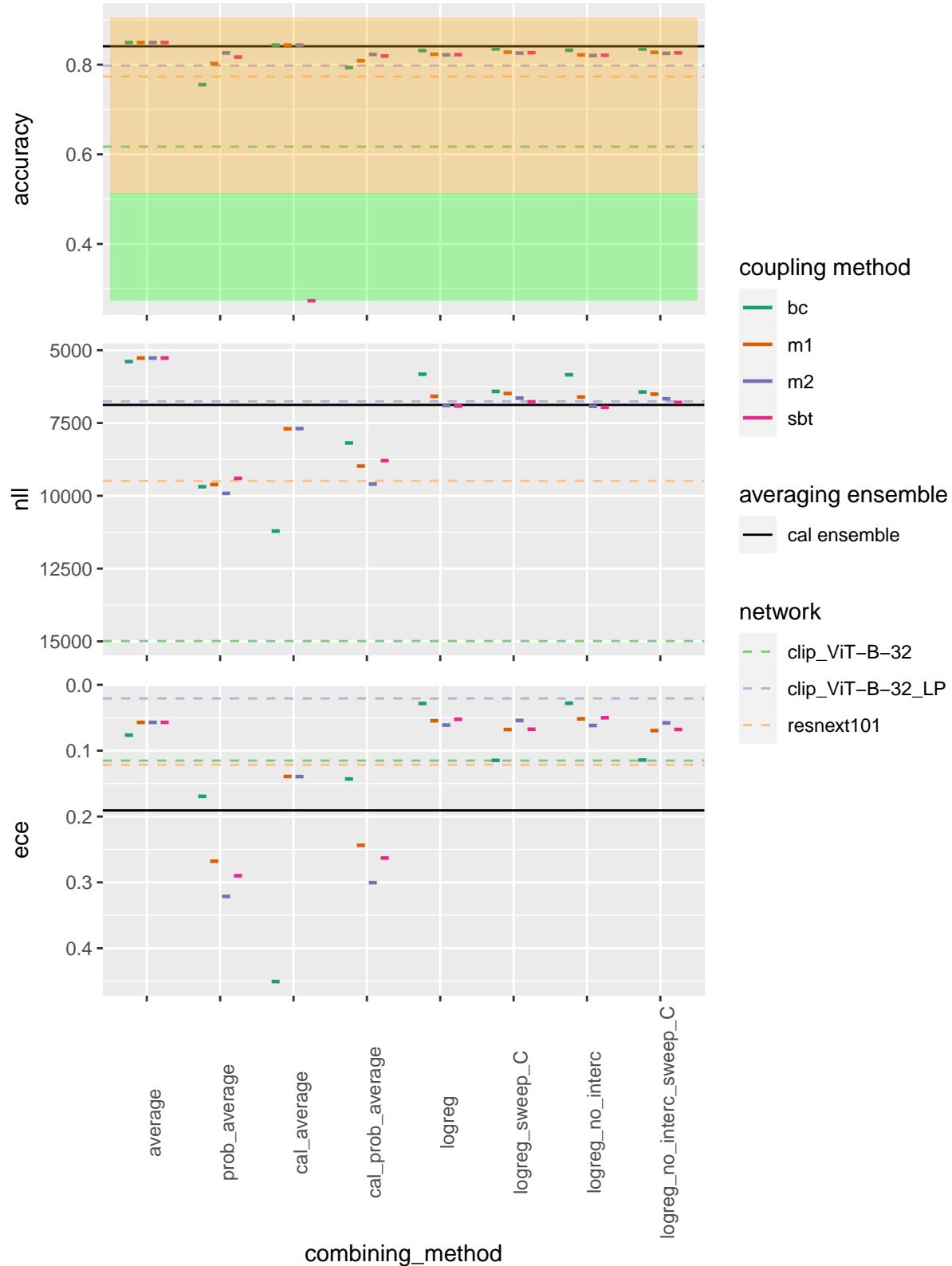
Ensemble metrics  
Error inconsistency 0.435699999332428



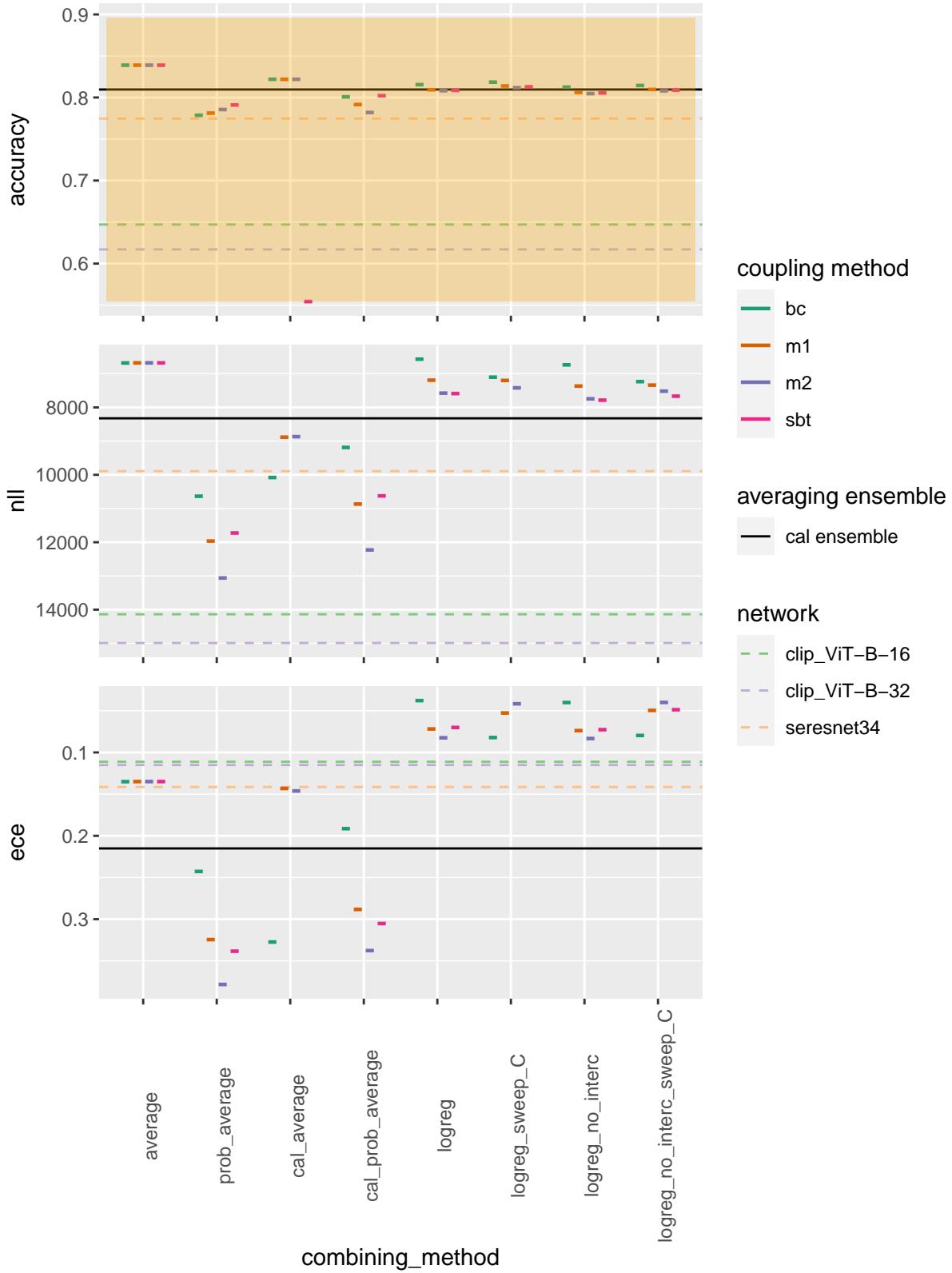
Ensemble metrics  
Error inconsistency 0.398900002241135



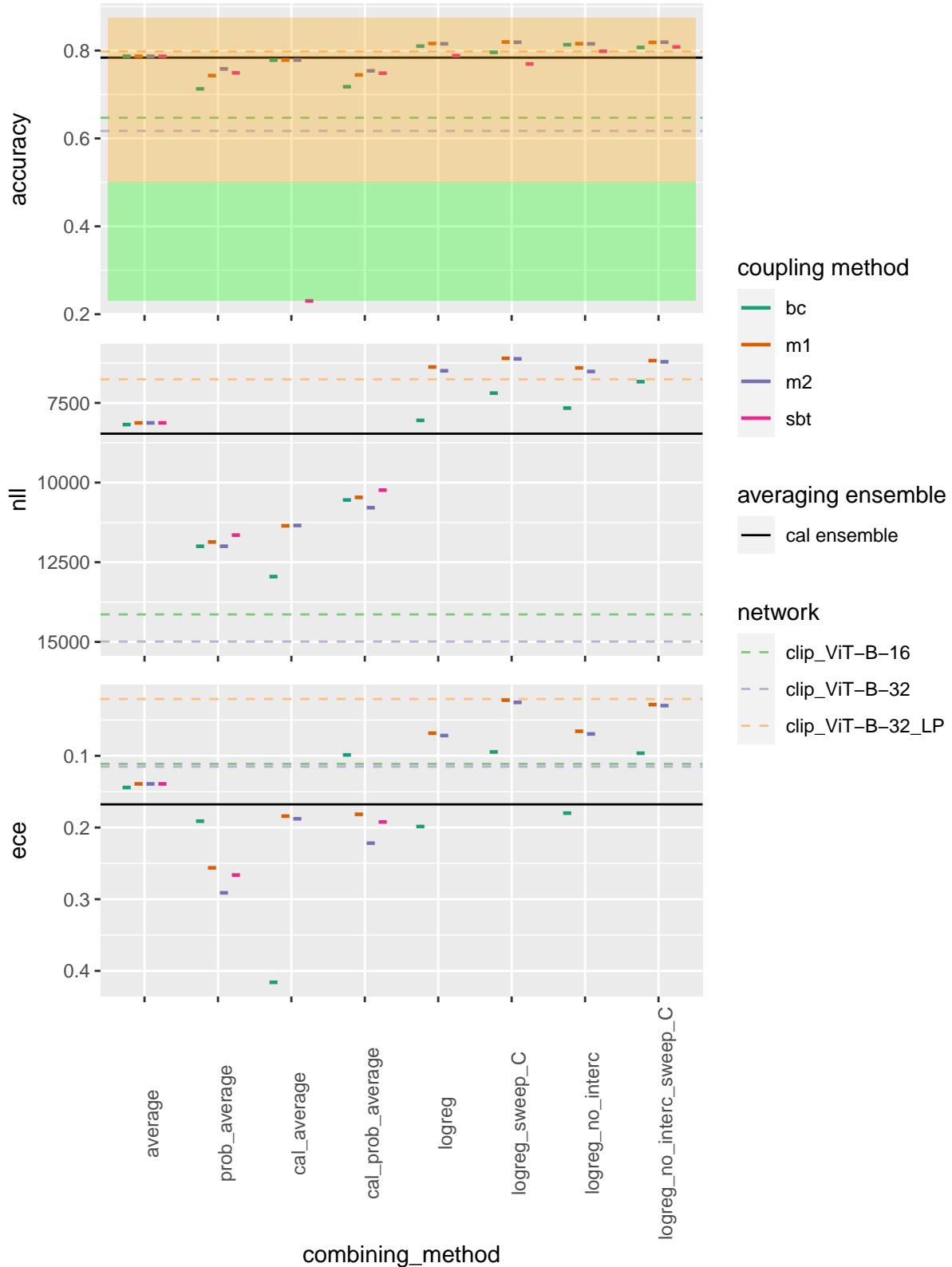
Ensemble metrics  
Error inconsistency 0.392499983310699



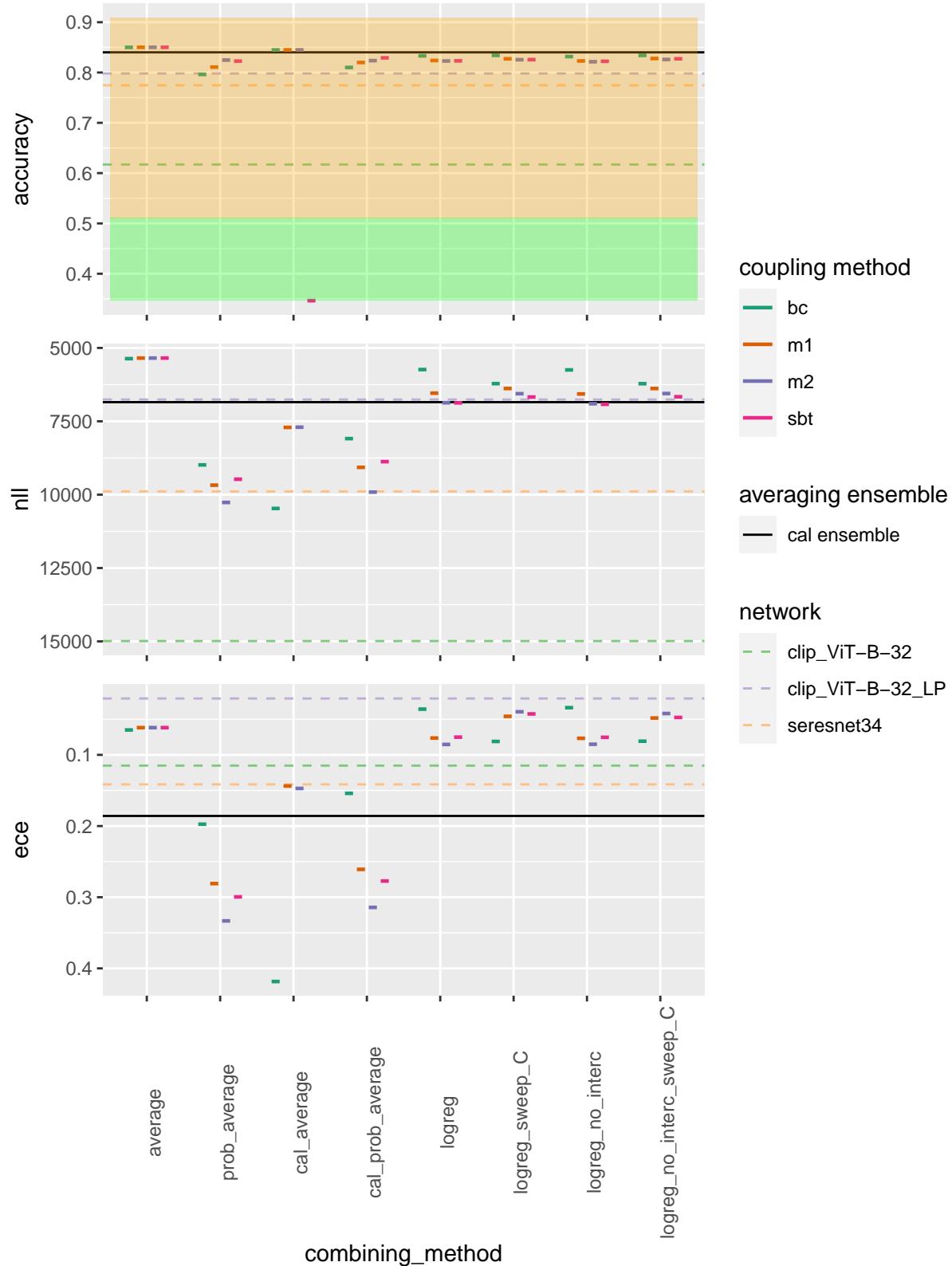
Ensemble metrics  
Error inconsistency 0.438999980688095



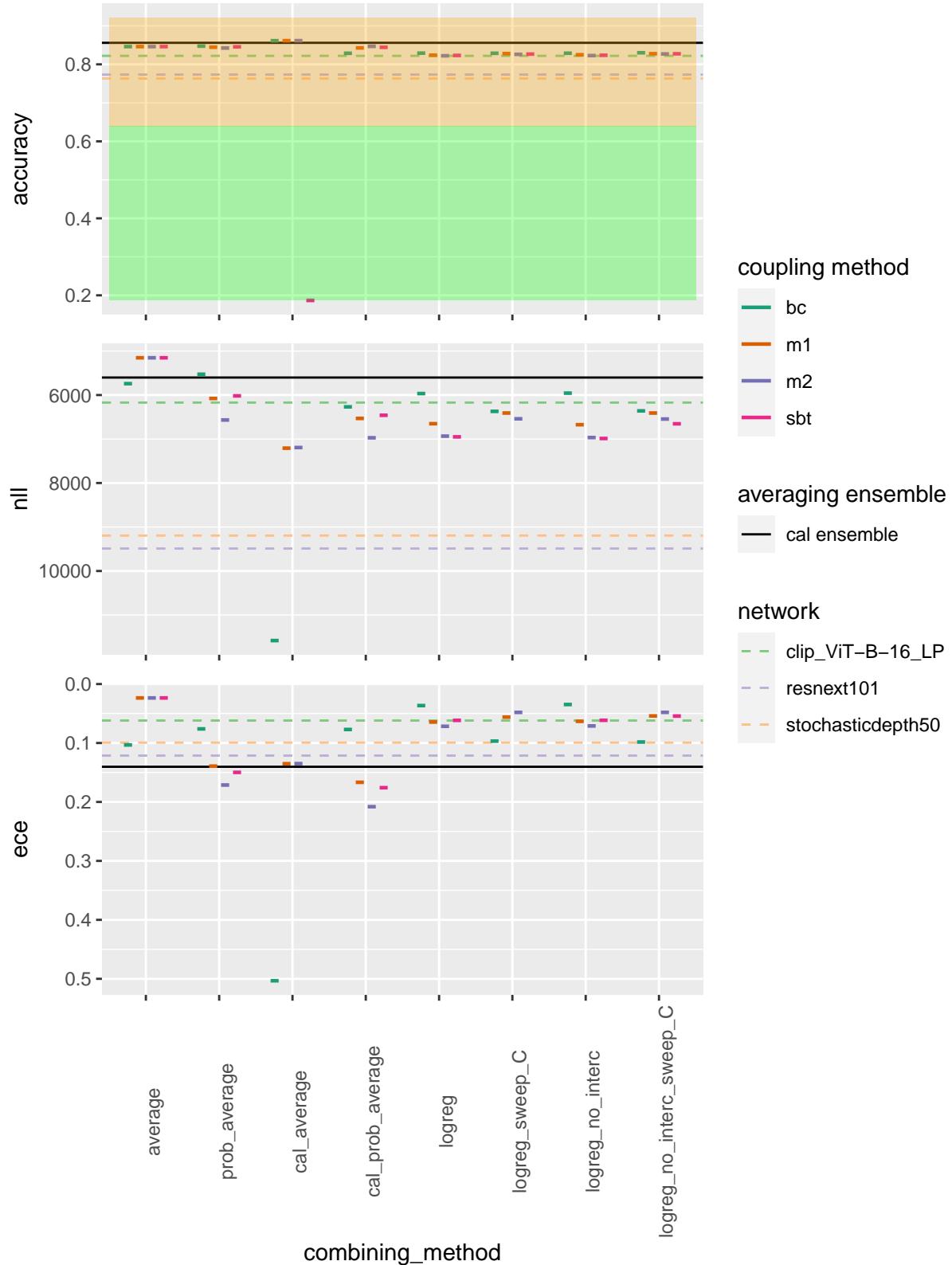
Ensemble metrics  
Error inconsistency 0.374499976634979



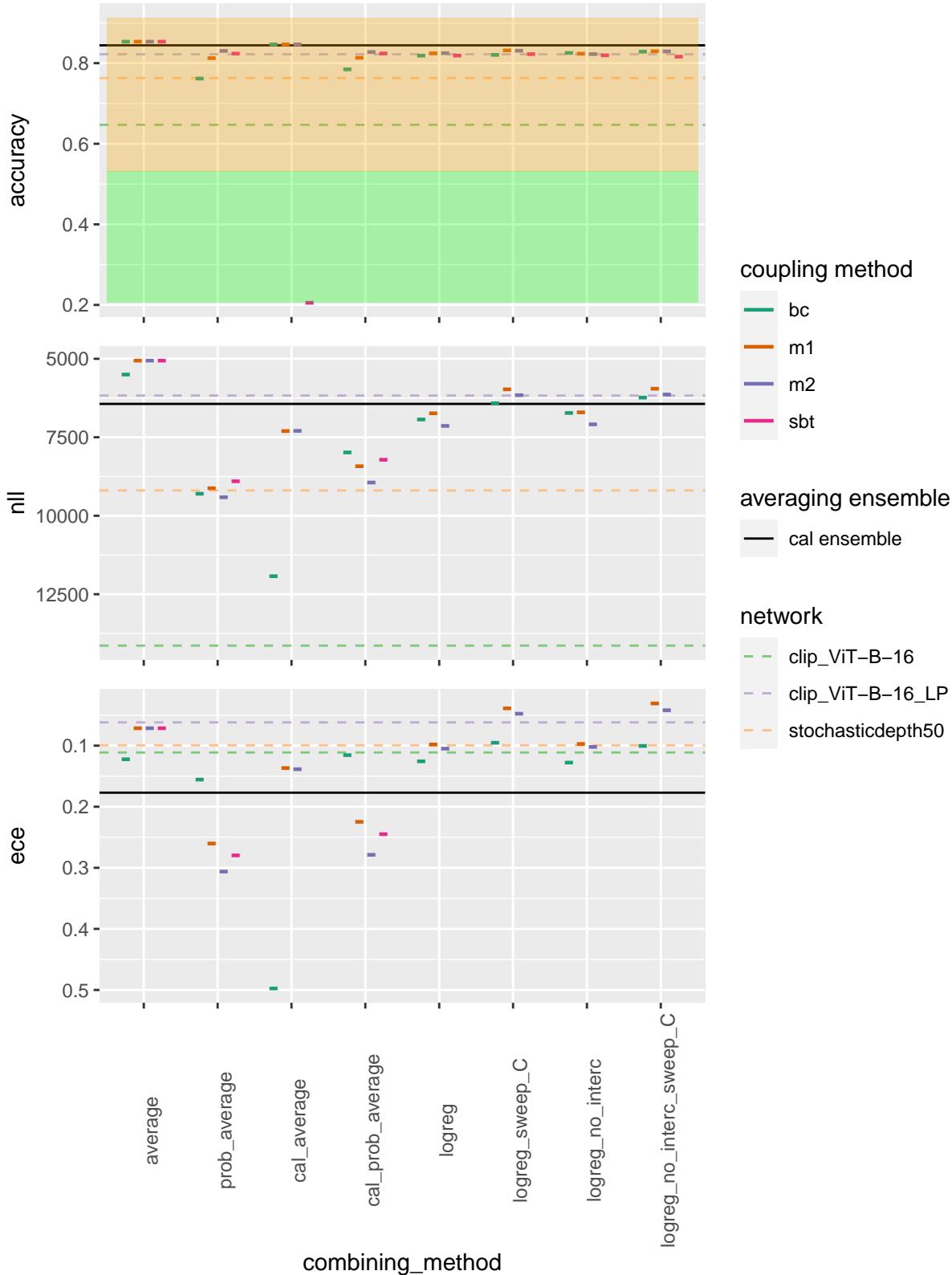
Ensemble metrics  
Error inconsistency 0.398099988698959



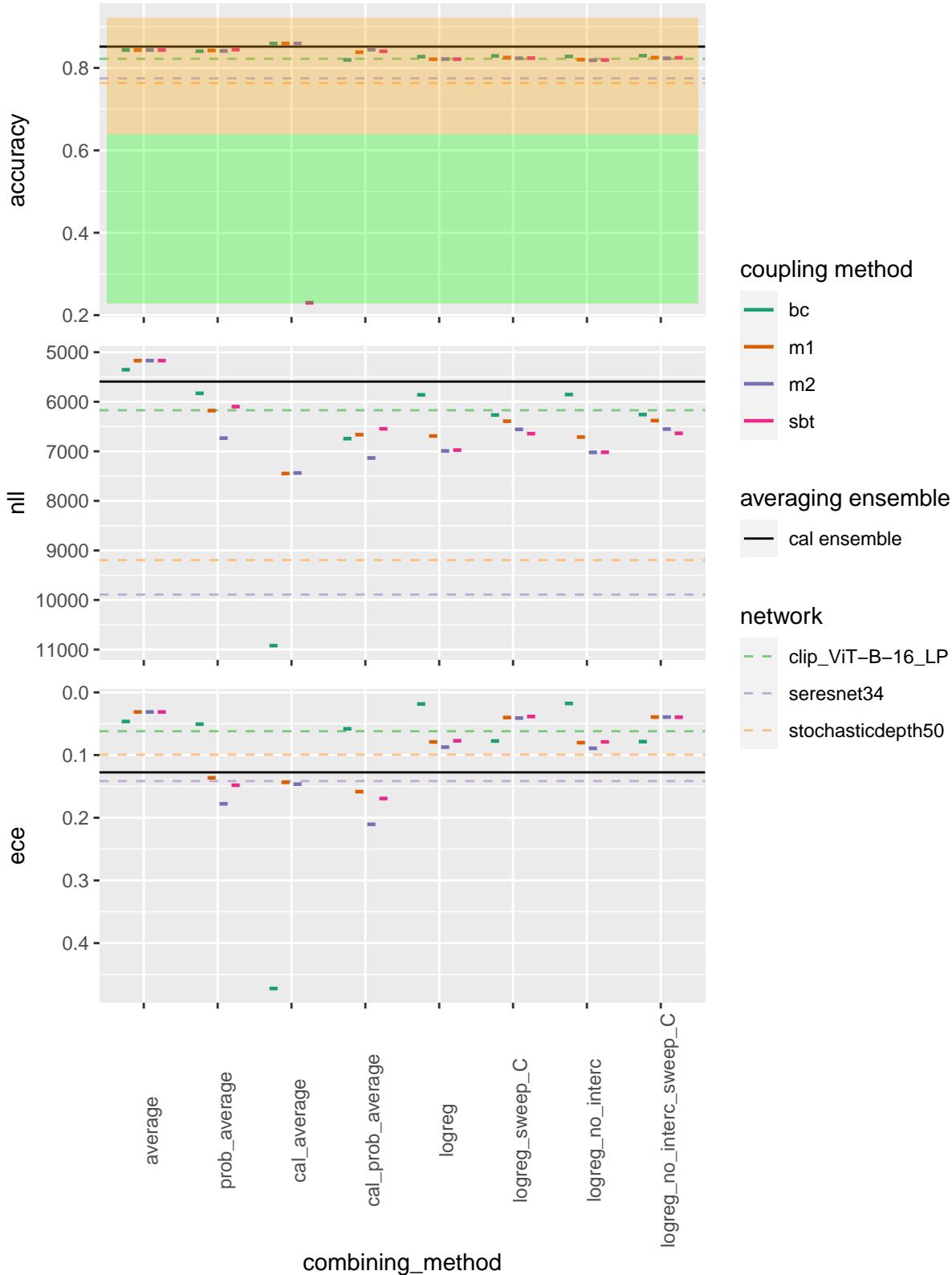
Ensemble metrics  
Error inconsistency 0.282700002193451



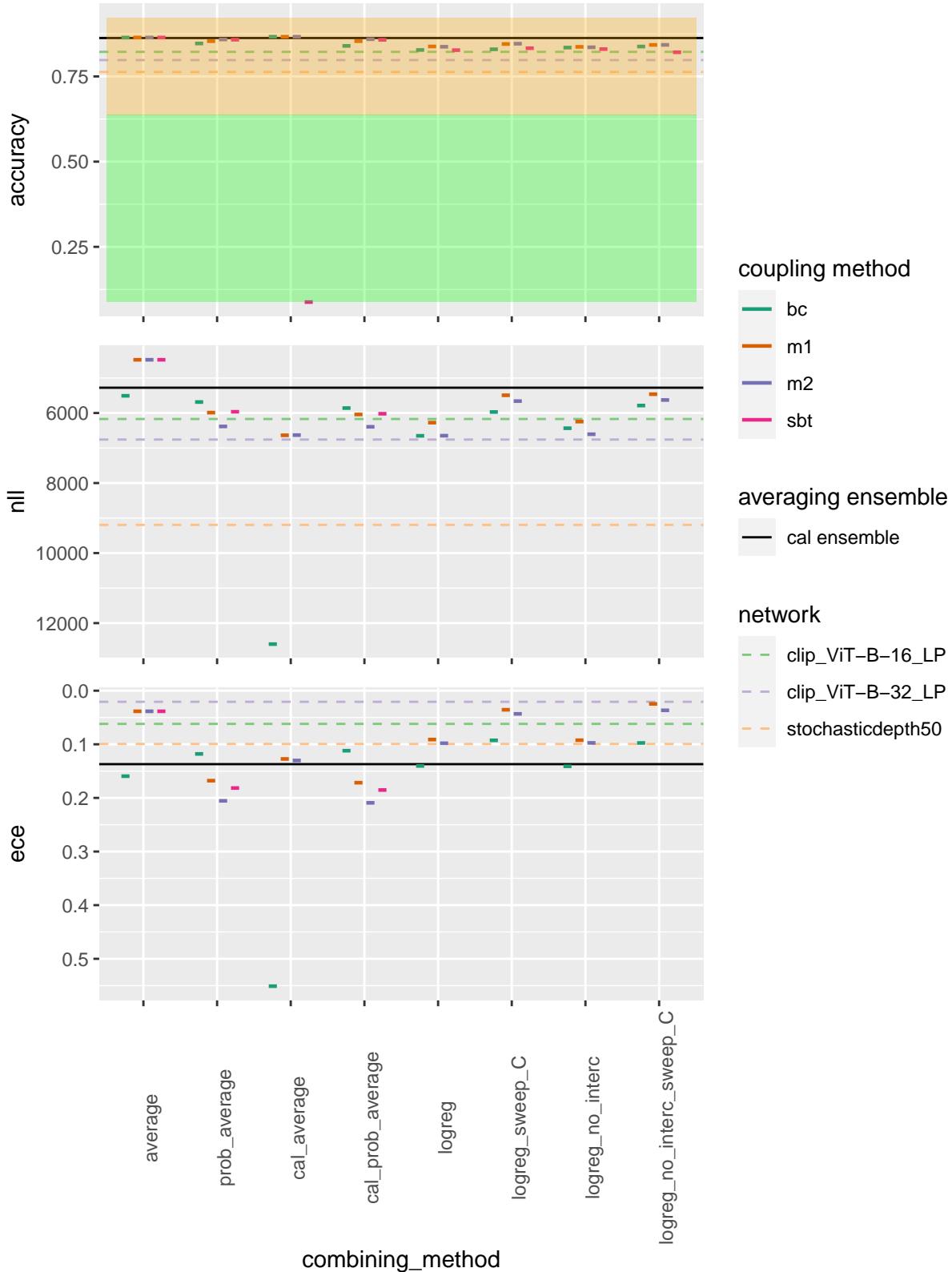
Ensemble metrics  
Error inconsistency 0.380899995565414



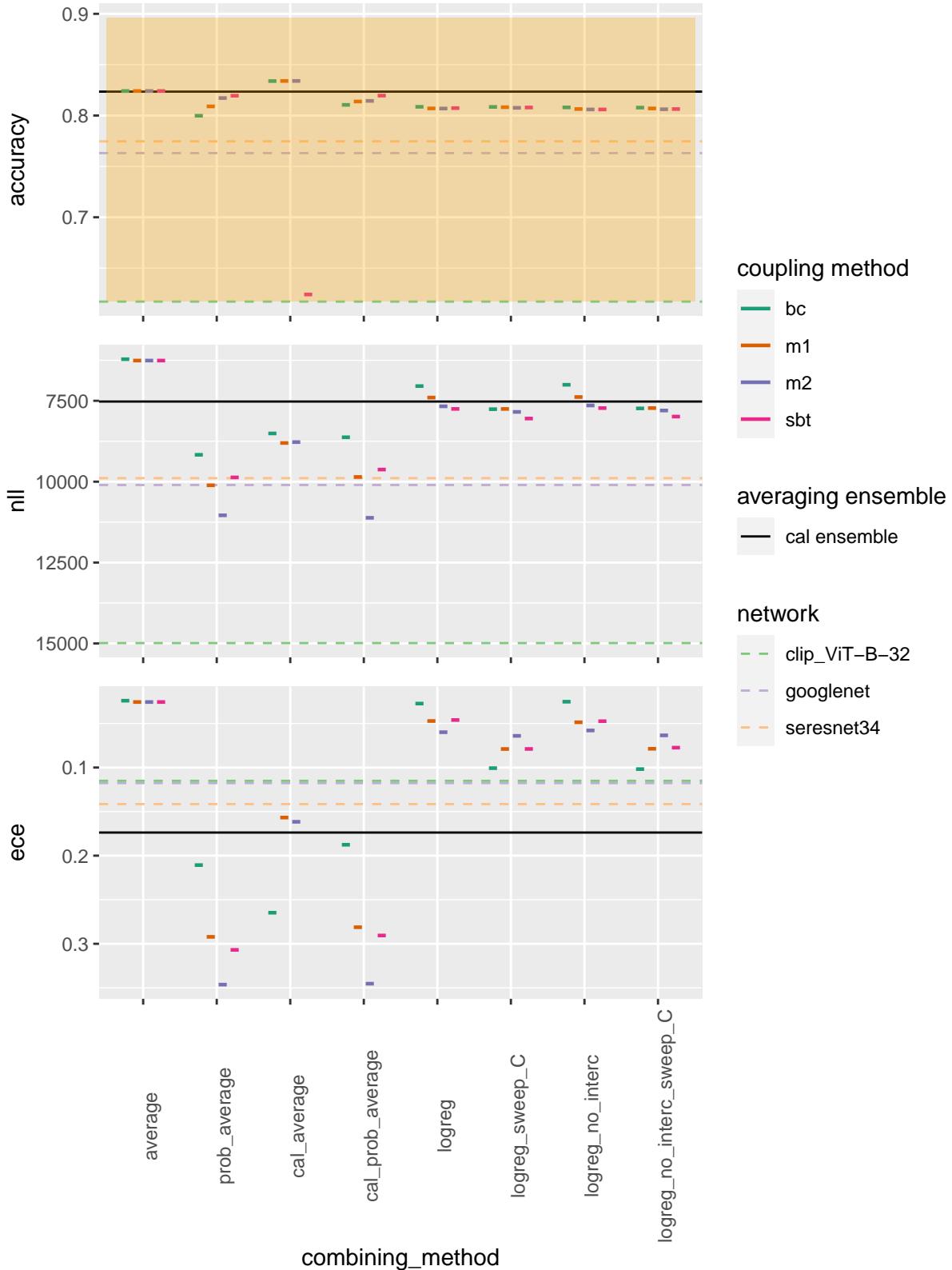
Ensemble metrics  
Error inconsistency 0.282299995422363



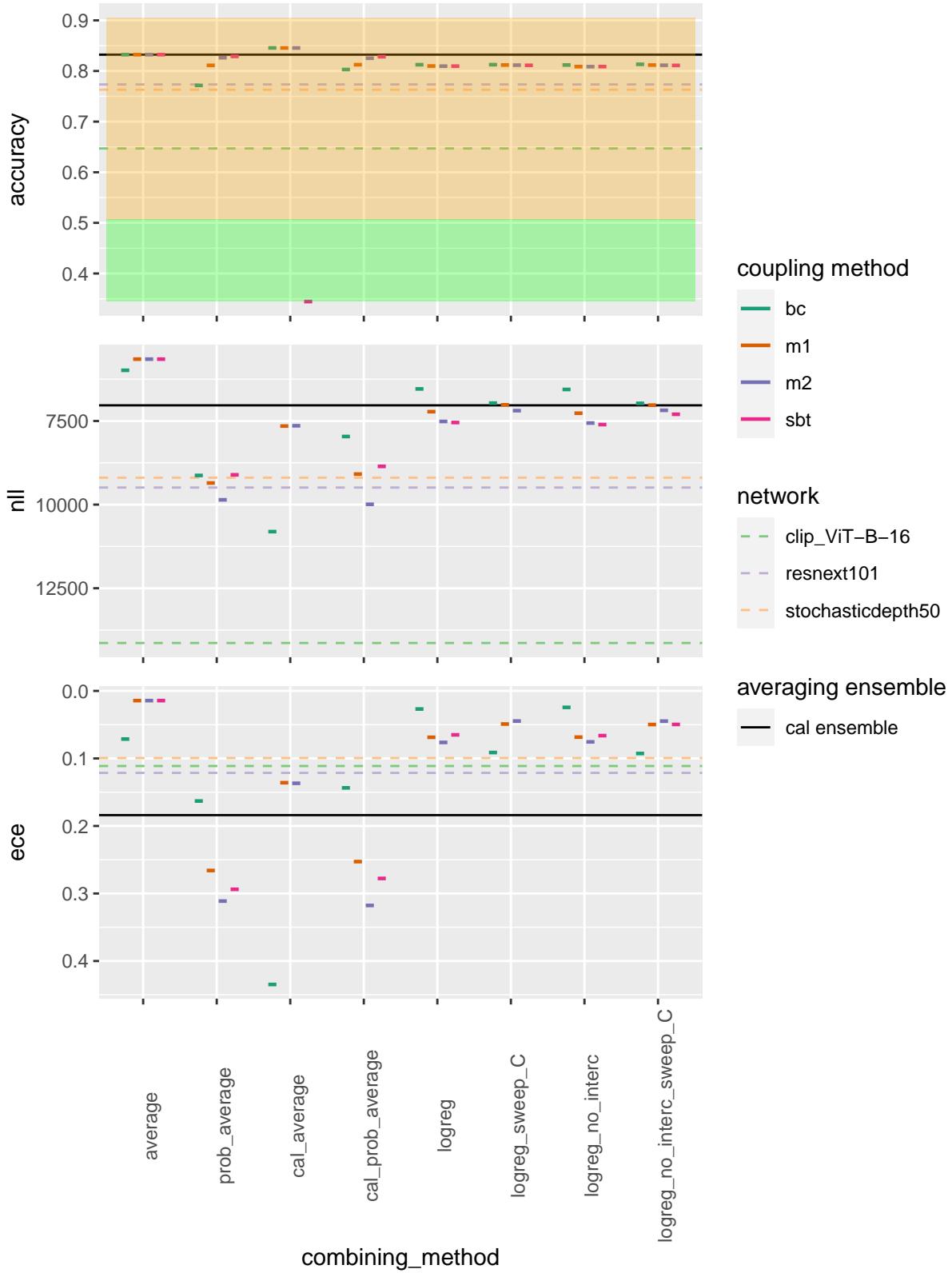
Ensemble metrics  
Error inconsistency 0.284999996423721



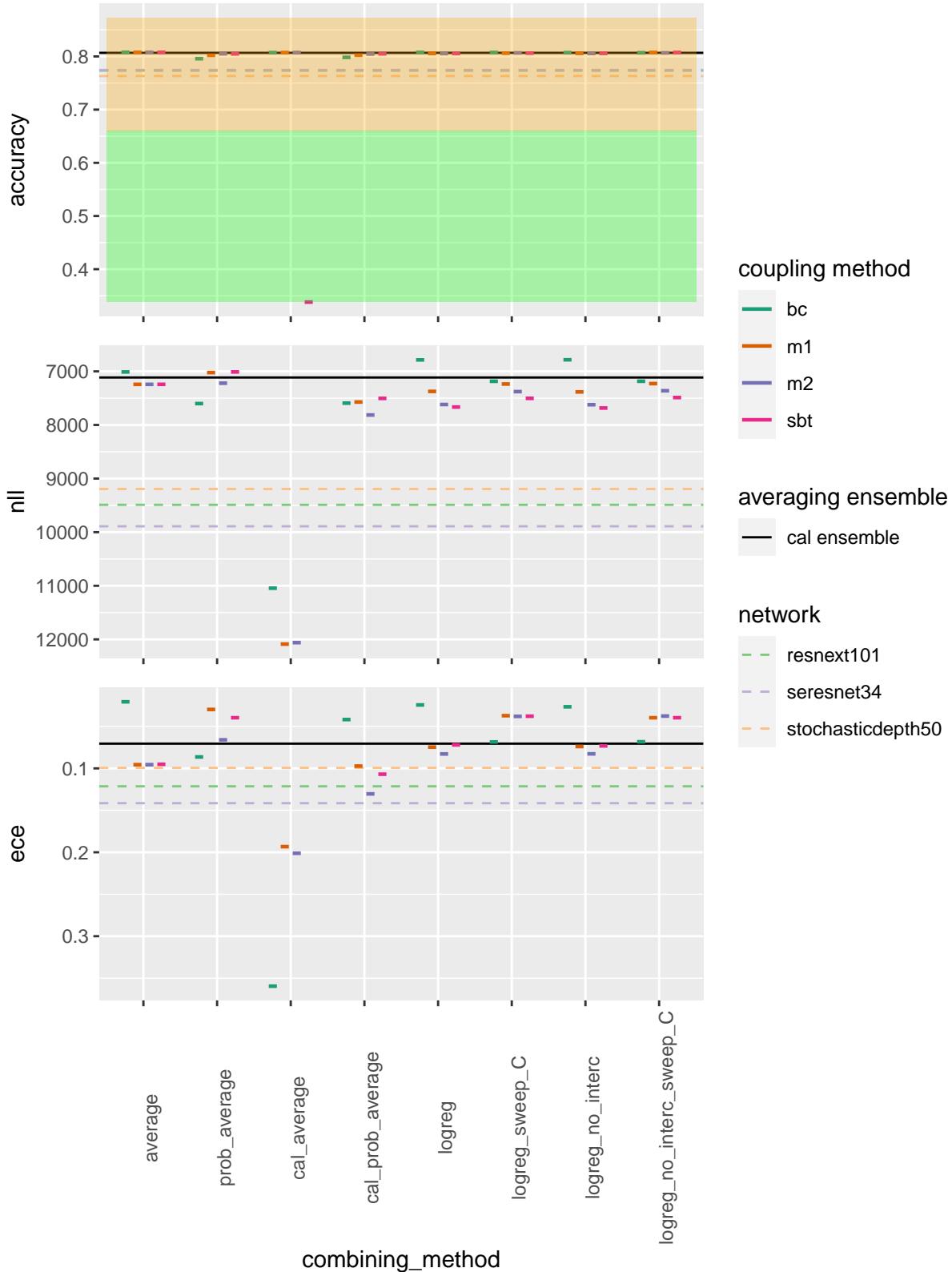
Ensemble metrics  
Error inconsistency 0.404499977827072



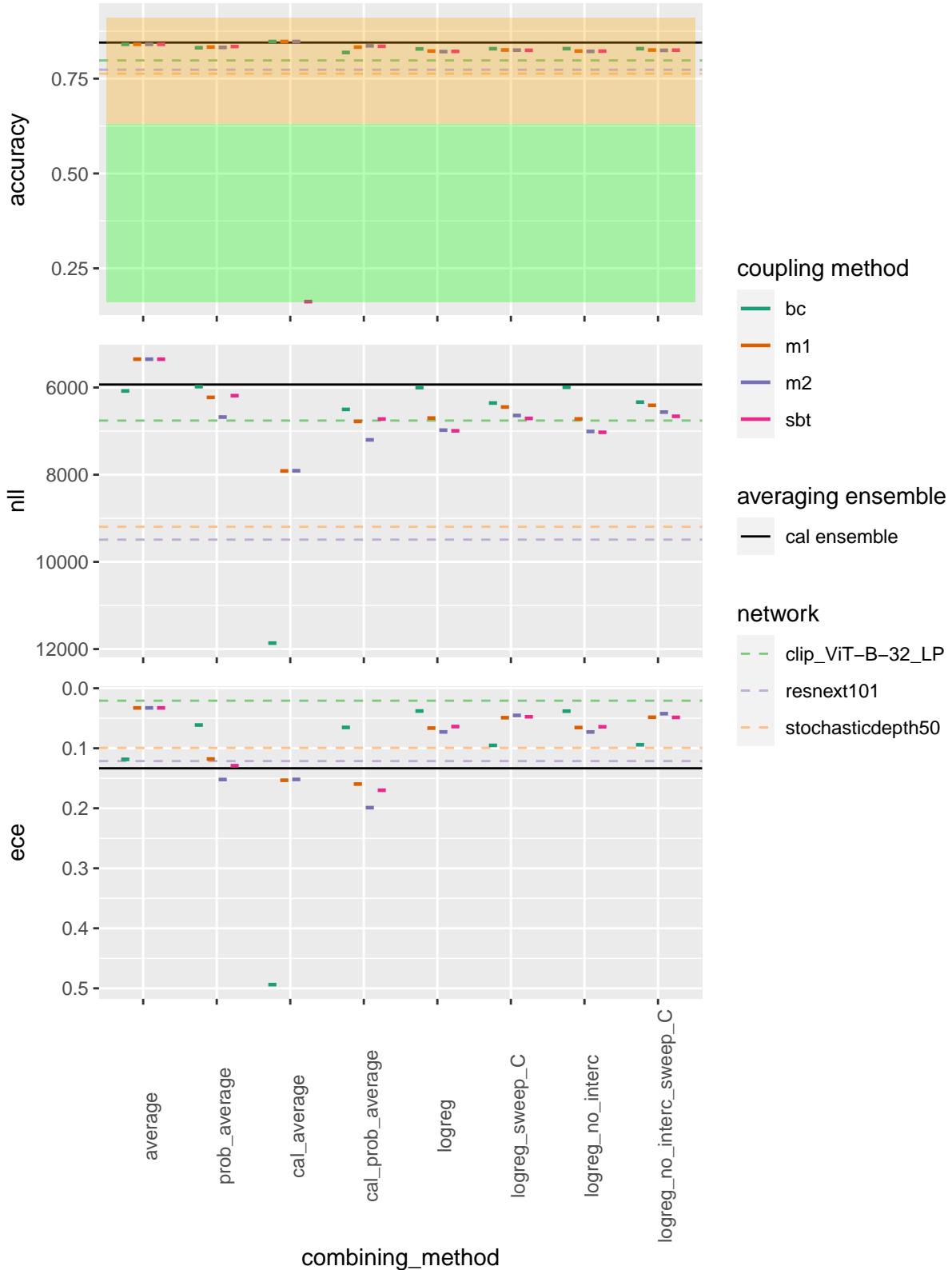
Ensemble metrics  
Error inconsistency 0.398599982261658



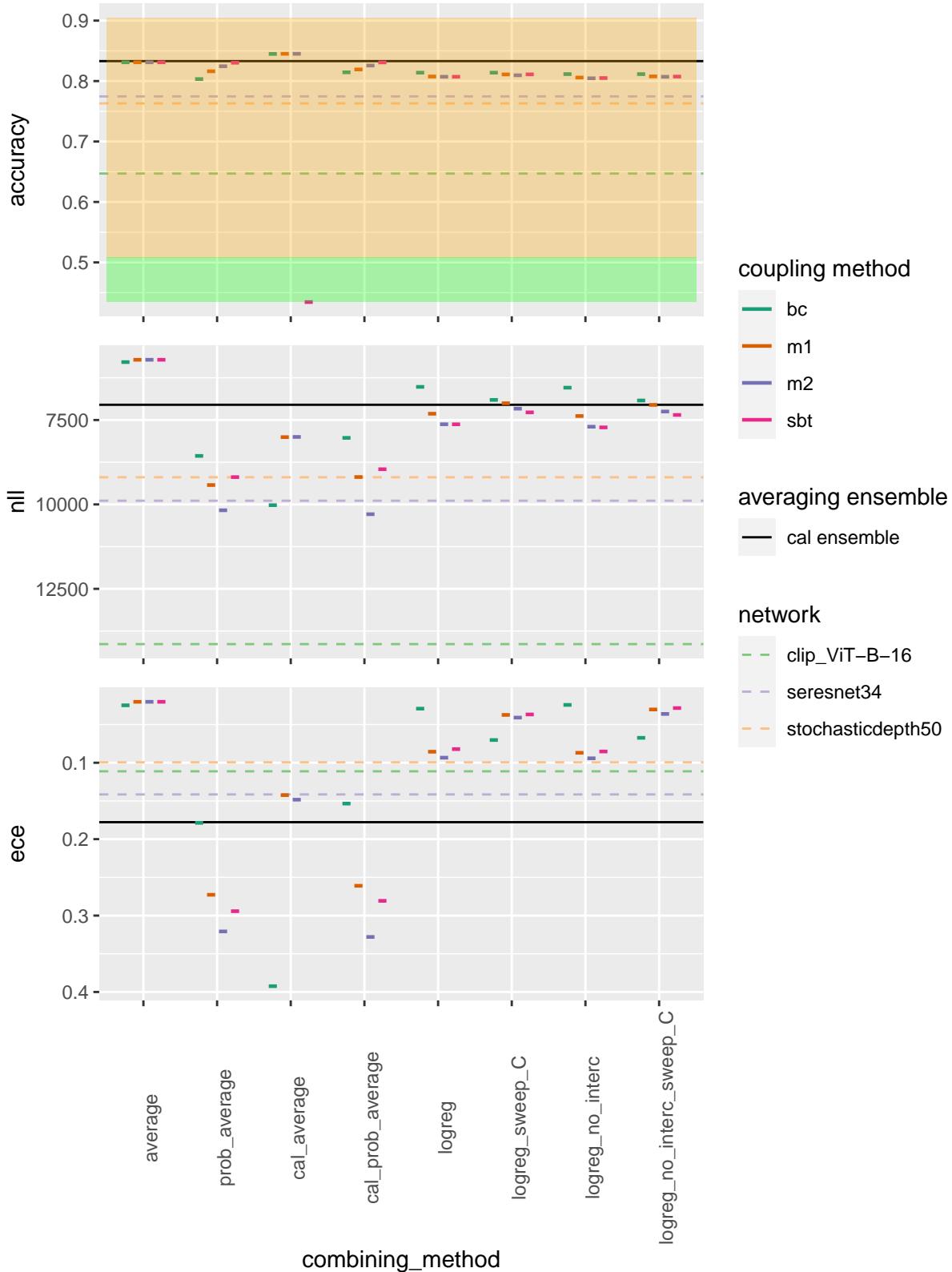
Ensemble metrics  
Error inconsistency 0.212799996137619



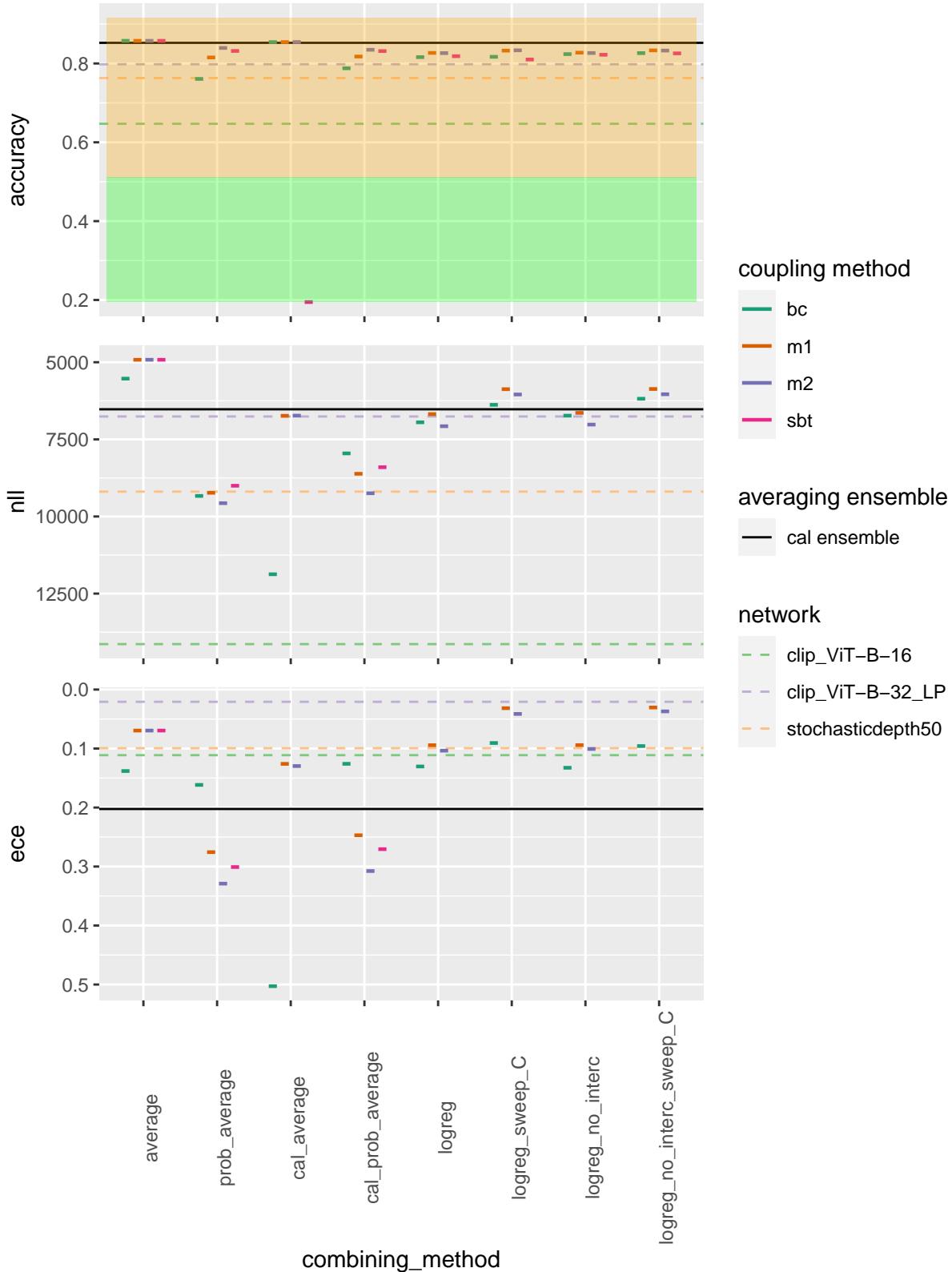
Ensemble metrics  
Error inconsistency 0.280099987983704



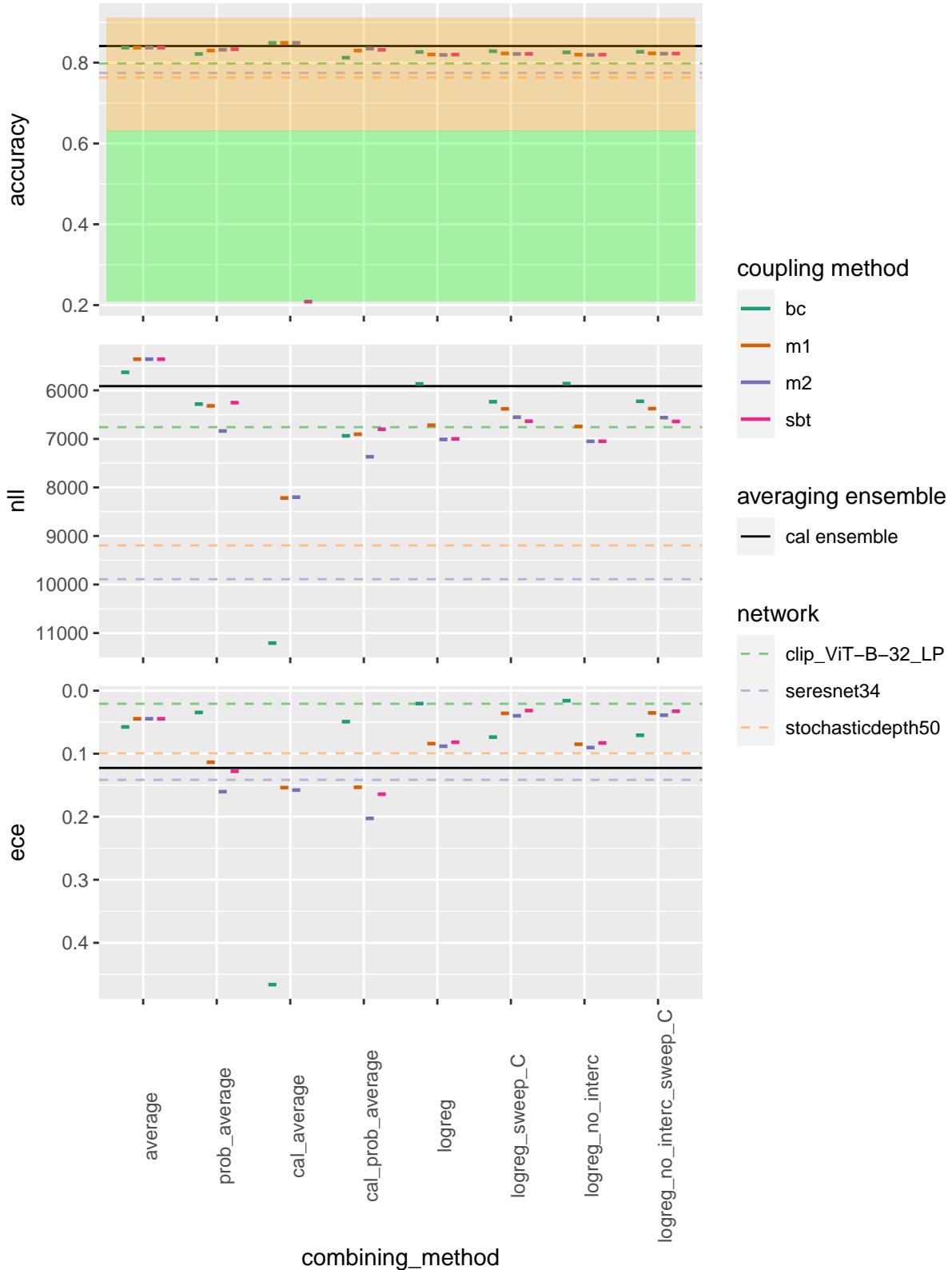
Ensemble metrics  
Error inconsistency 0.396200001239777



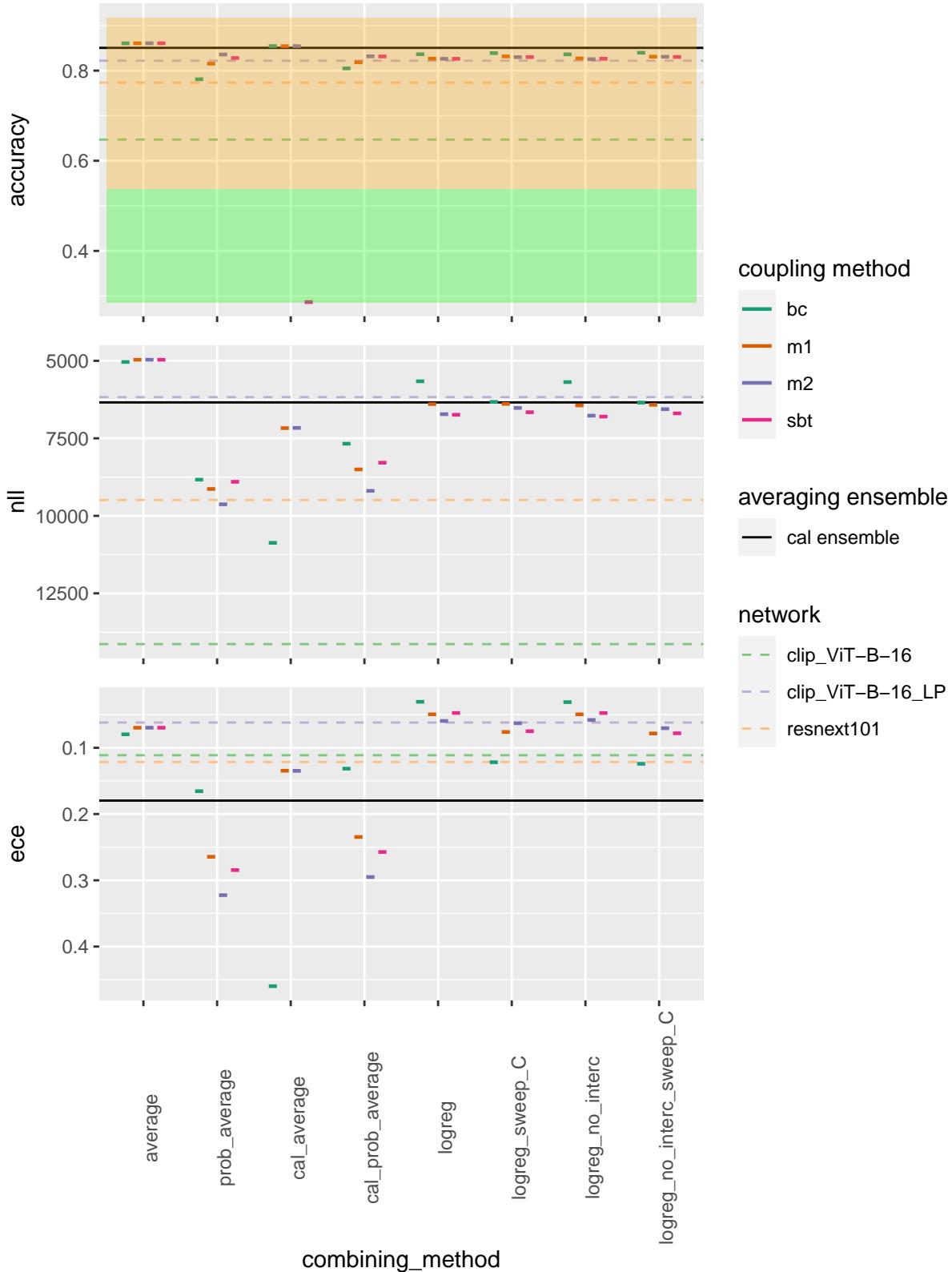
Ensemble metrics  
Error inconsistency 0.405399978160858



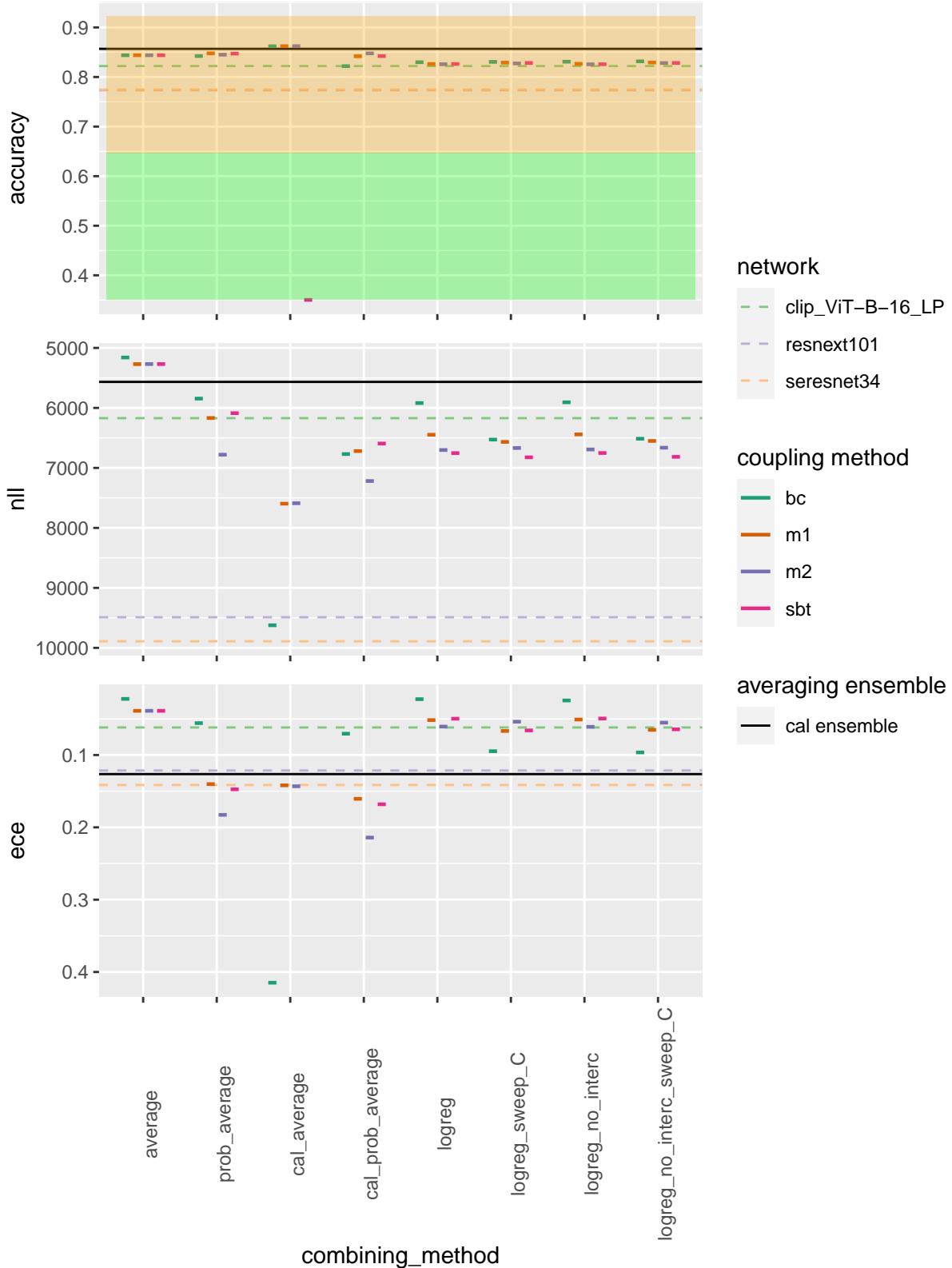
Ensemble metrics  
Error inconsistency 0.280000001192093



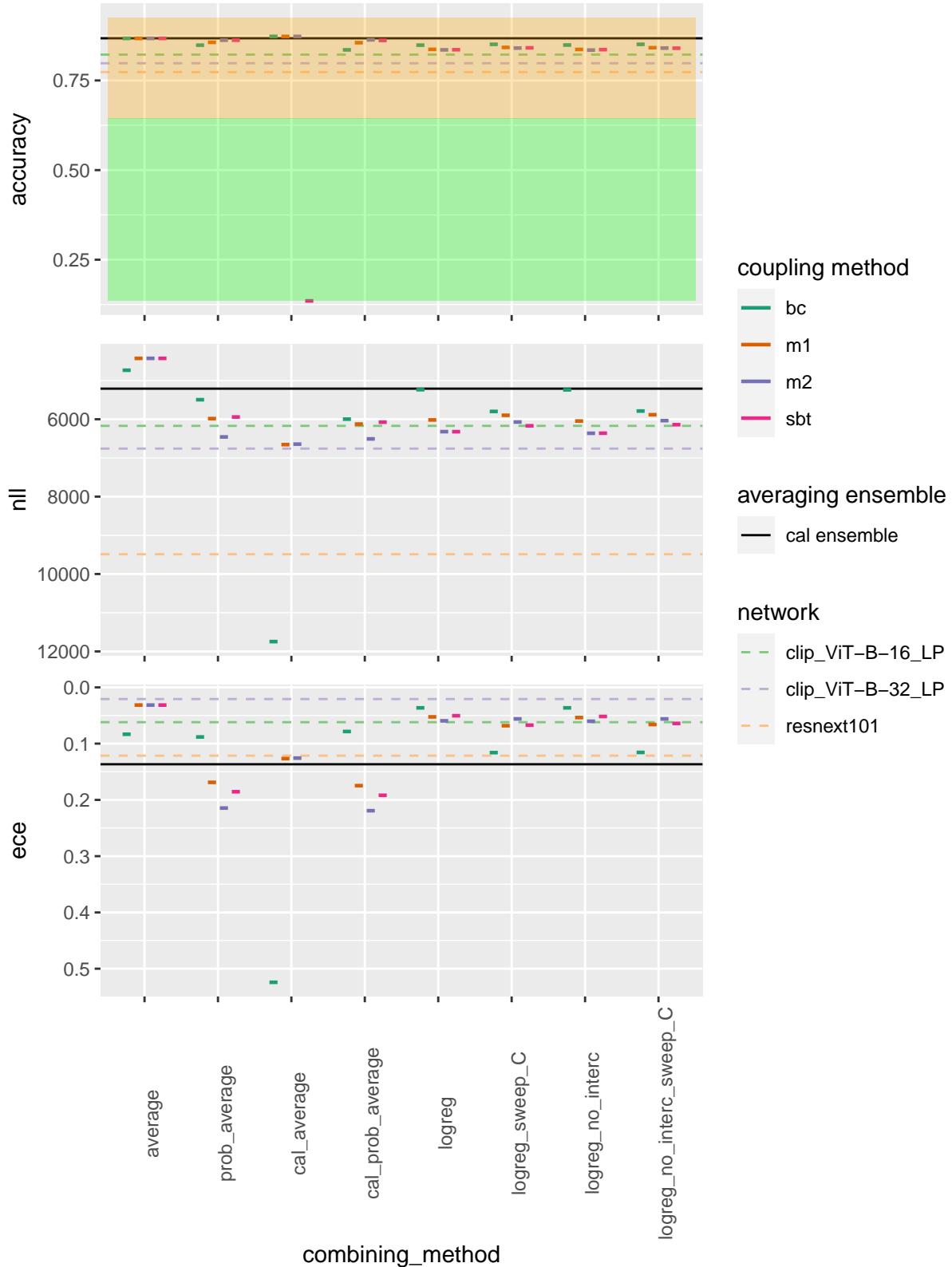
Ensemble metrics  
Error inconsistency 0.380199998617172



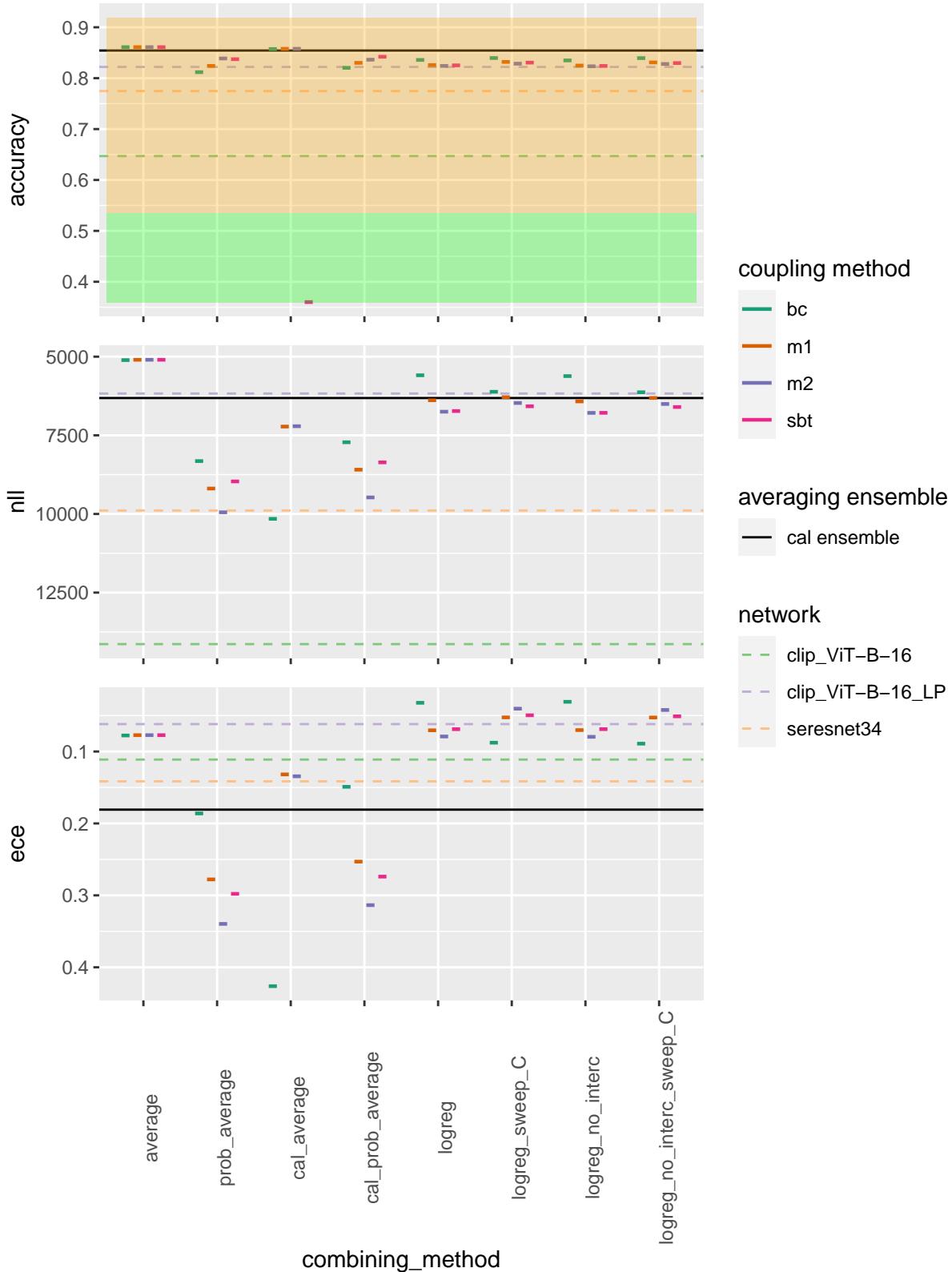
Ensemble metrics  
Error inconsistency 0.276600003242493



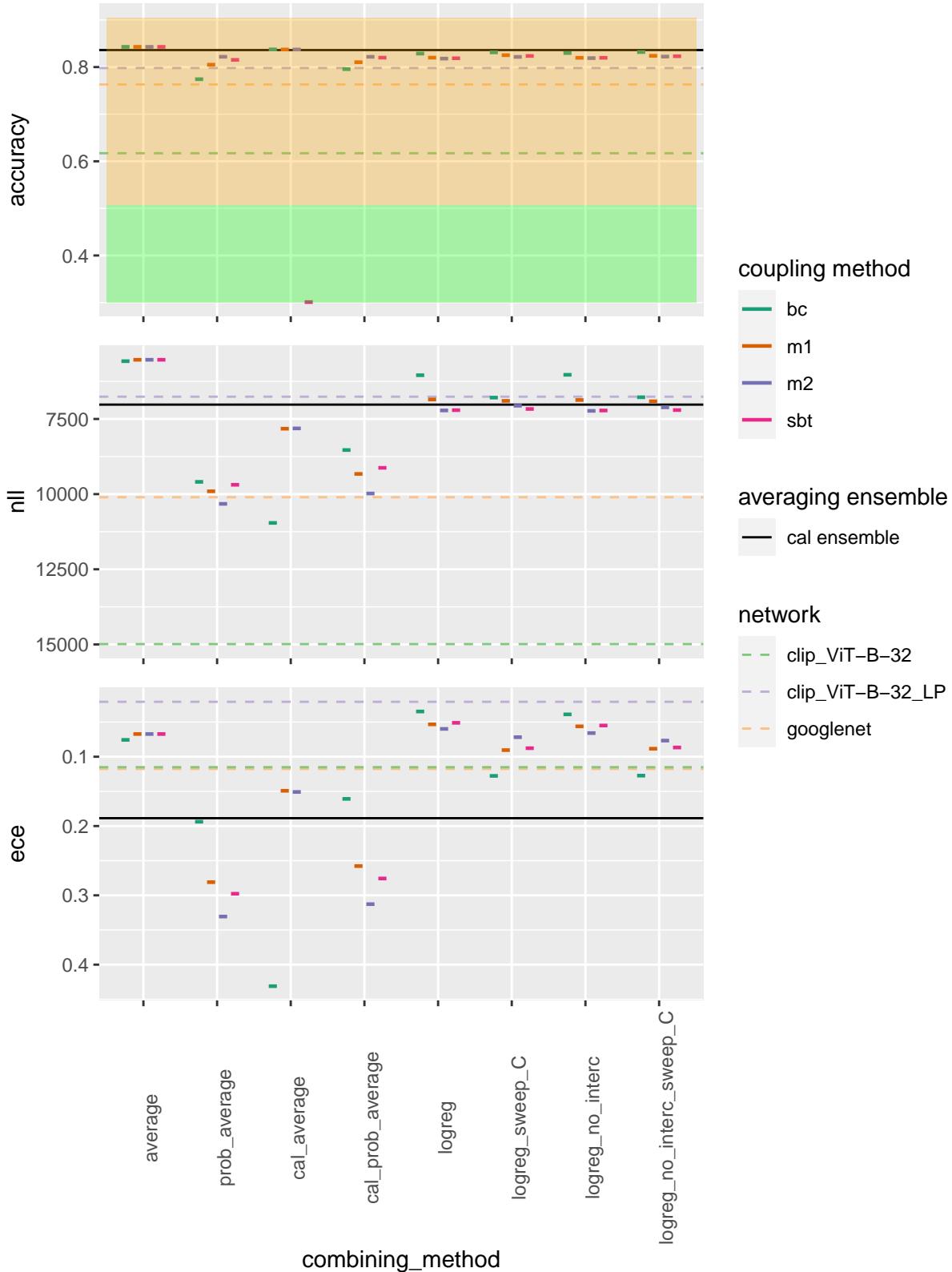
Ensemble metrics  
Error inconsistency 0.28099998831749



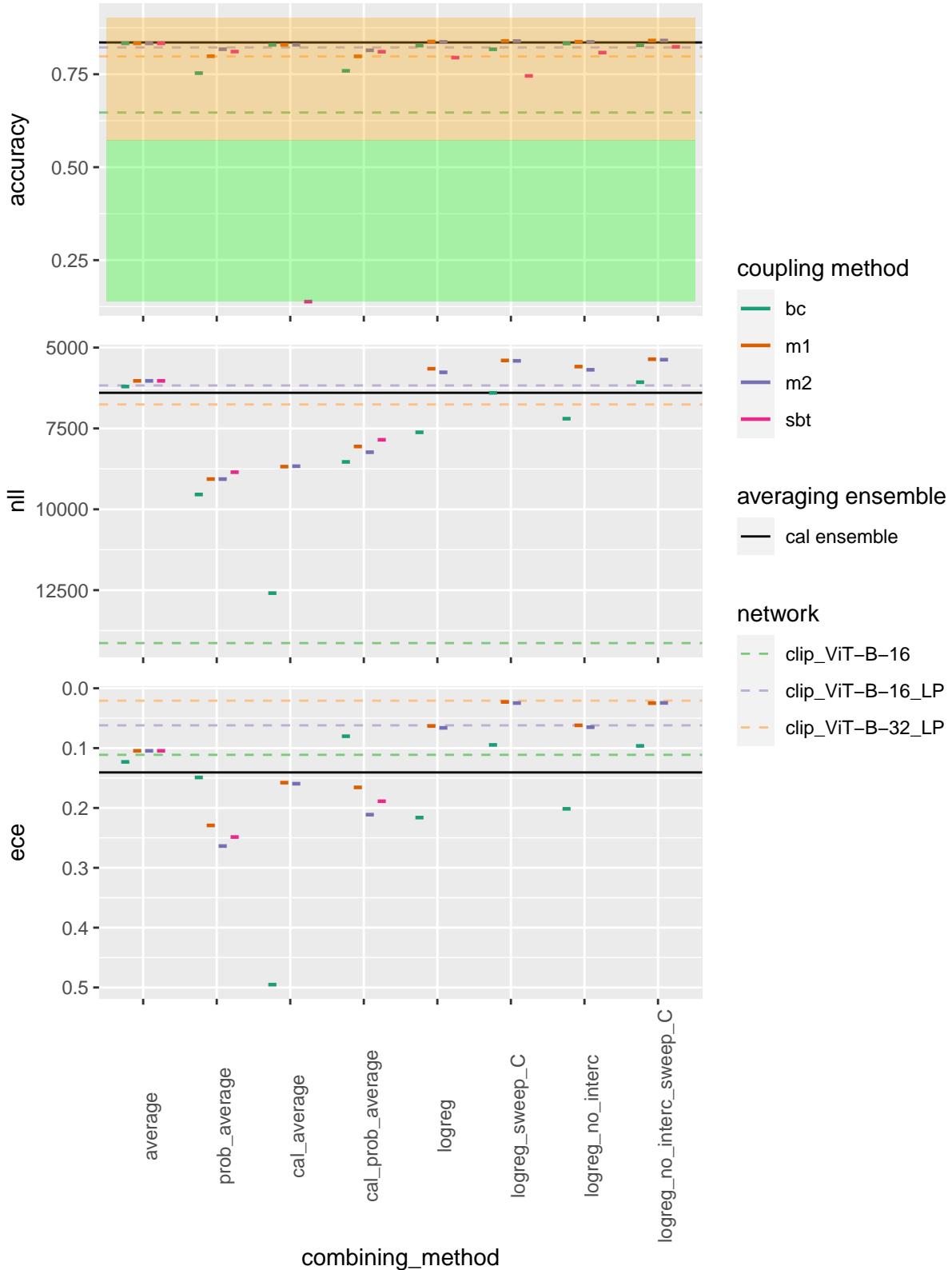
Ensemble metrics  
Error inconsistency 0.38400000333786



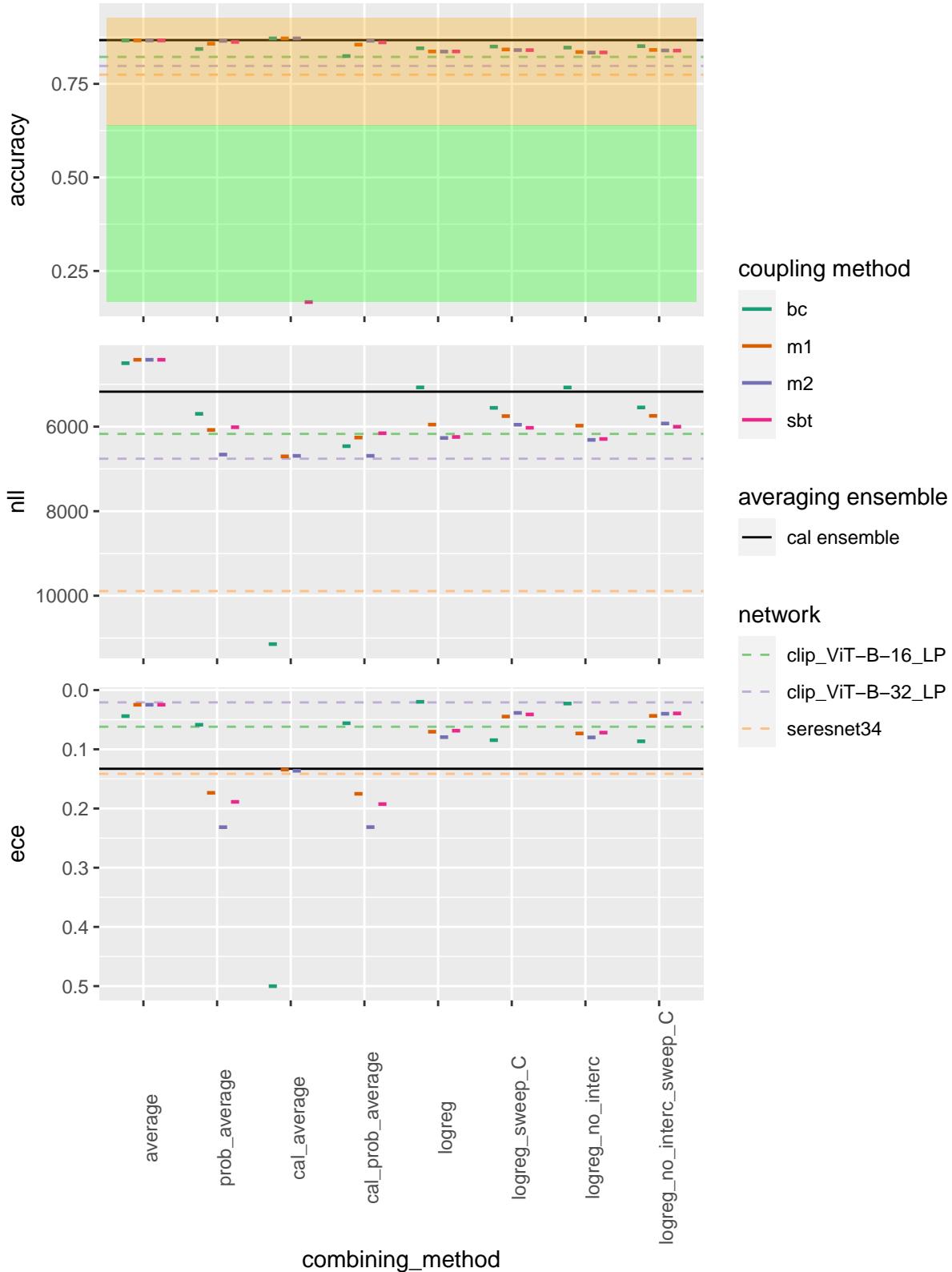
Ensemble metrics  
 Error inconsistency 0.397699981927872



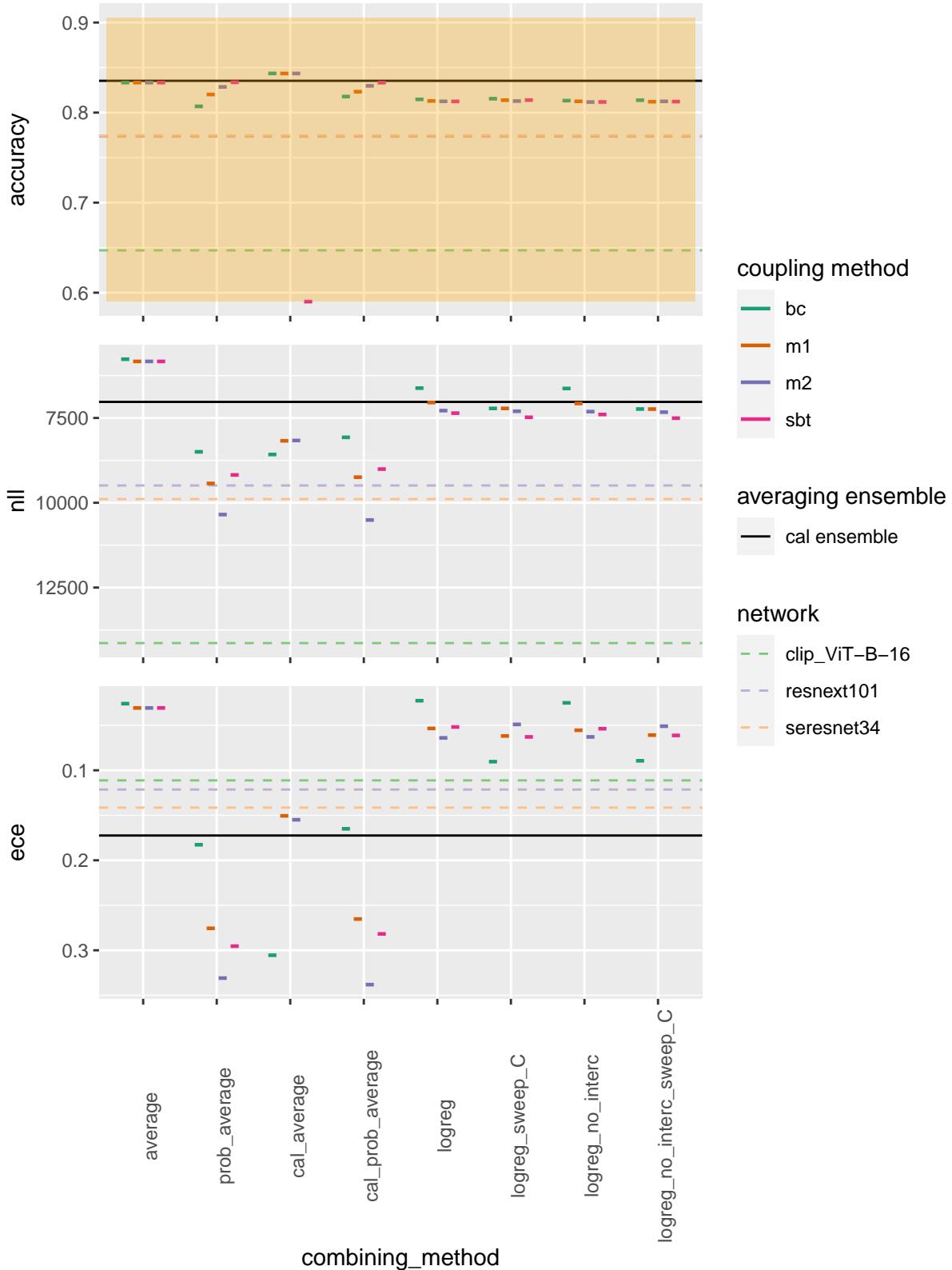
Ensemble metrics  
 Error inconsistency 0.330300003290176



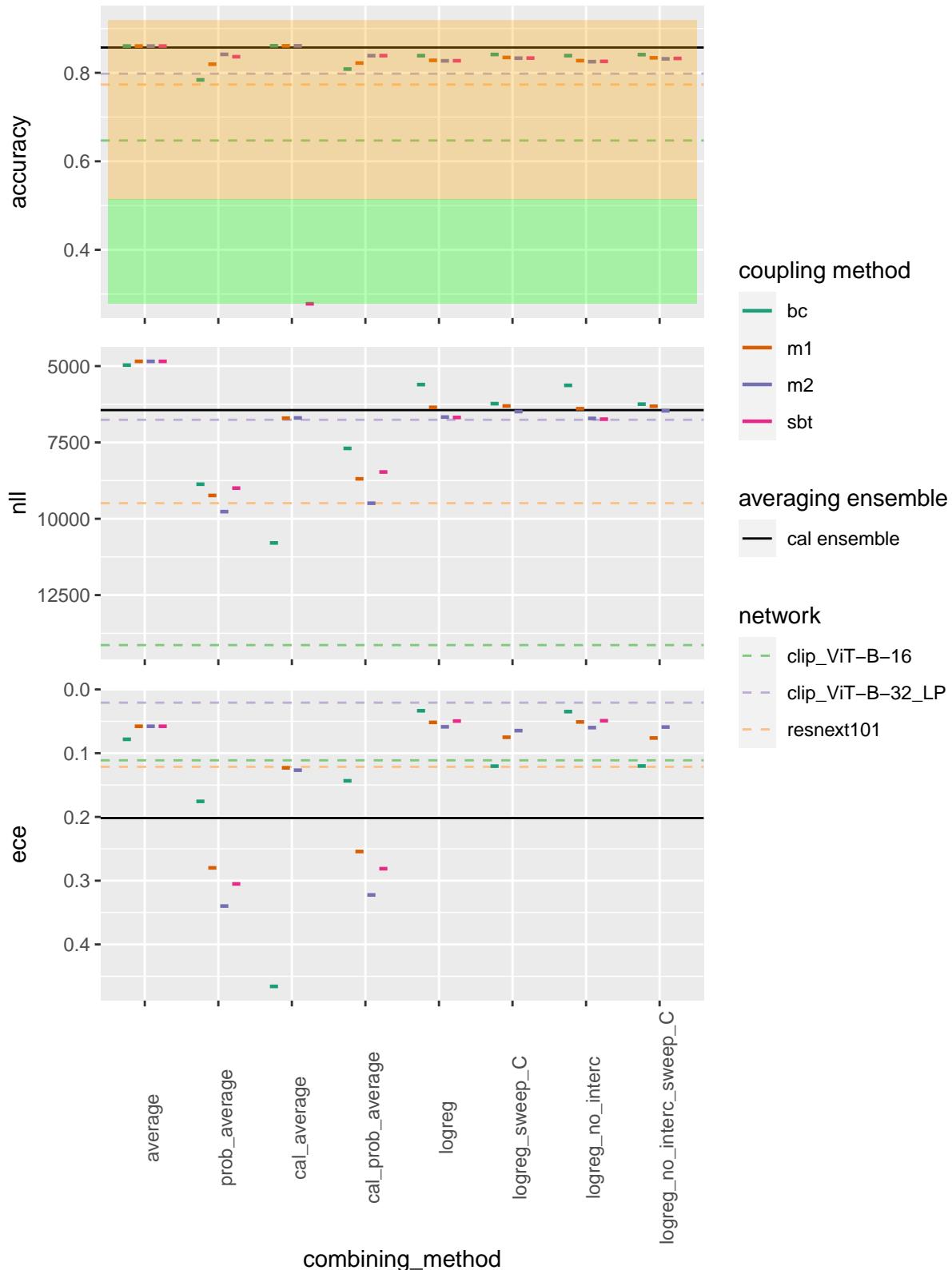
Ensemble metrics  
Error inconsistency 0.287099987268448



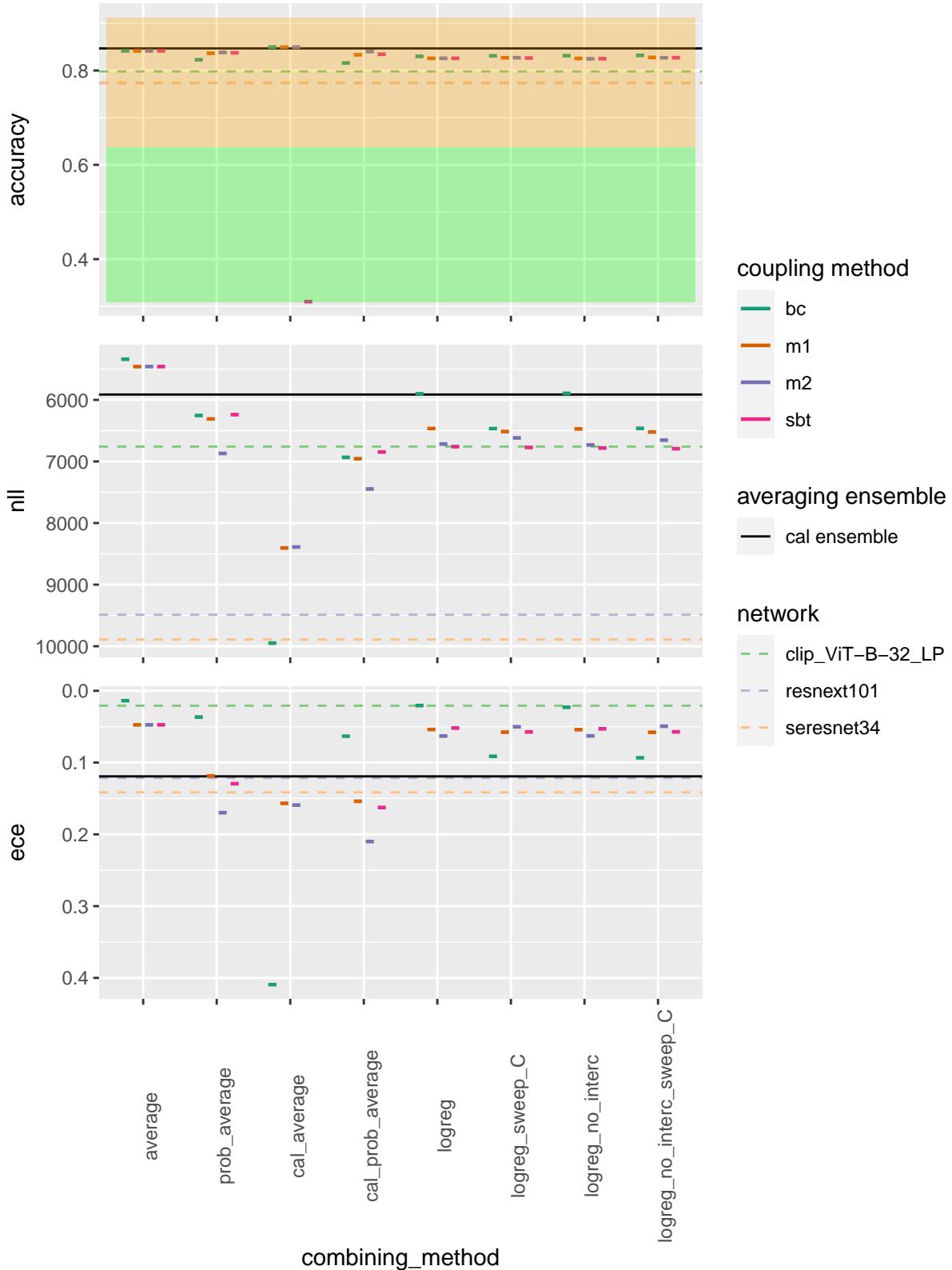
Ensemble metrics  
Error inconsistency 0.393899977207184



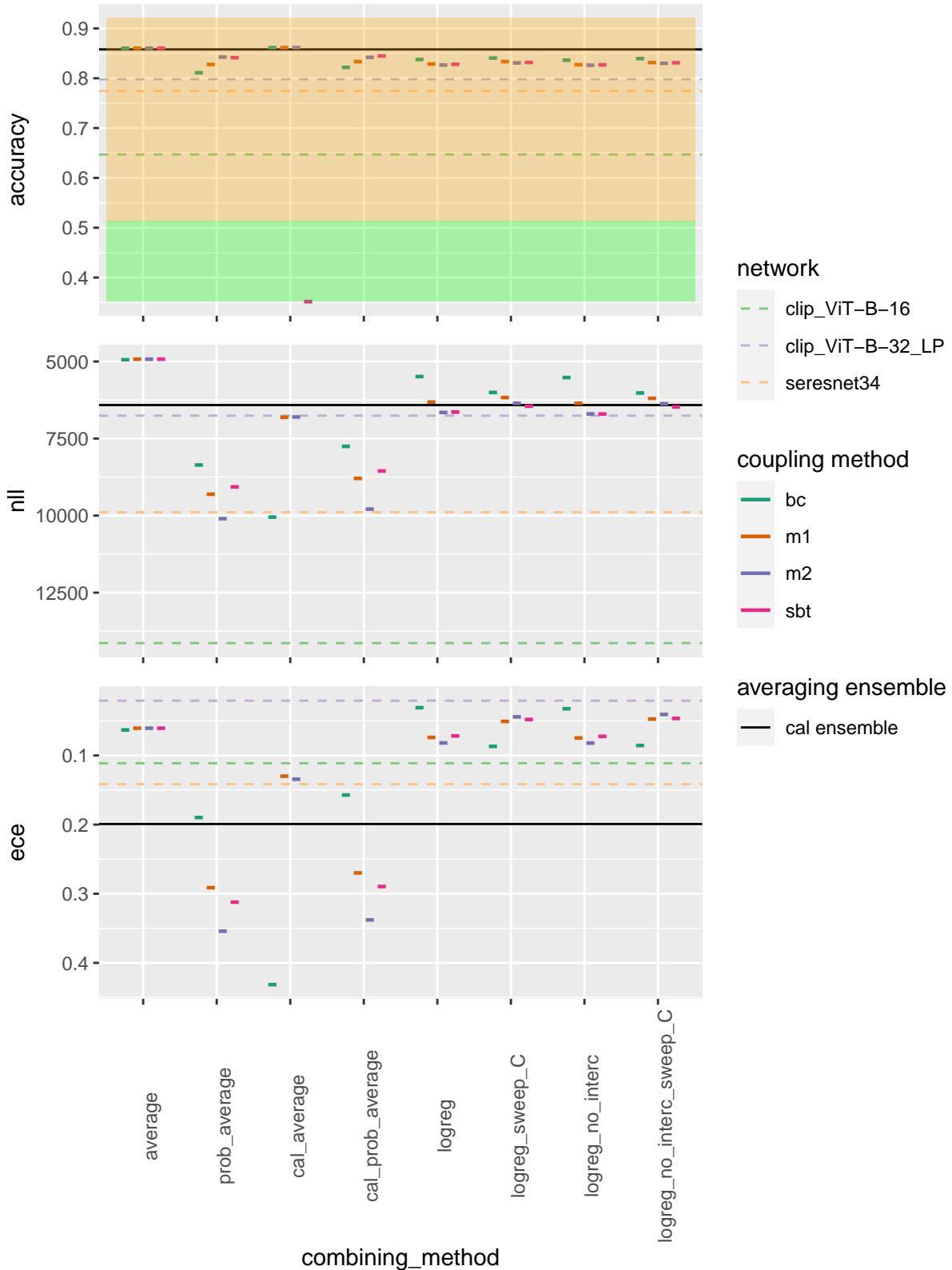
Ensemble metrics  
Error inconsistency 0.404799997806549



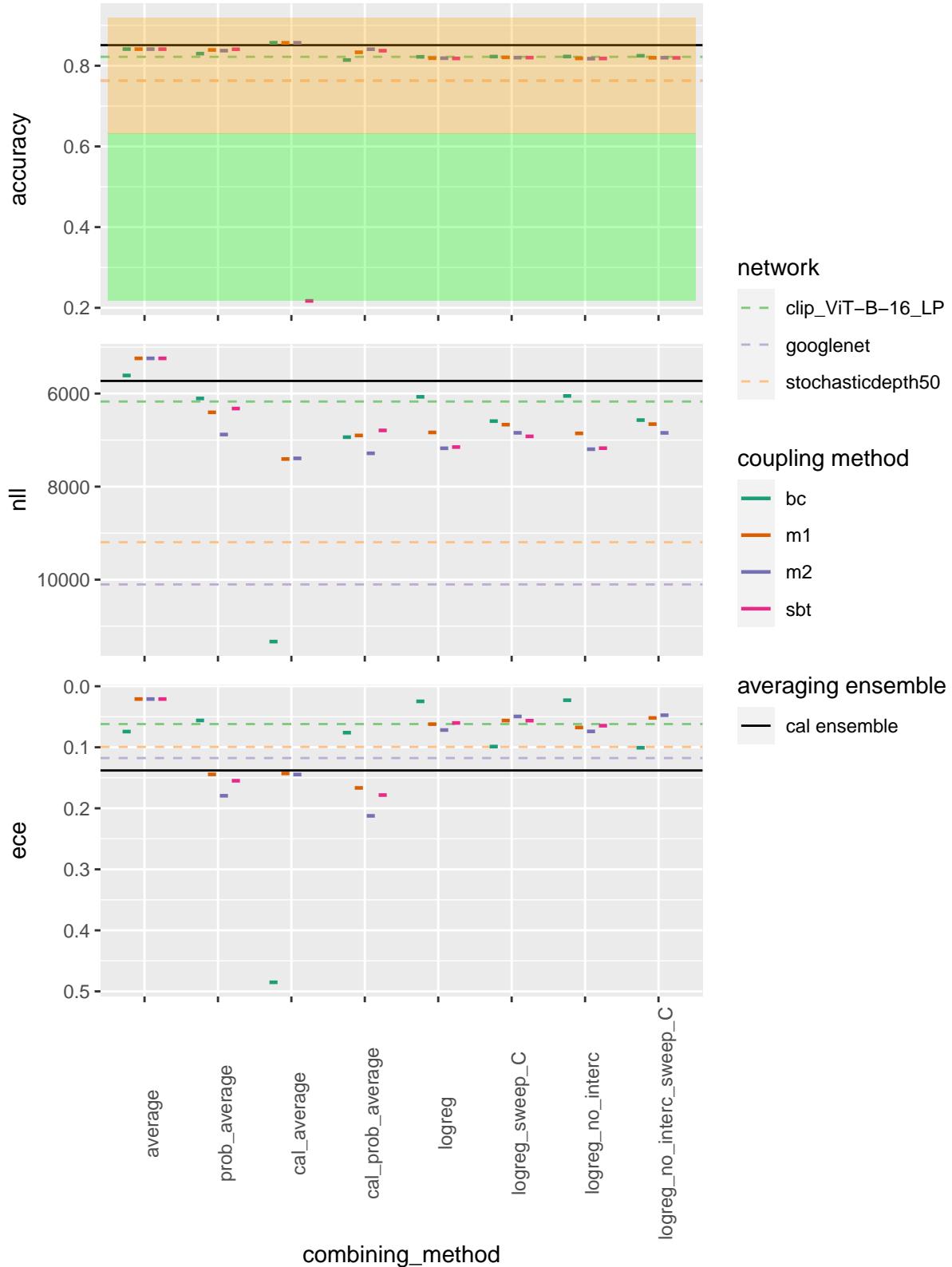
Ensemble metrics  
Error inconsistency 0.274399995803833



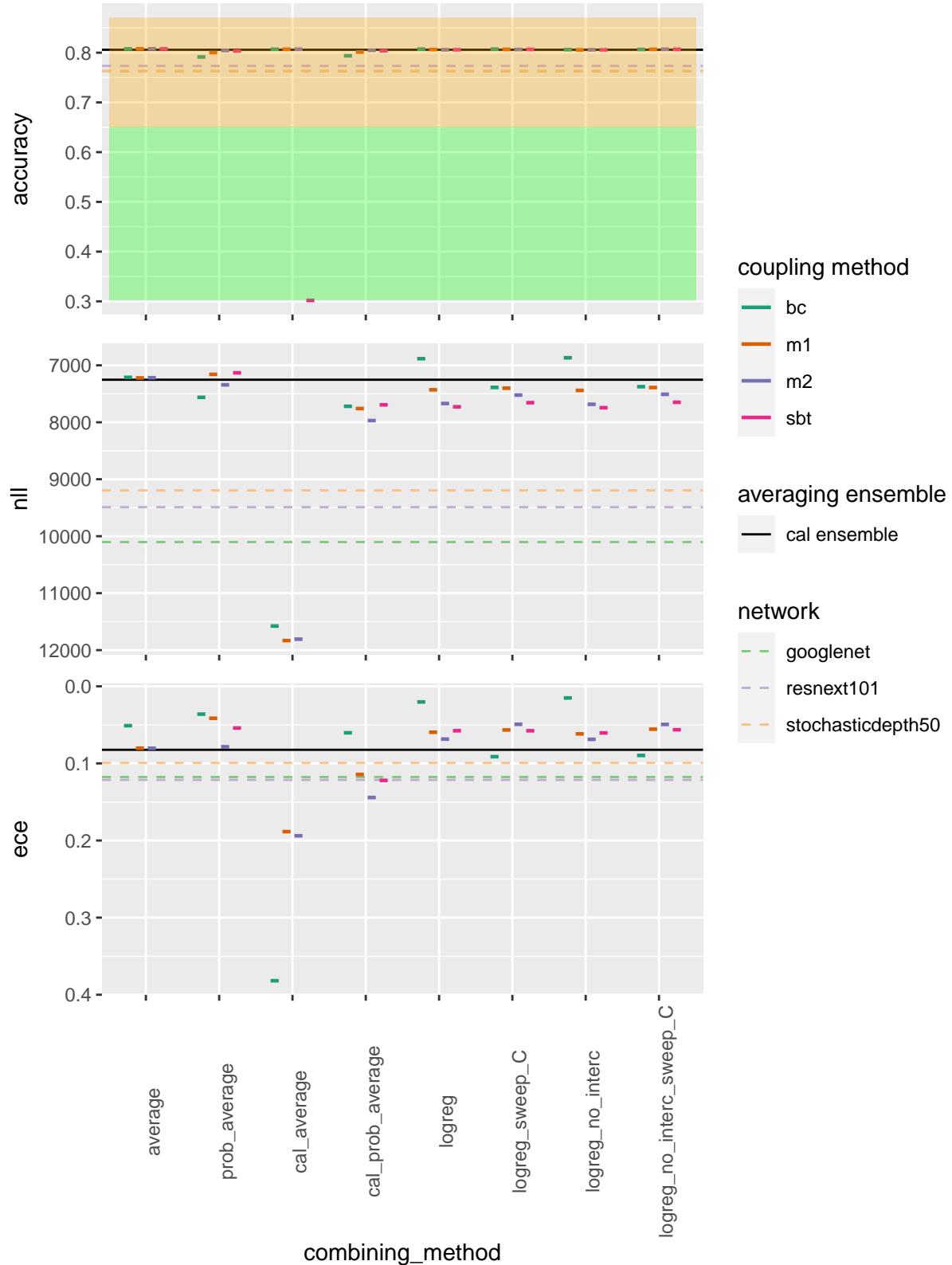
Ensemble metrics  
Error inconsistency 0.408899992704391



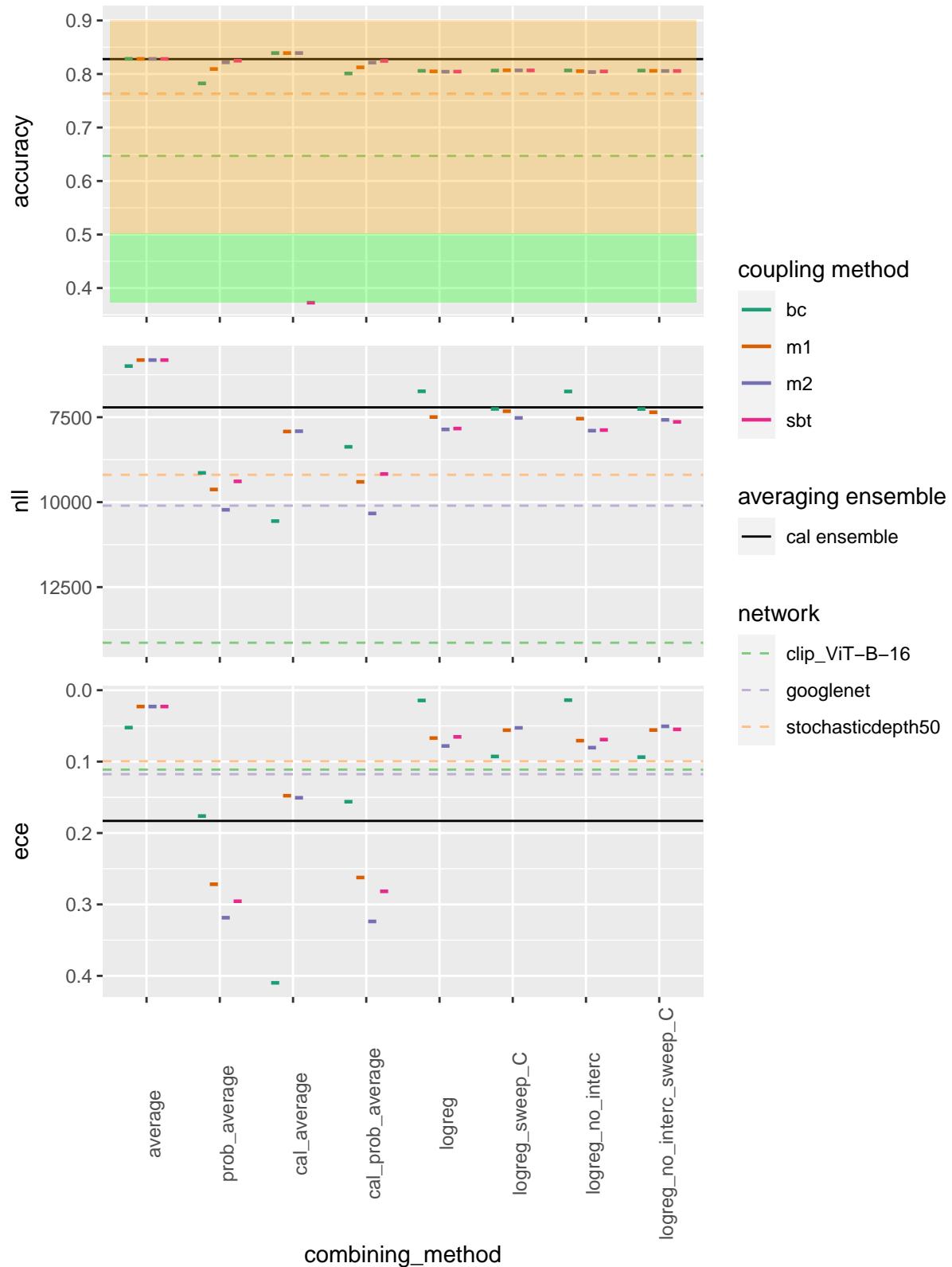
Ensemble metrics  
Error inconsistency 0.286799997091293



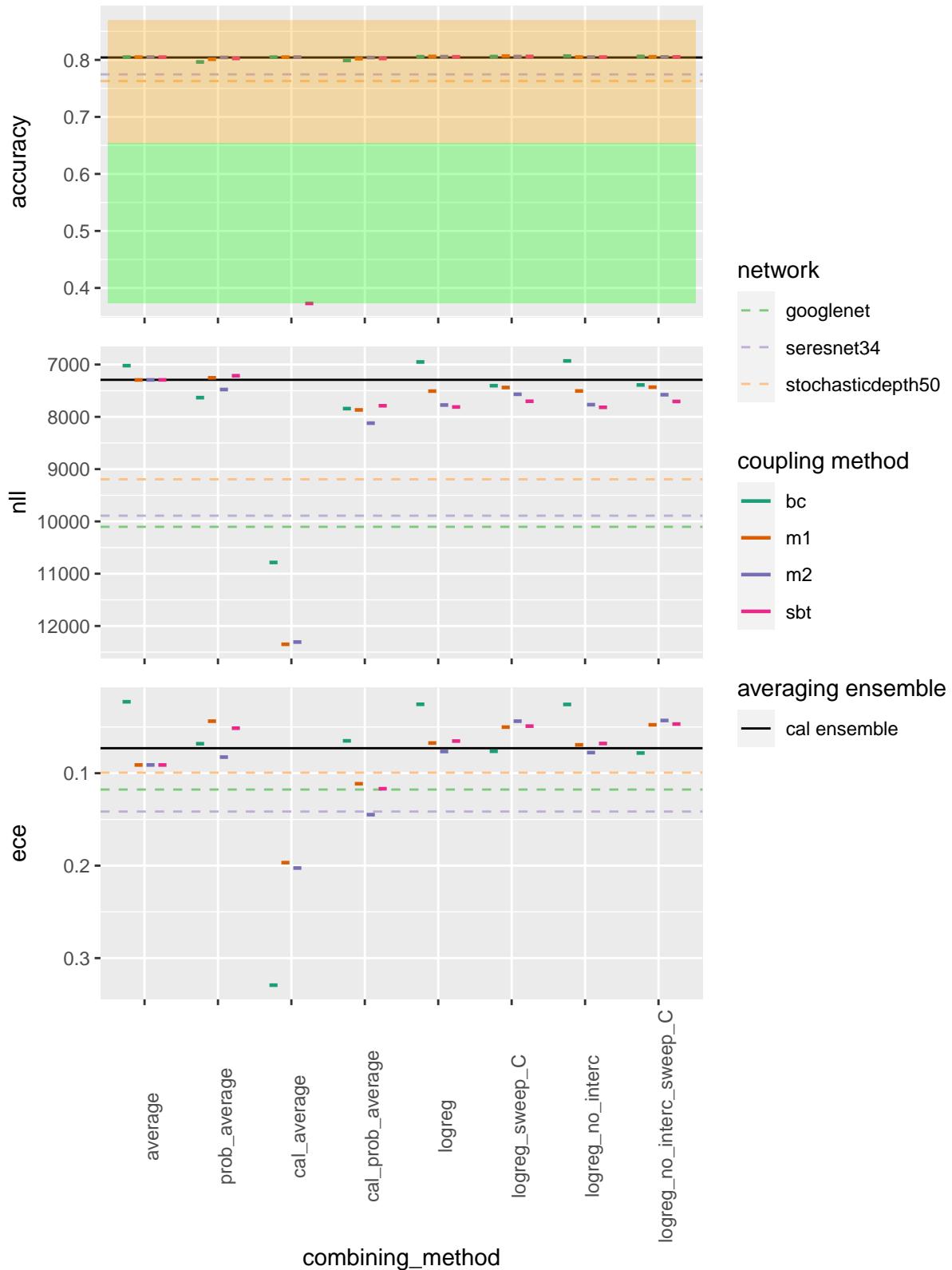
Ensemble metrics  
Error inconsistency 0.220100000500679



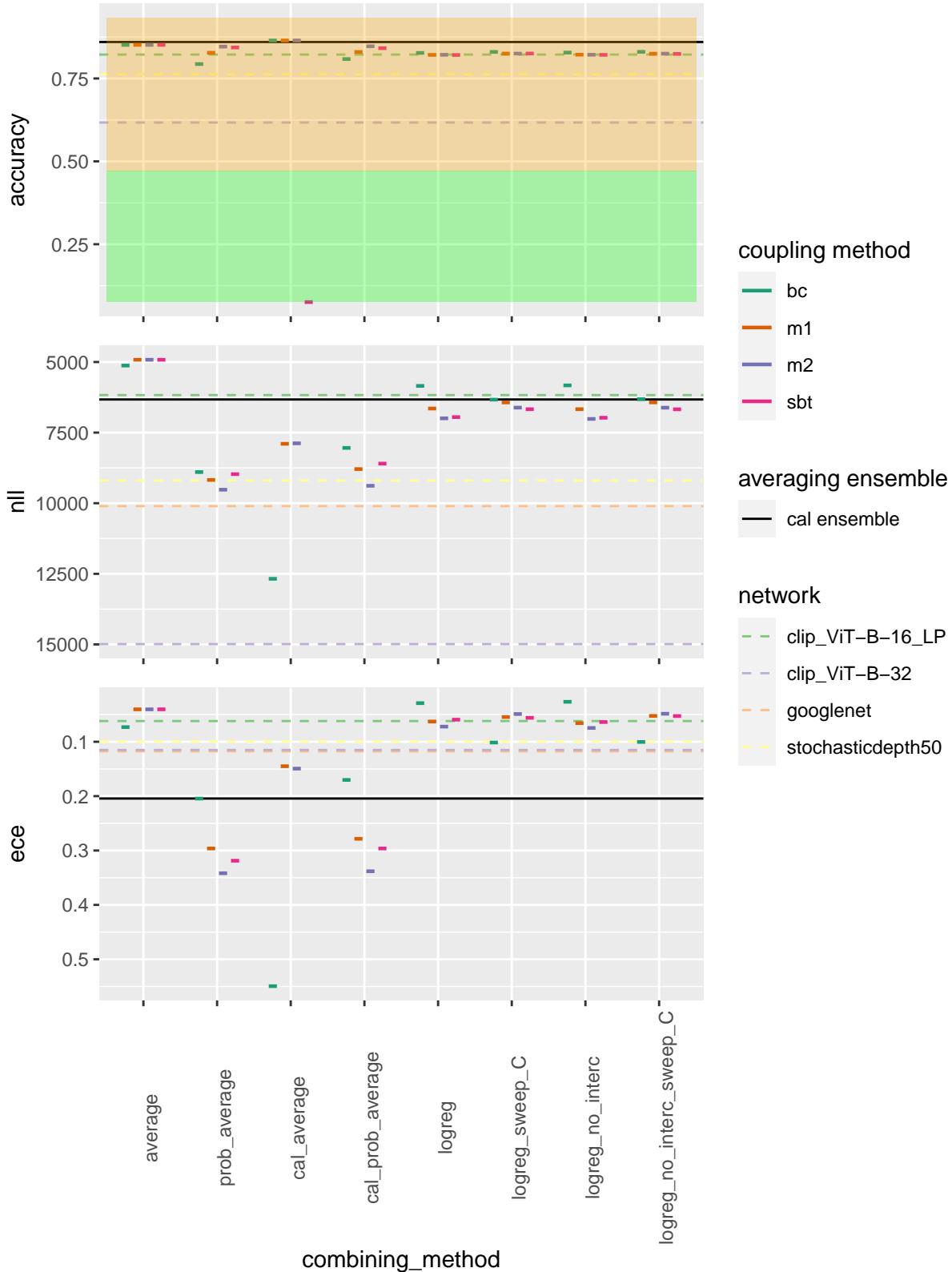
Ensemble metrics  
Error inconsistency 0.399199992418289



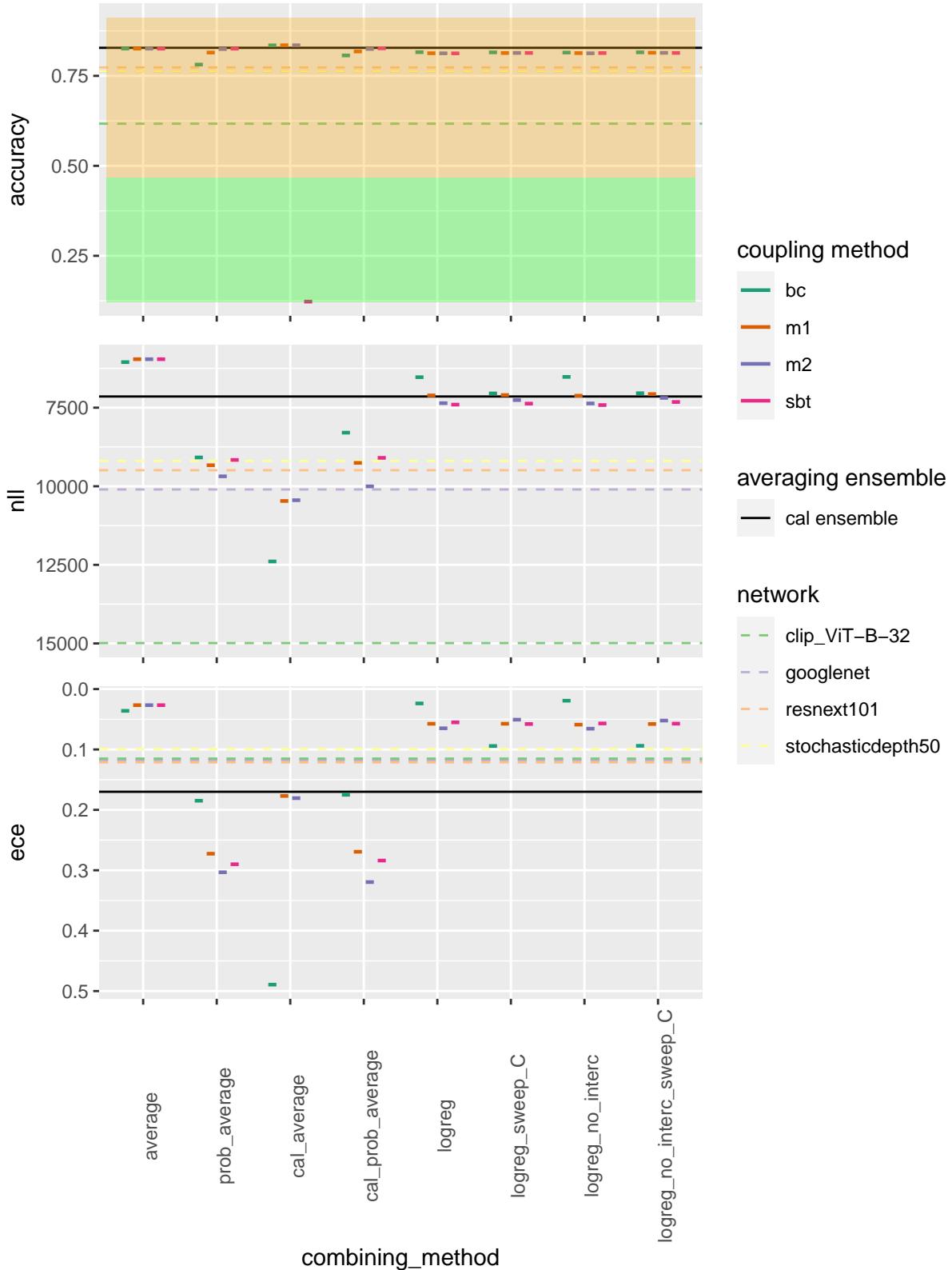
Ensemble metrics  
Error inconsistency 0.215799987316132



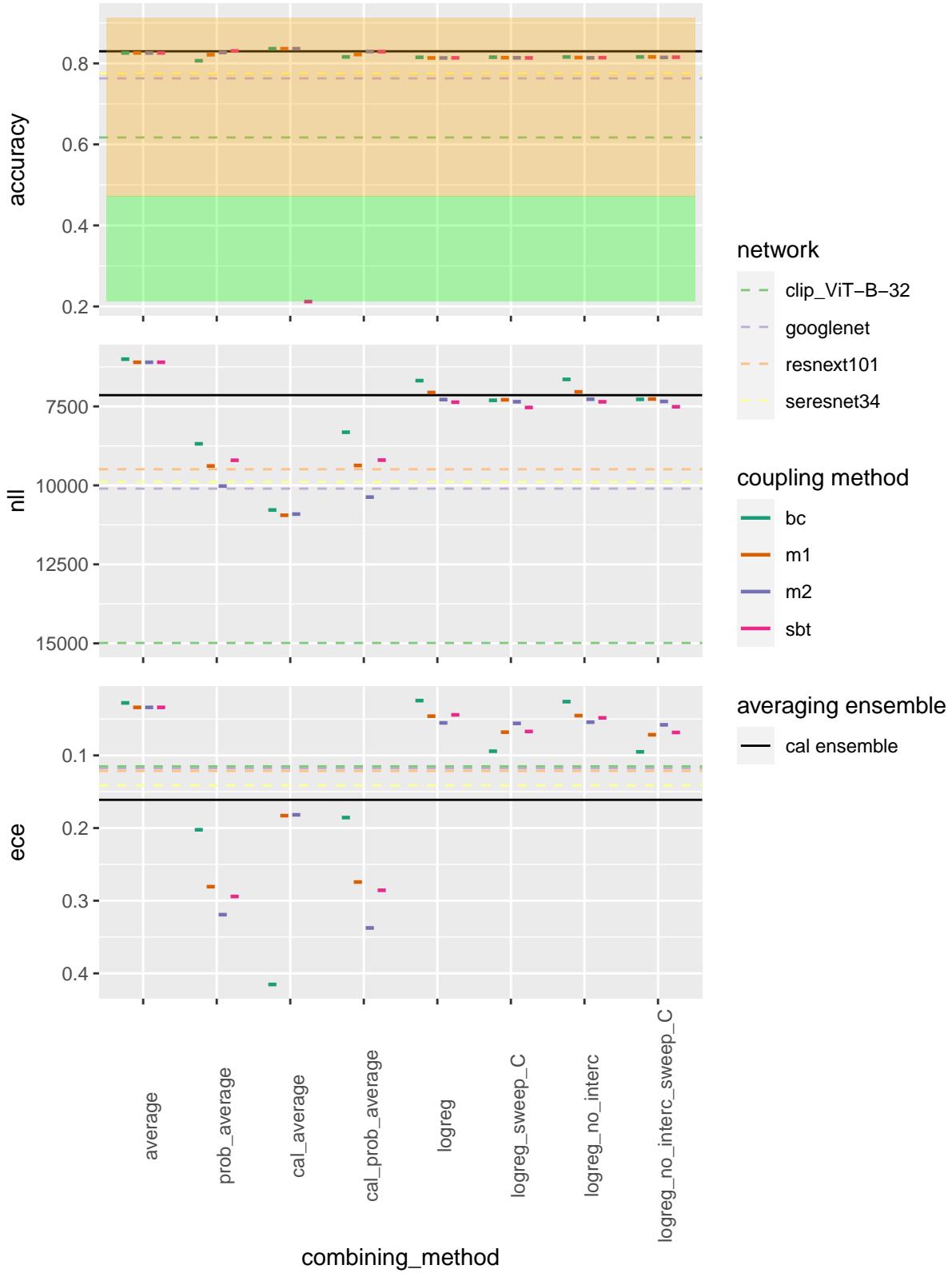
Ensemble metrics  
Error inconsistency 0.462199985980987



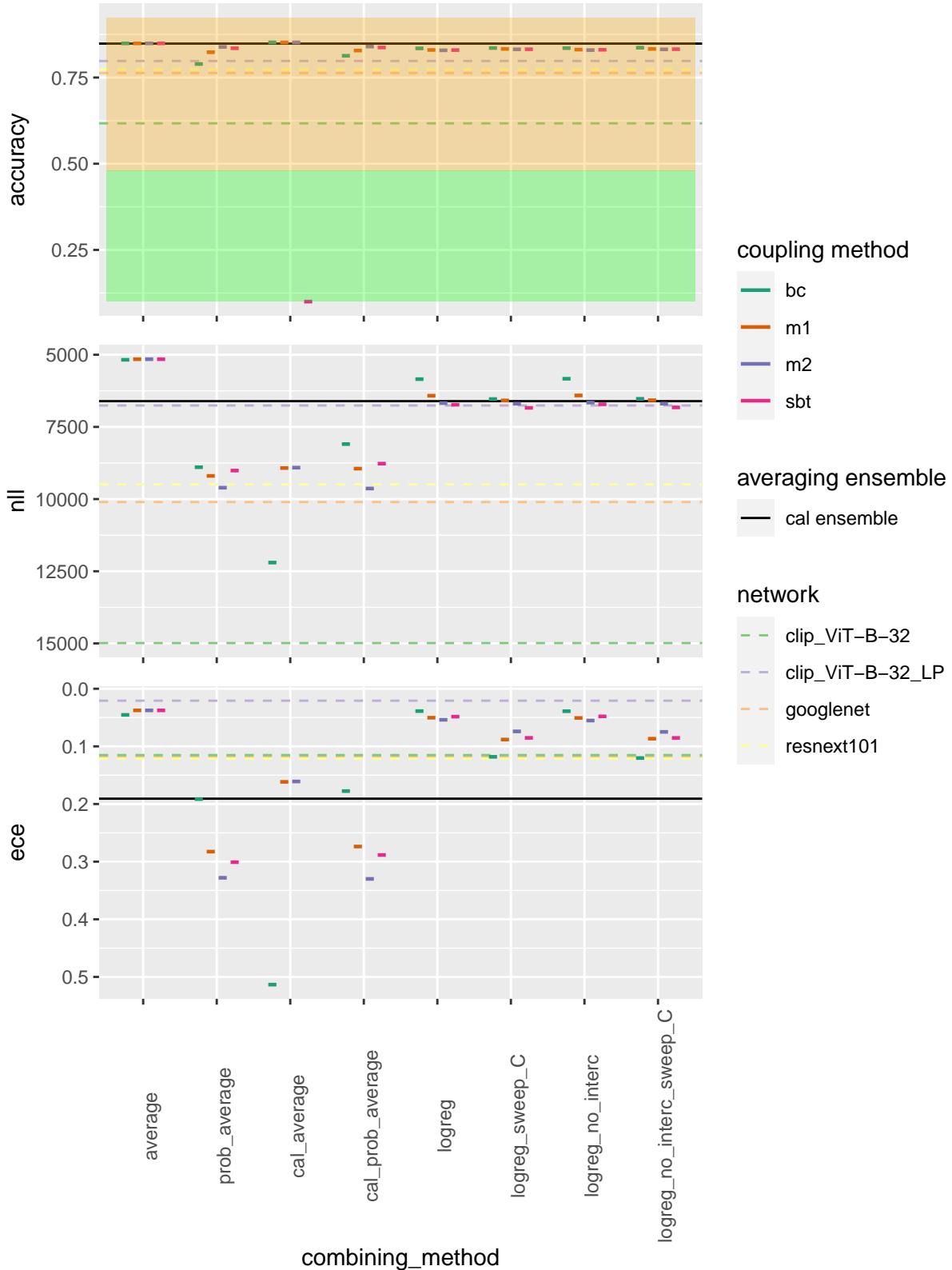
Ensemble metrics  
Error inconsistency 0.443999975919724



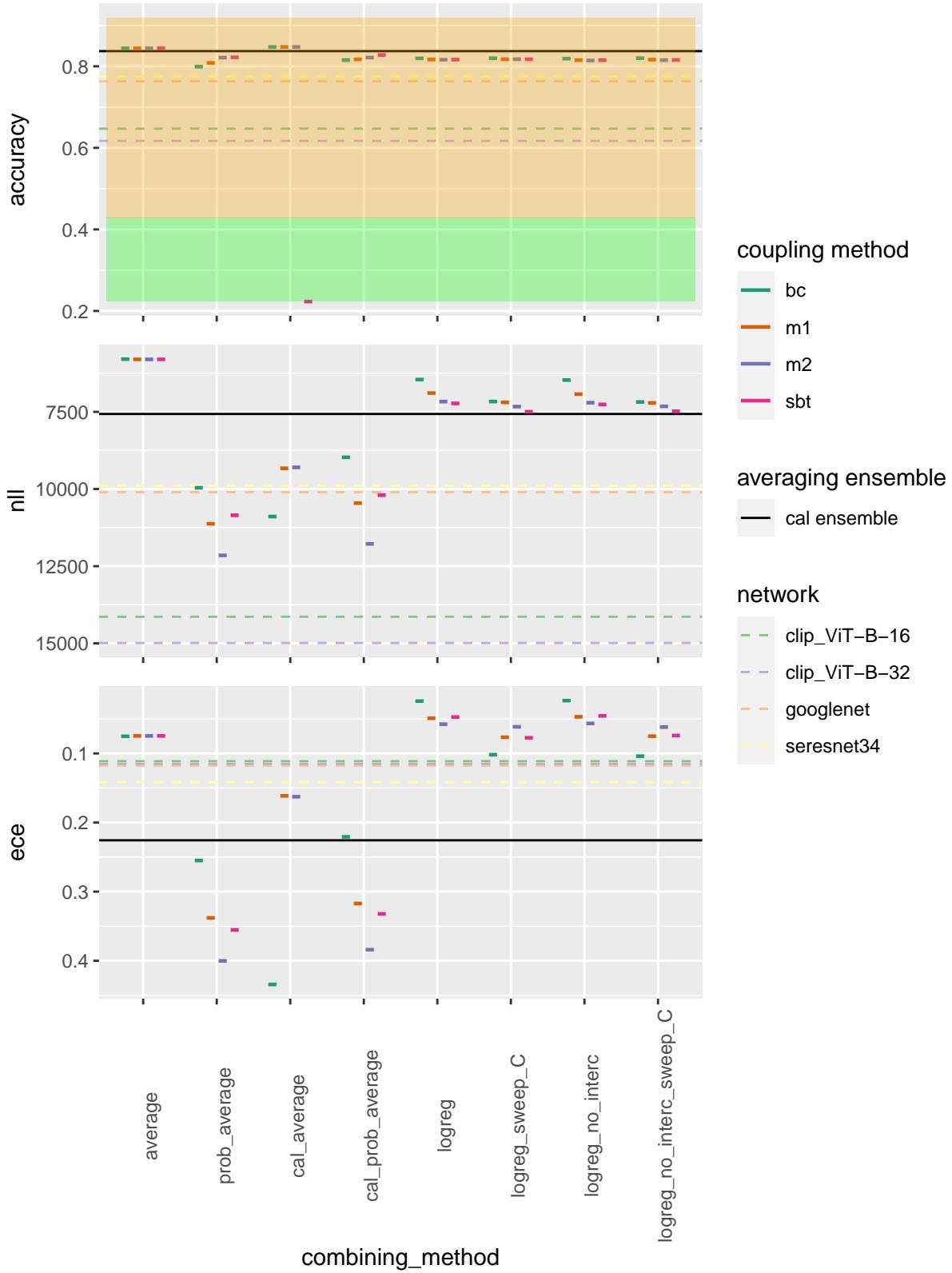
Ensemble metrics  
Error inconsistency 0.440200001001358



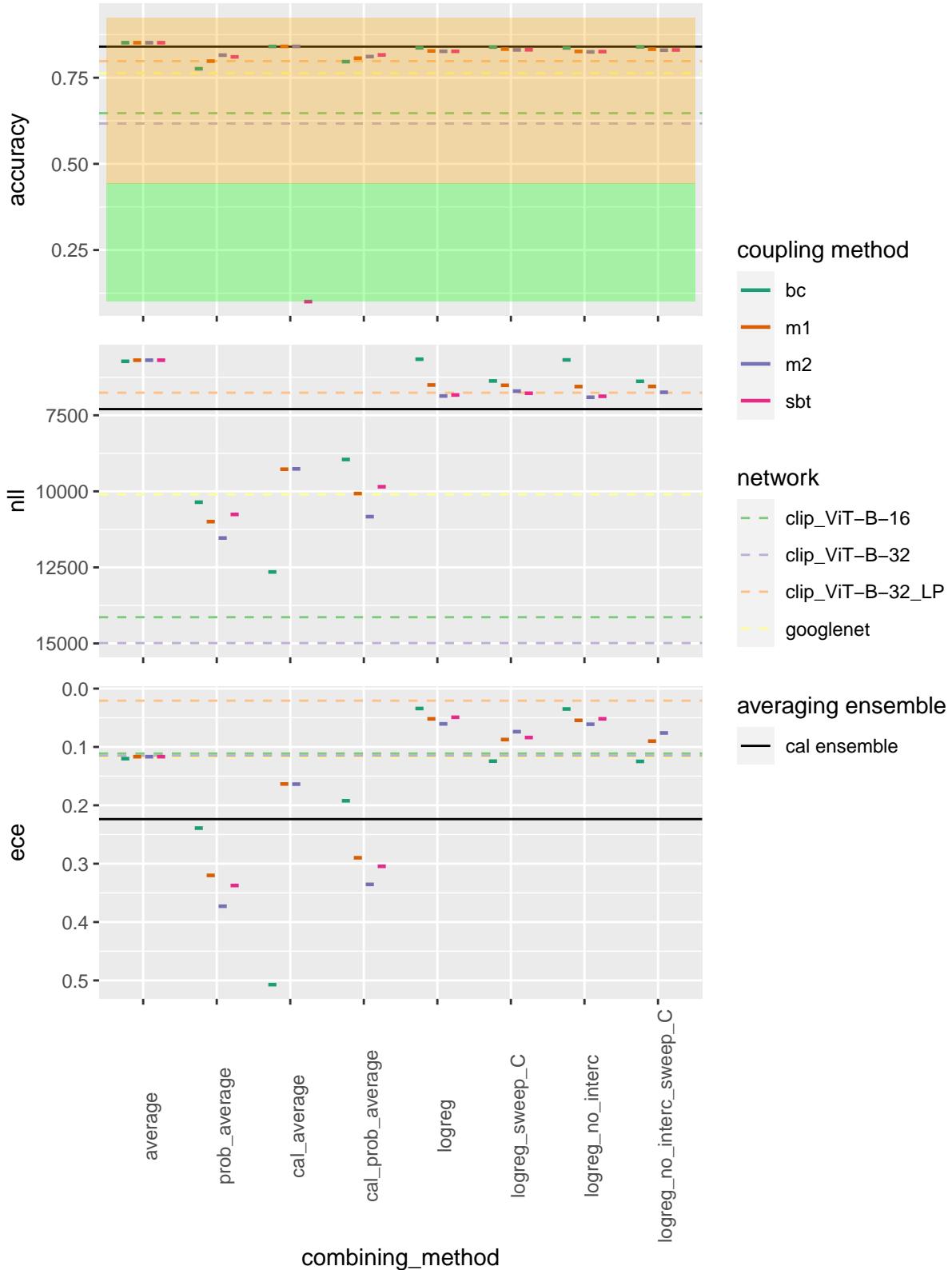
Ensemble metrics  
Error inconsistency 0.444499999284744



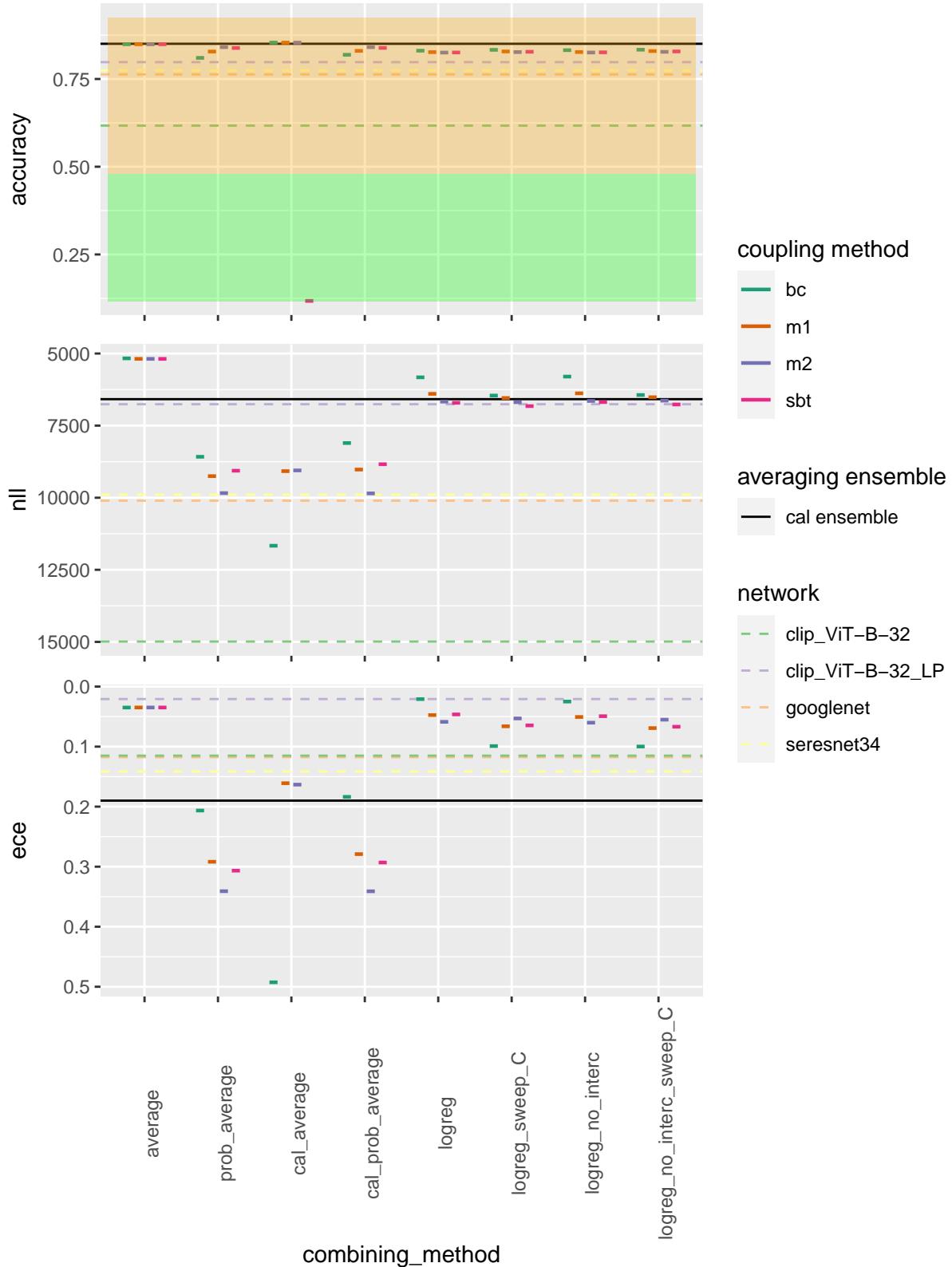
Ensemble metrics  
Error inconsistency 0.490599989891052



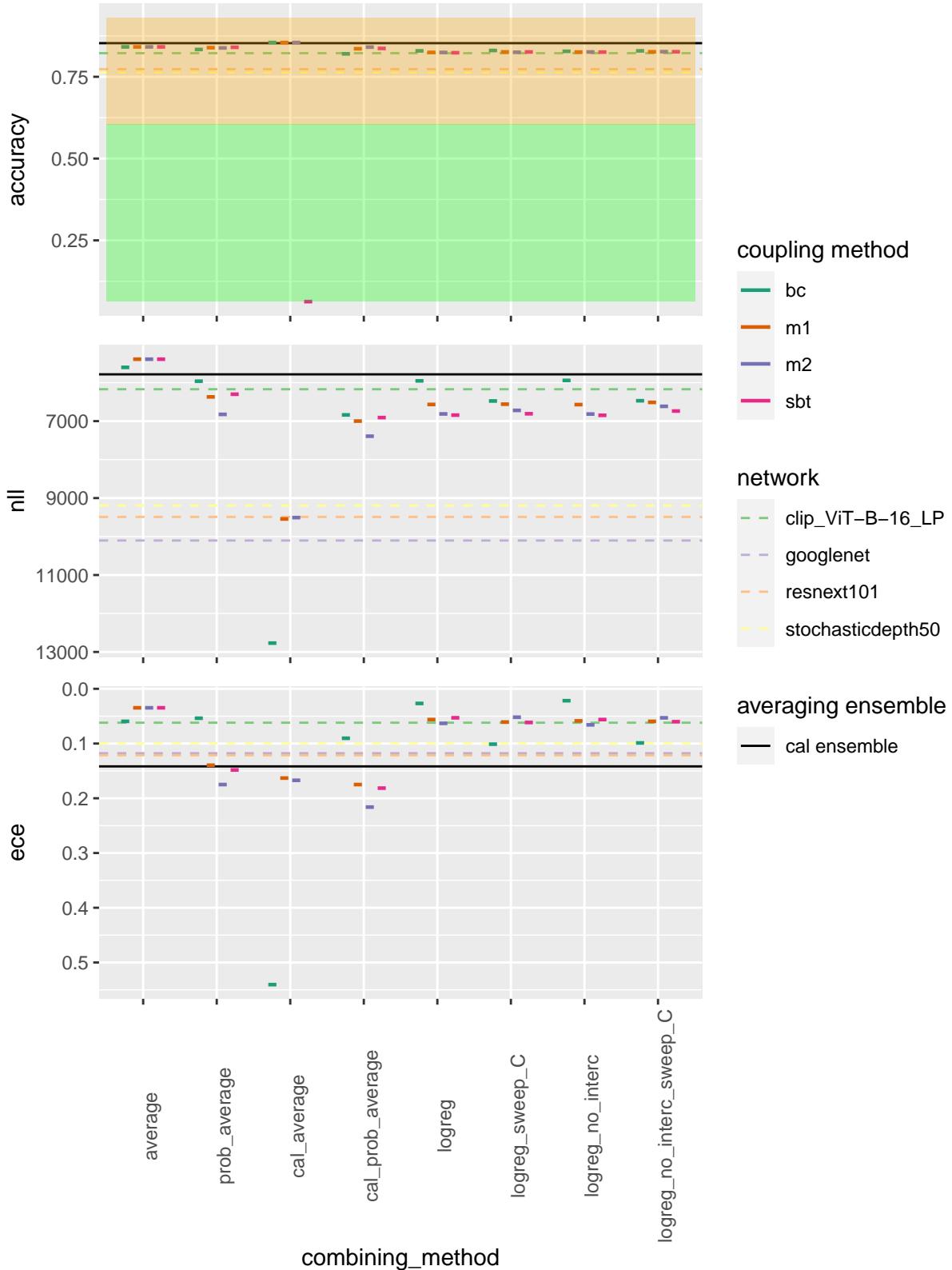
Ensemble metrics  
 Error inconsistency 0.481299996376038



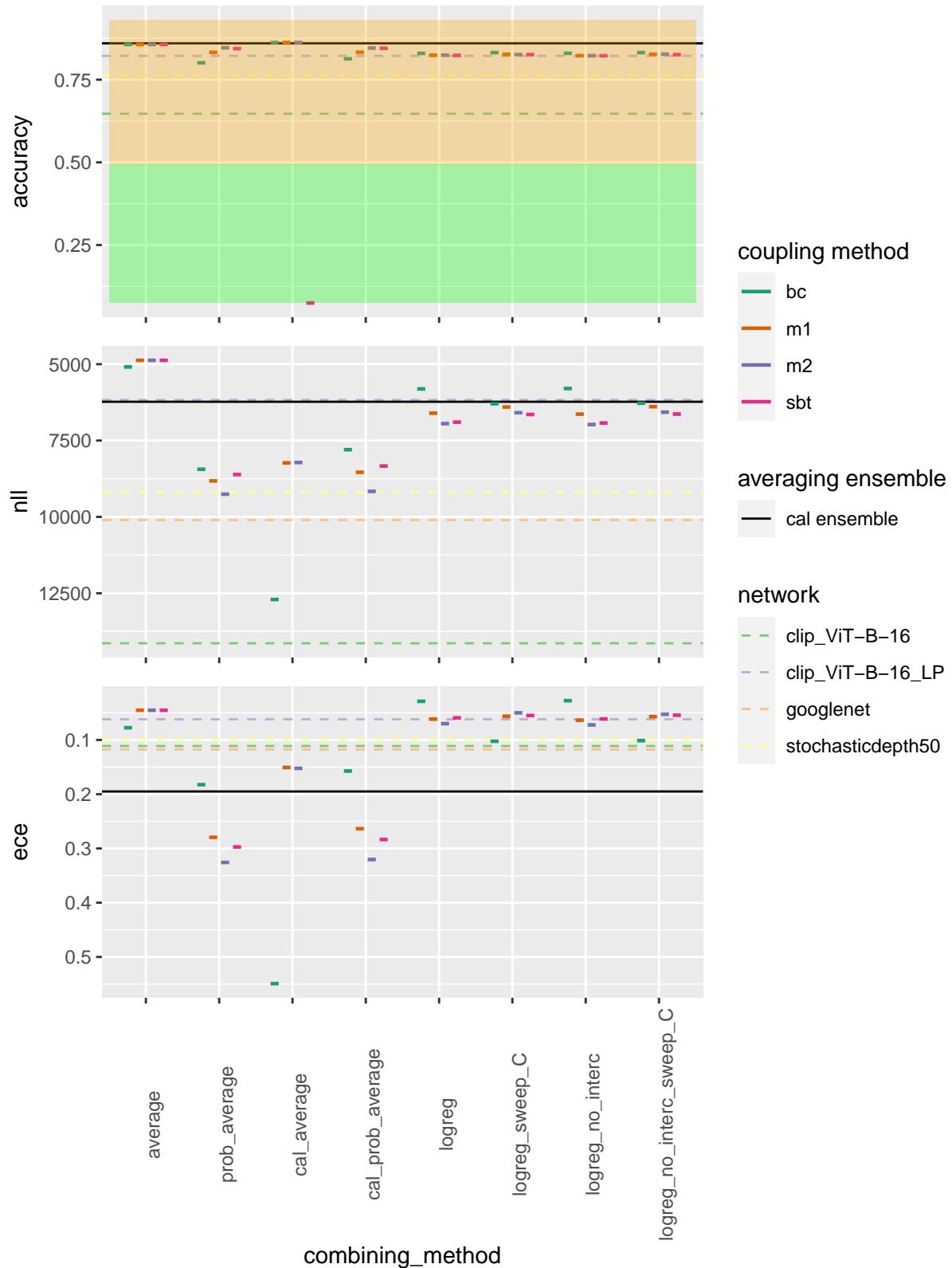
Ensemble metrics  
Error inconsistency 0.44539999961853



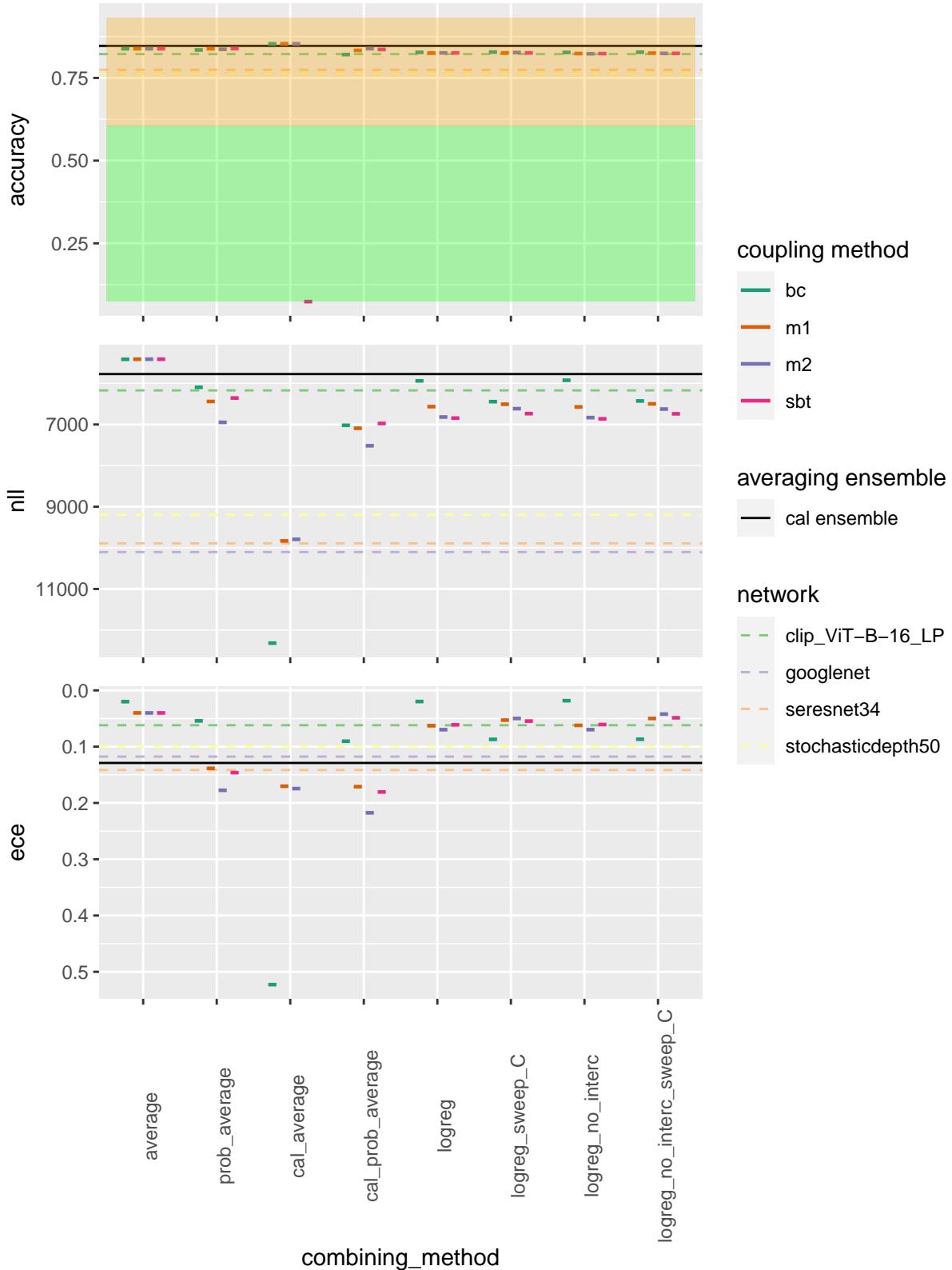
Ensemble metrics  
Error inconsistency 0.326900005340576



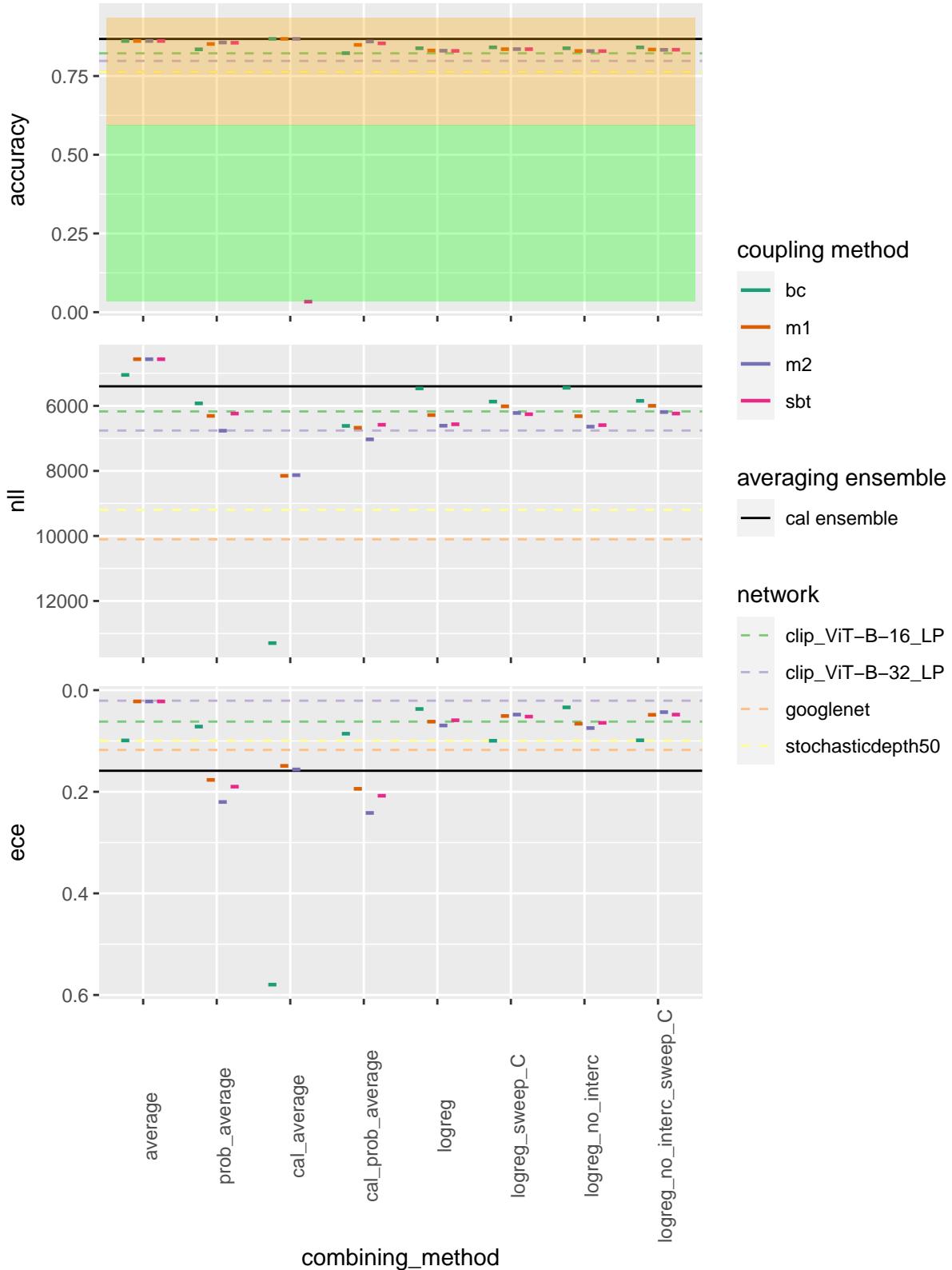
Ensemble metrics  
Error inconsistency 0.436599999666214



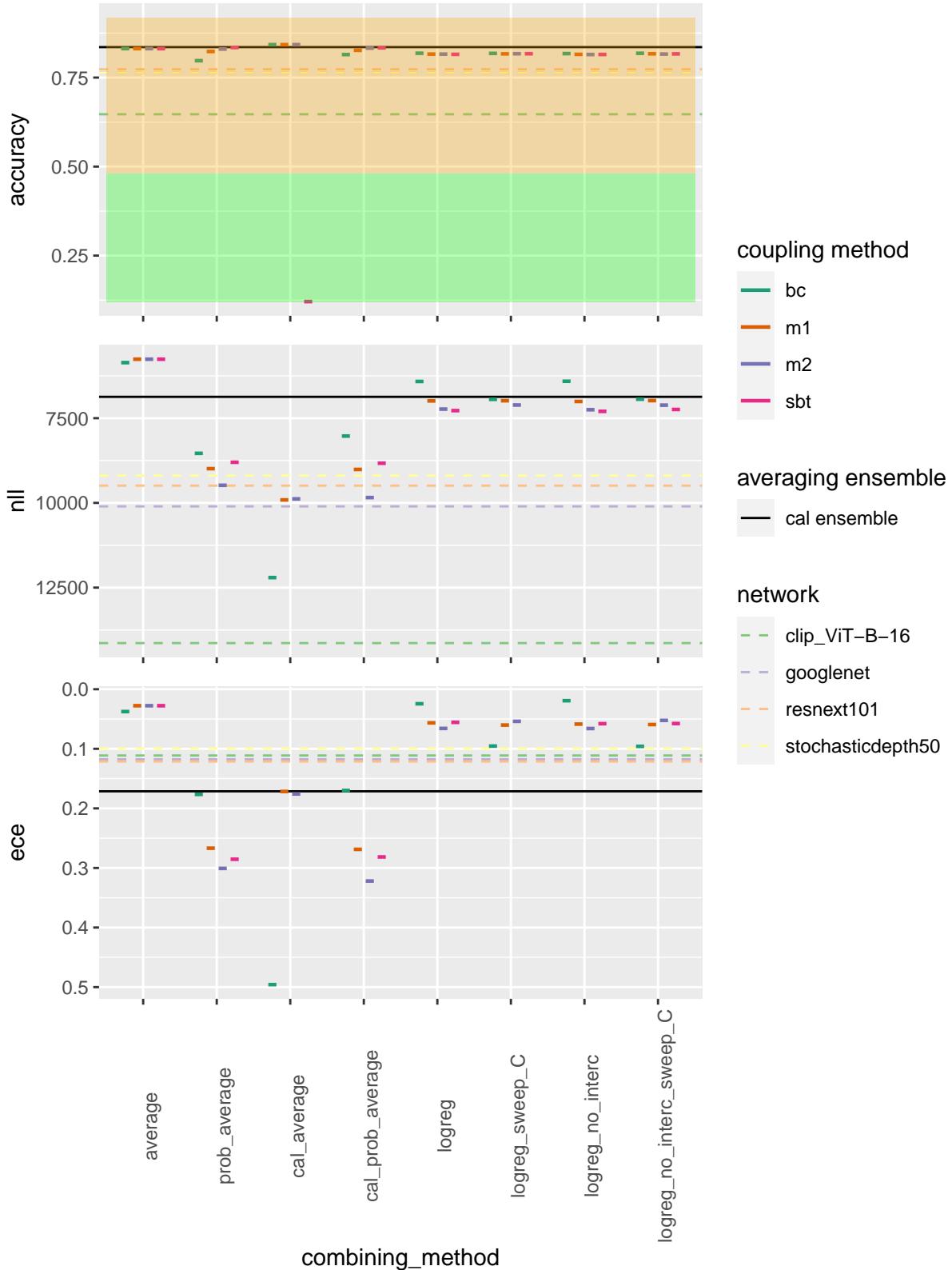
Ensemble metrics  
Error inconsistency 0.326499998569489



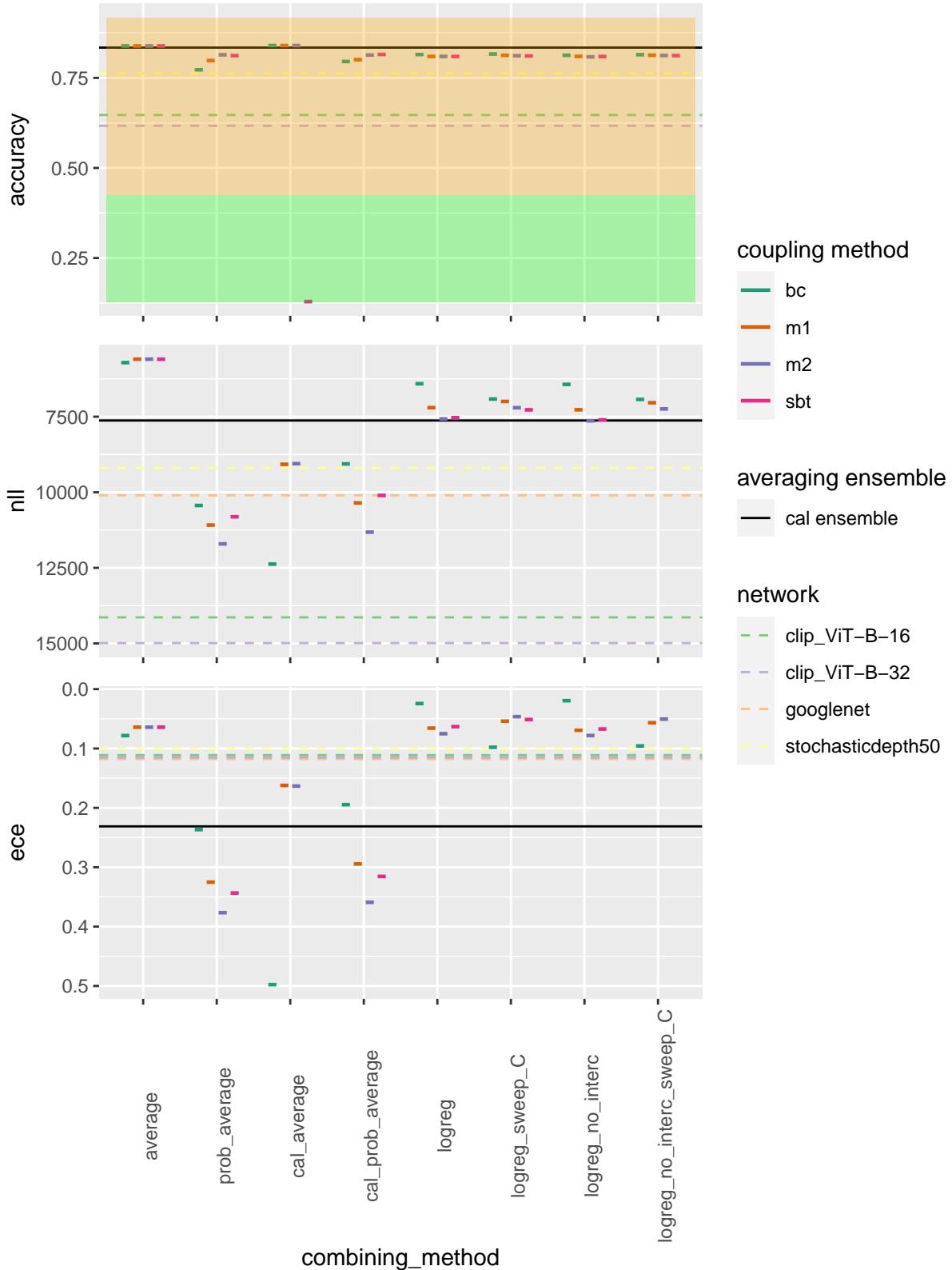
Ensemble metrics  
Error inconsistency 0.341999977827072



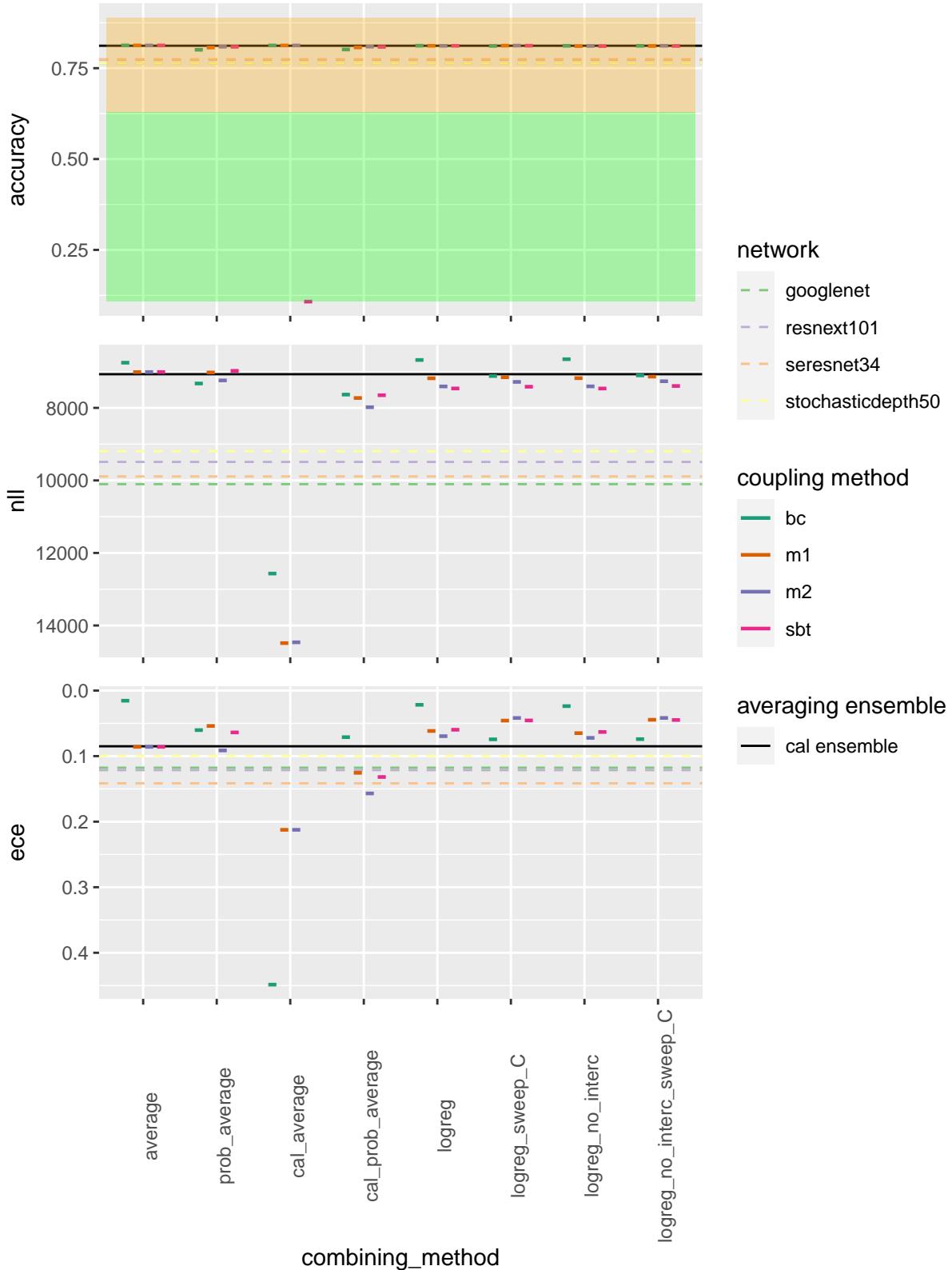
Ensemble metrics  
Error inconsistency 0.437899976968765



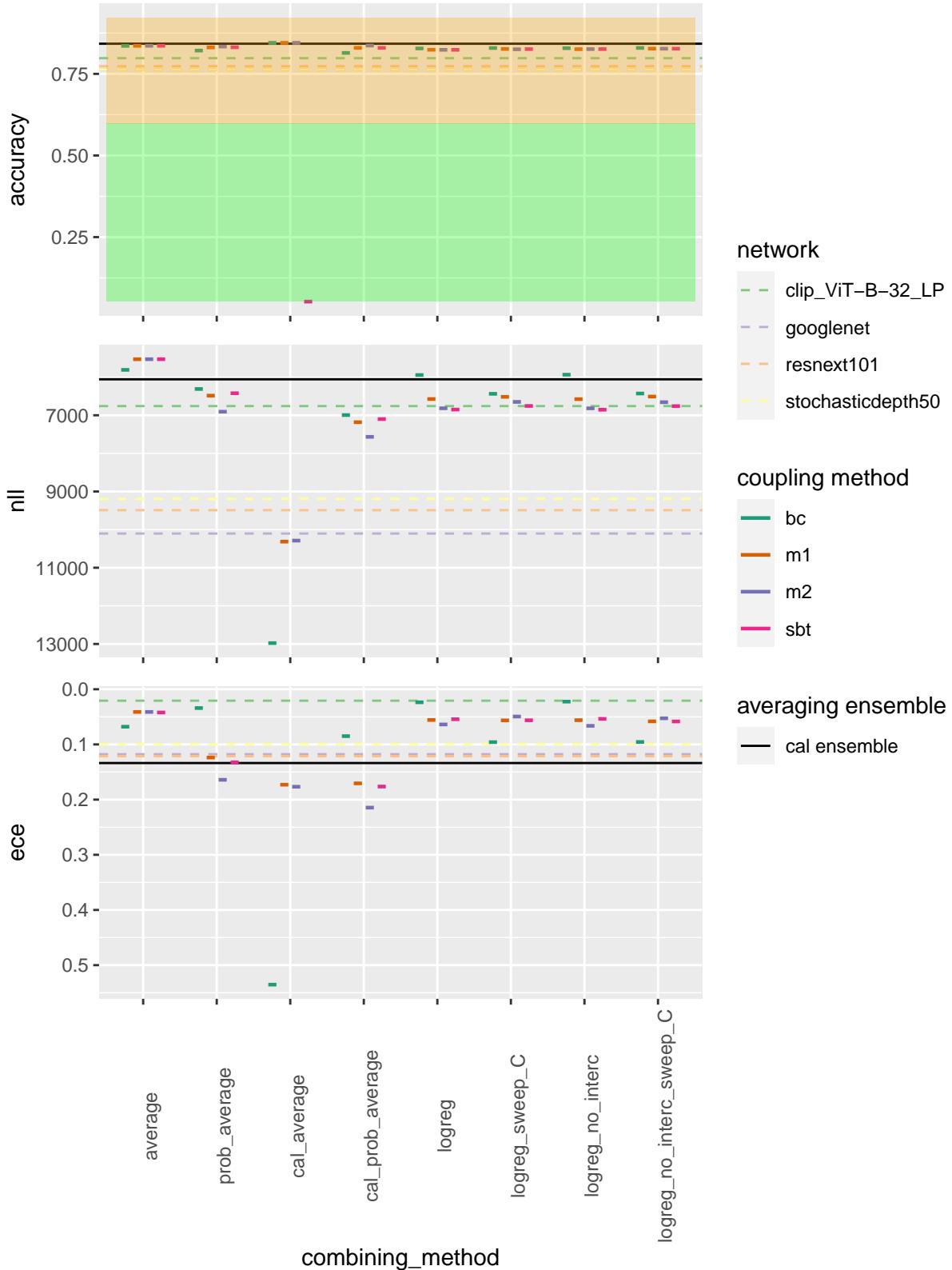
Ensemble metrics  
Error inconsistency 0.492699980735779



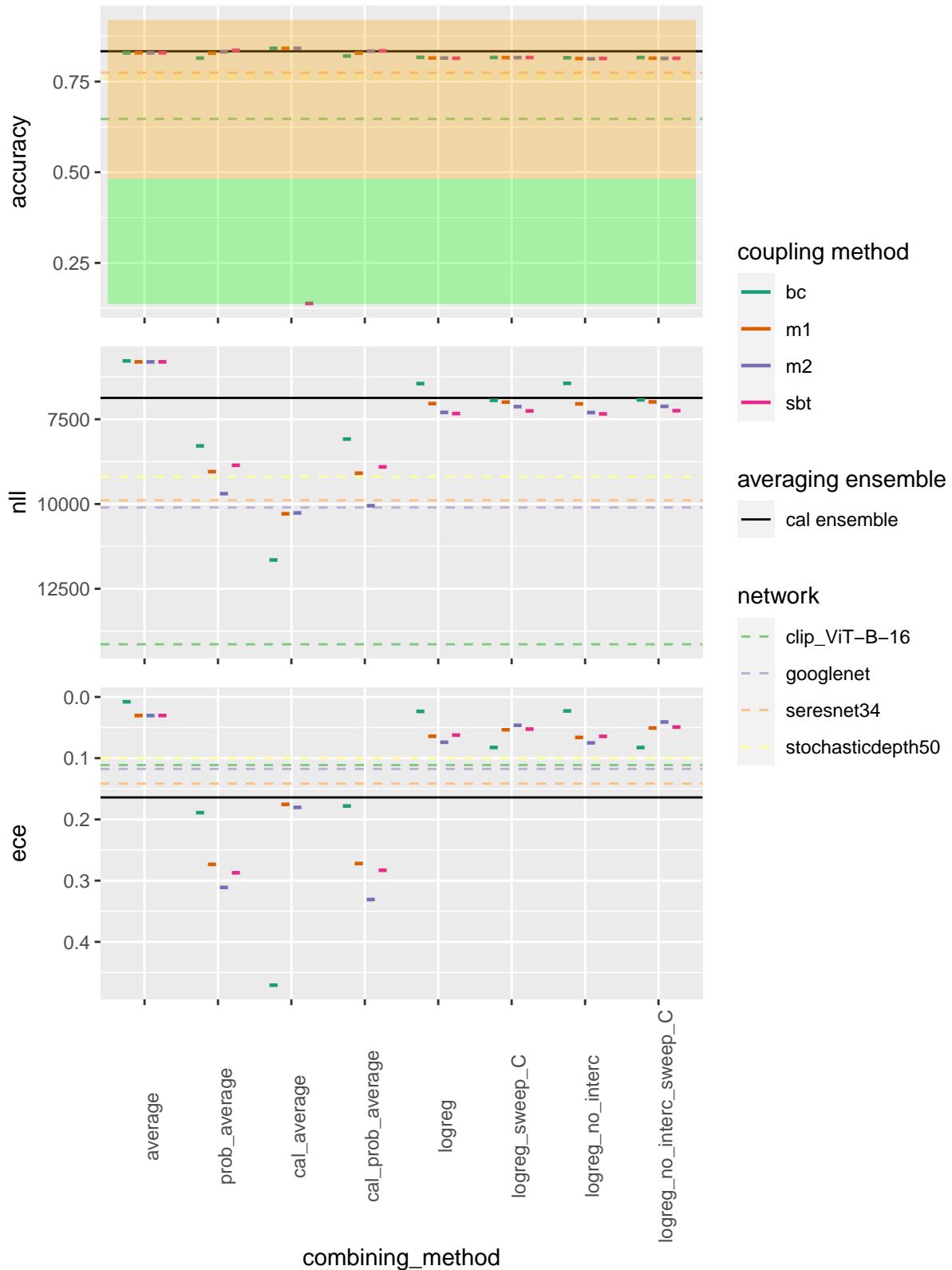
Ensemble metrics  
Error inconsistency 0.260600000619888



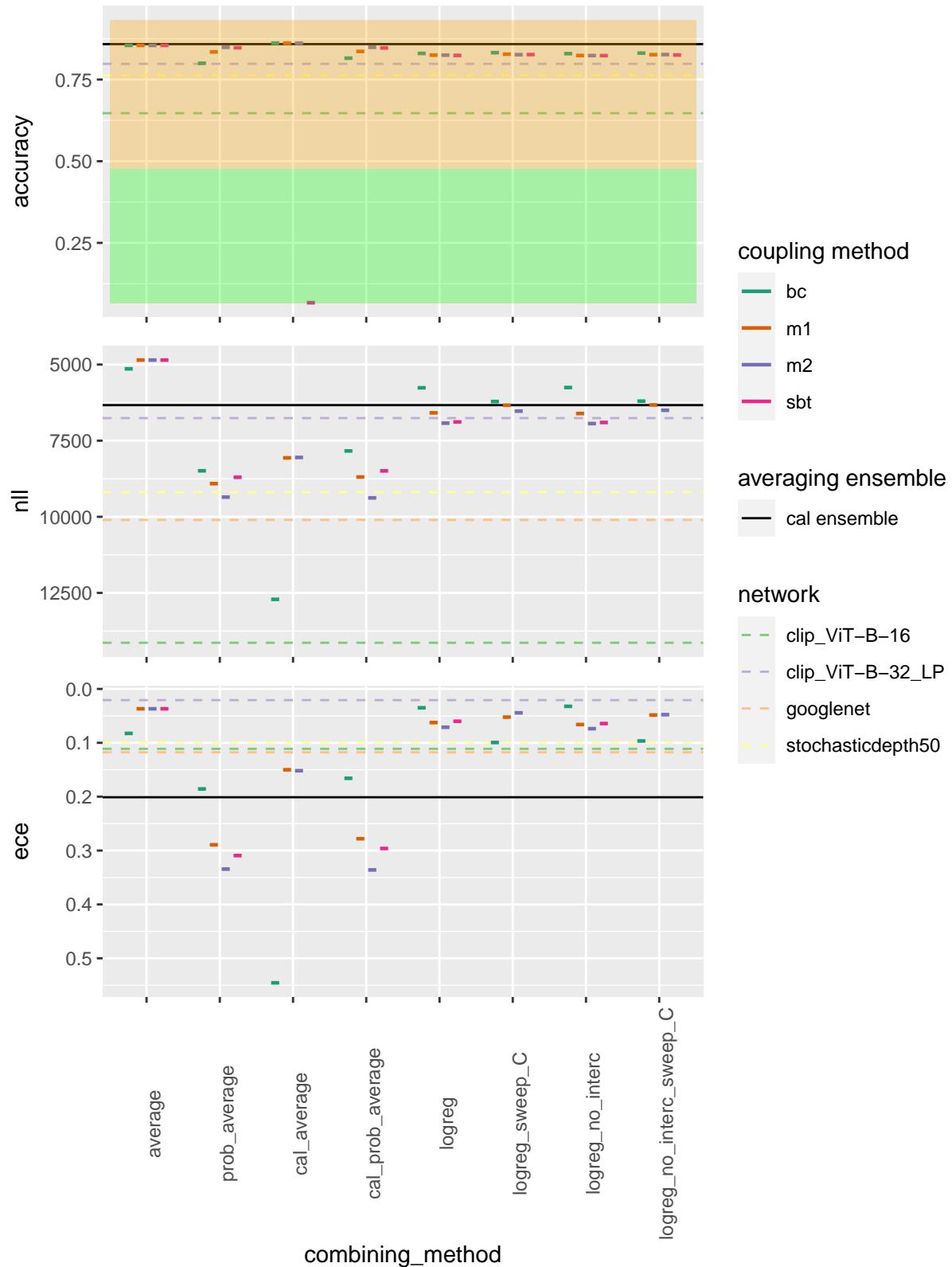
Ensemble metrics  
Error inconsistency 0.324099987745285



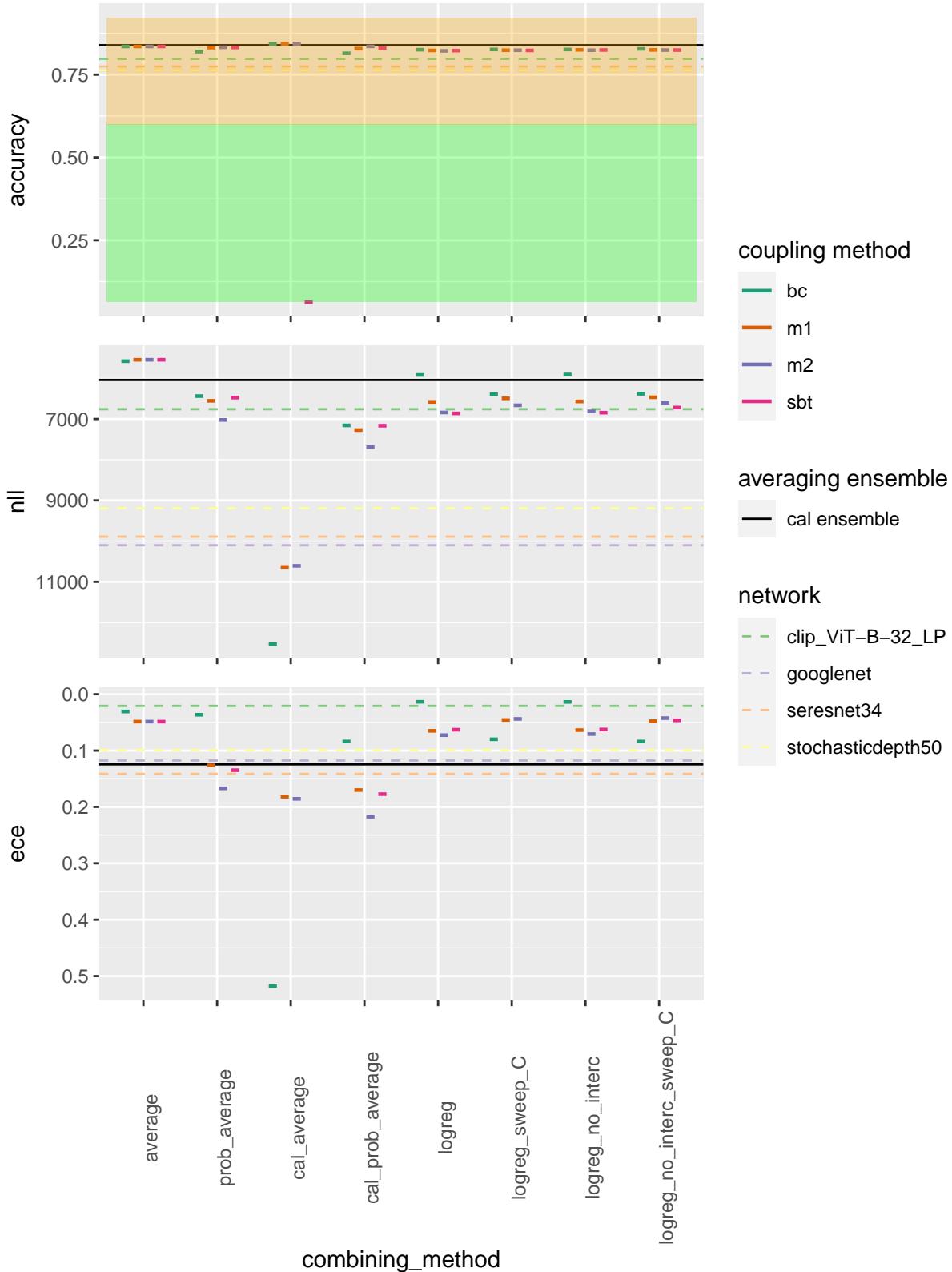
Ensemble metrics  
Error inconsistency 0.437700003385544



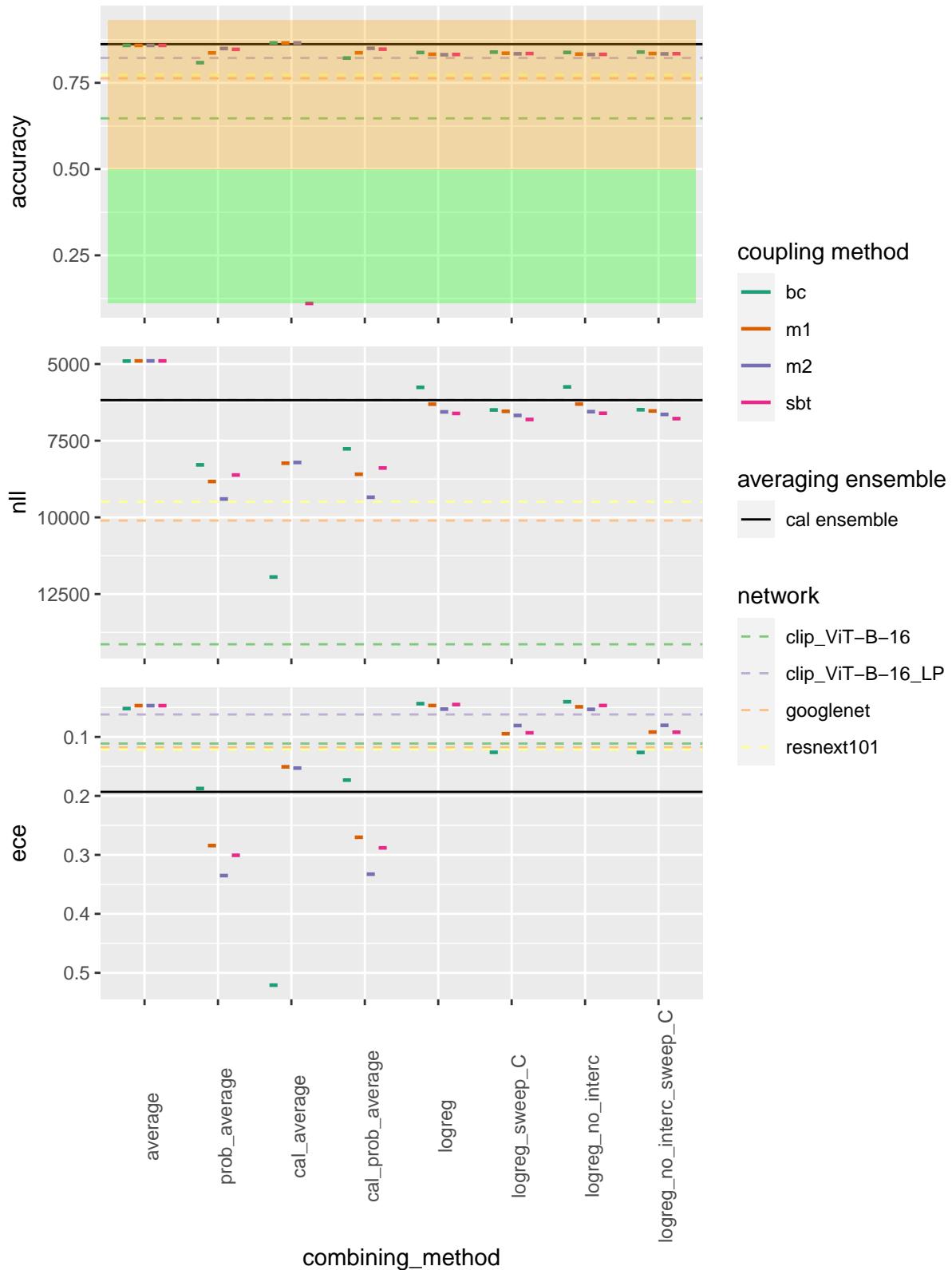
Ensemble metrics  
Error inconsistency 0.455699980258942



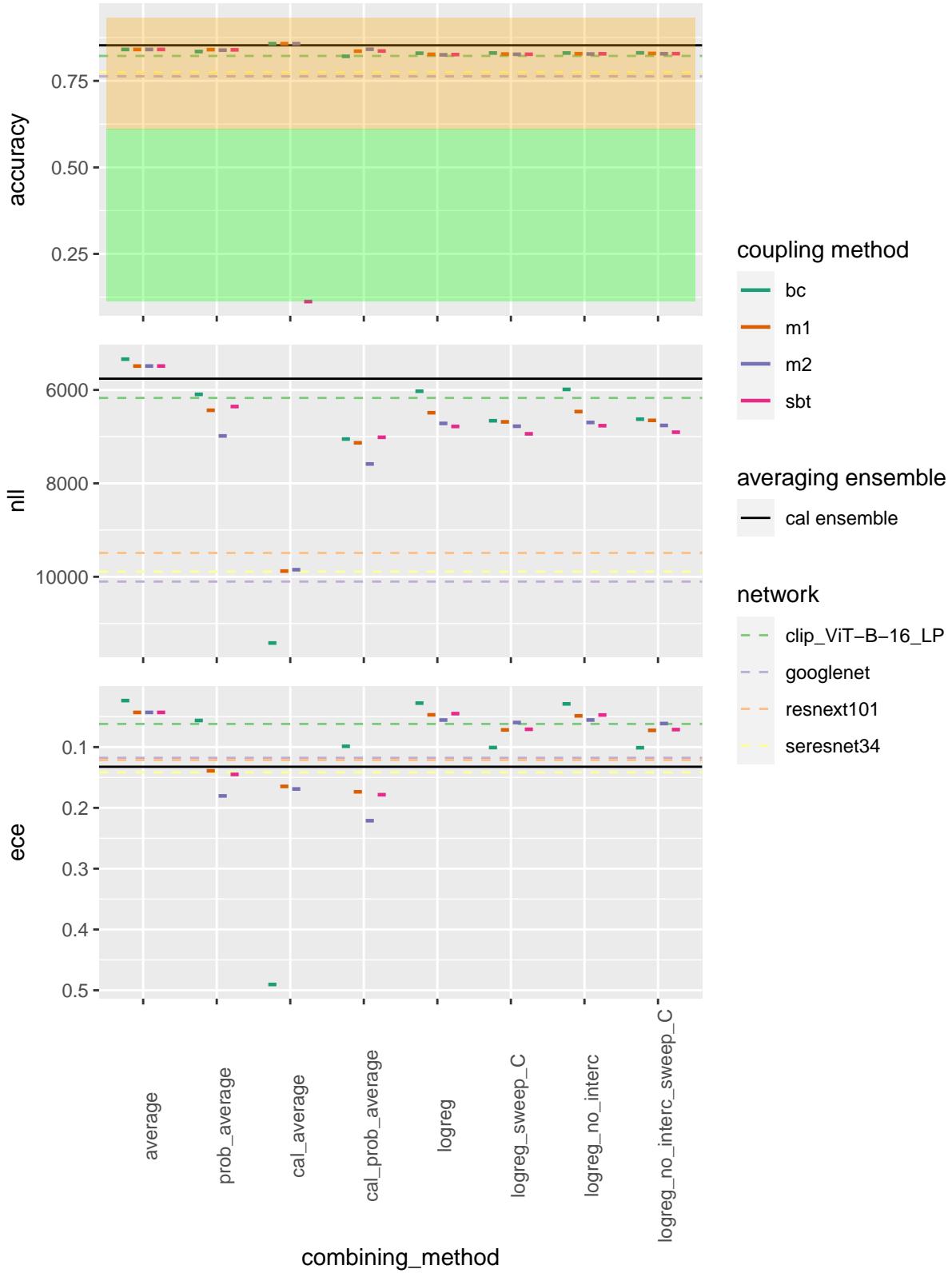
Ensemble metrics  
Error inconsistency 0.323499977588654



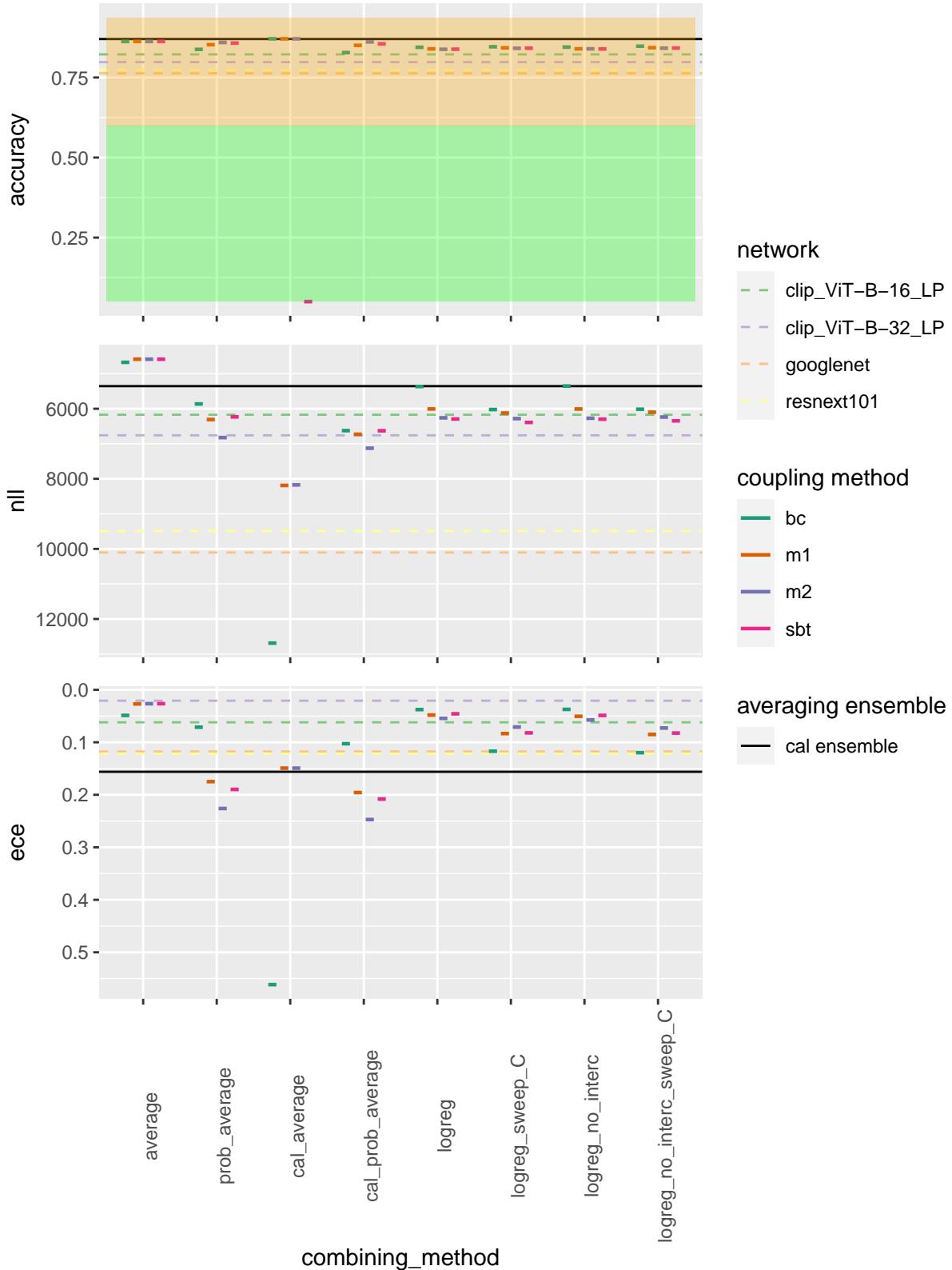
Ensemble metrics  
Error inconsistency 0.433799982070923



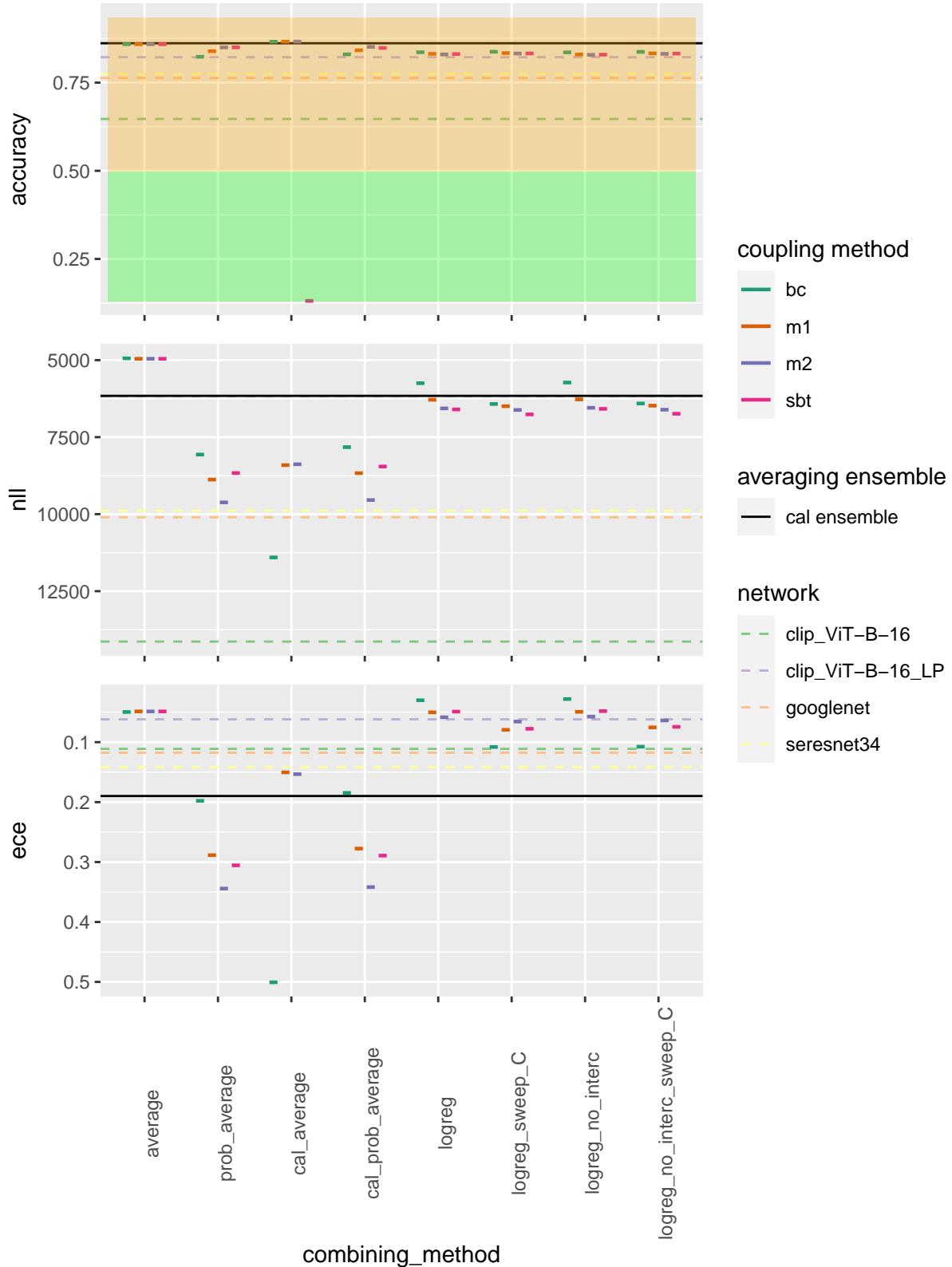
Ensemble metrics  
Error inconsistency 0.322099983692169



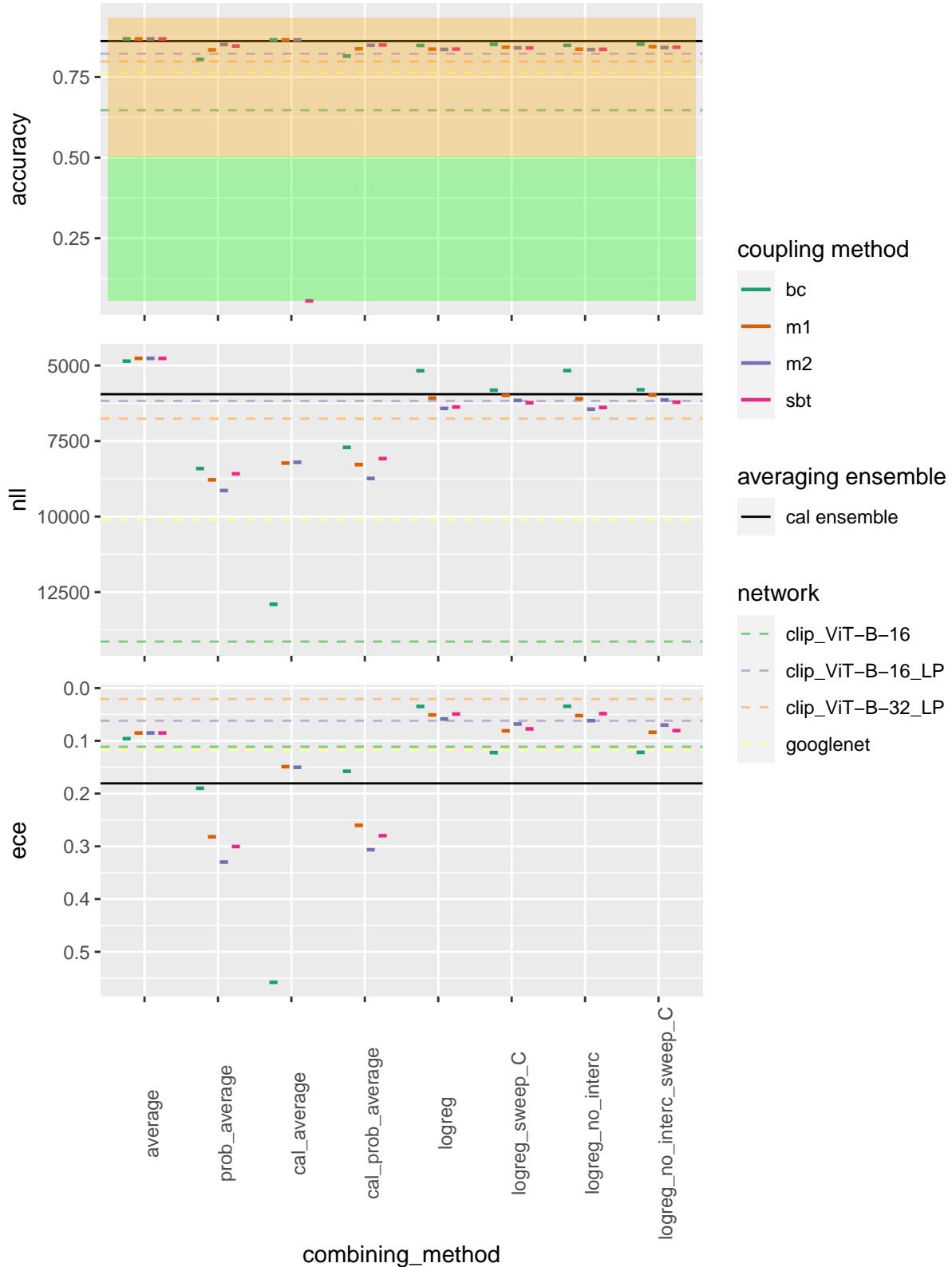
Ensemble metrics  
Error inconsistency 0.337799996137619



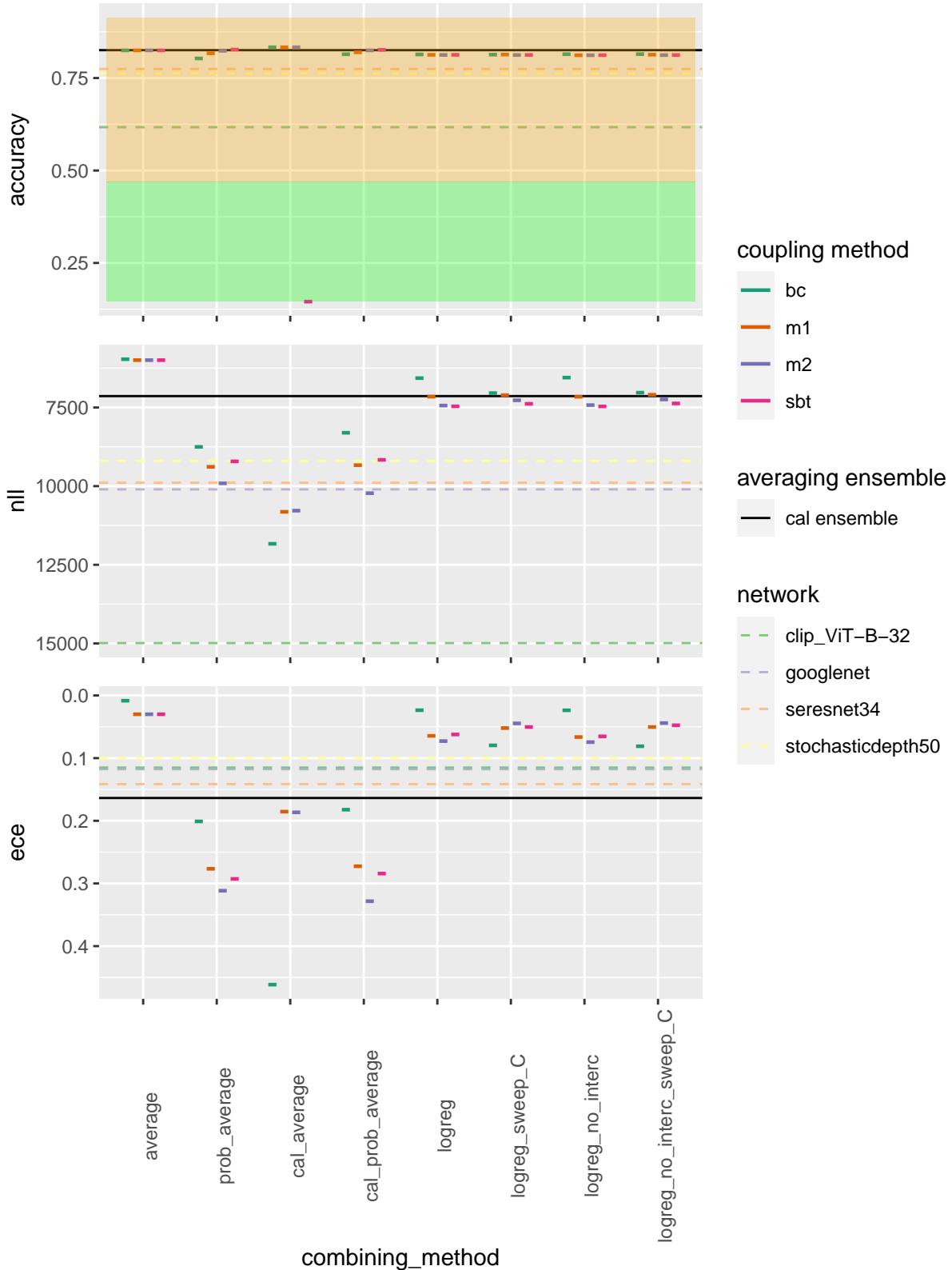
Ensemble metrics  
Error inconsistency 0.436299979686737



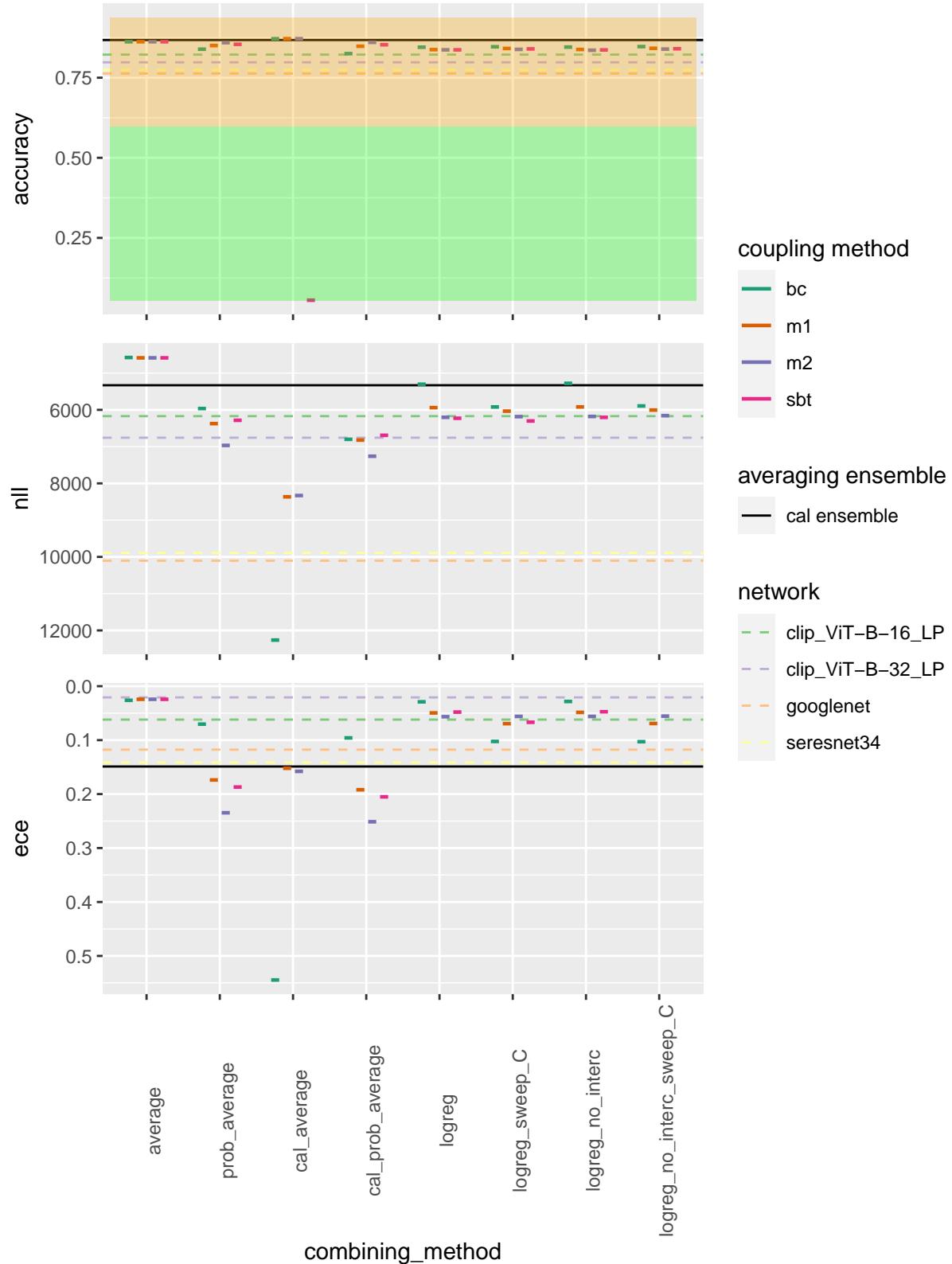
Ensemble metrics  
Error inconsistency 0.43079999089241



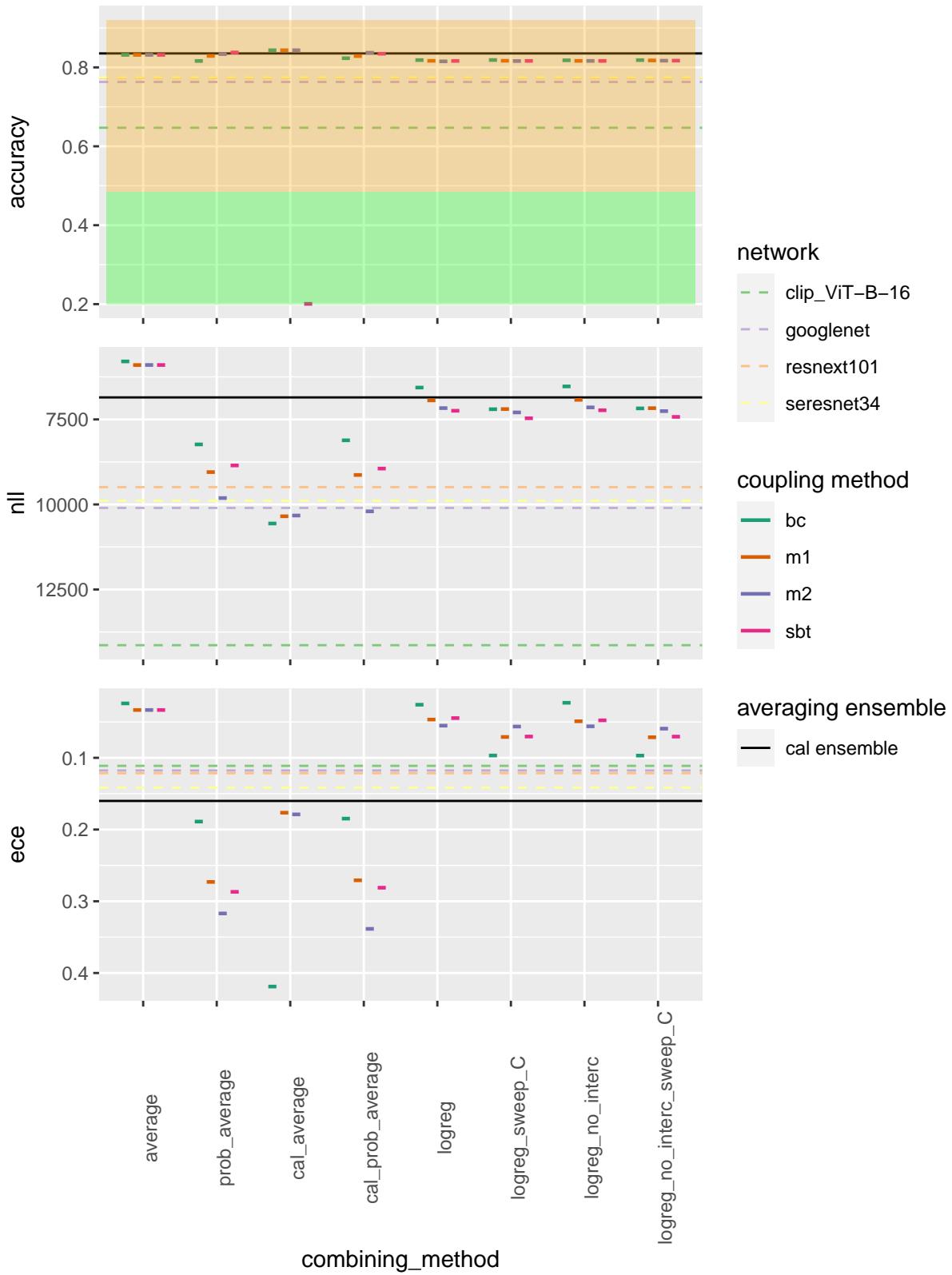
Ensemble metrics  
Error inconsistency 0.443999975919724



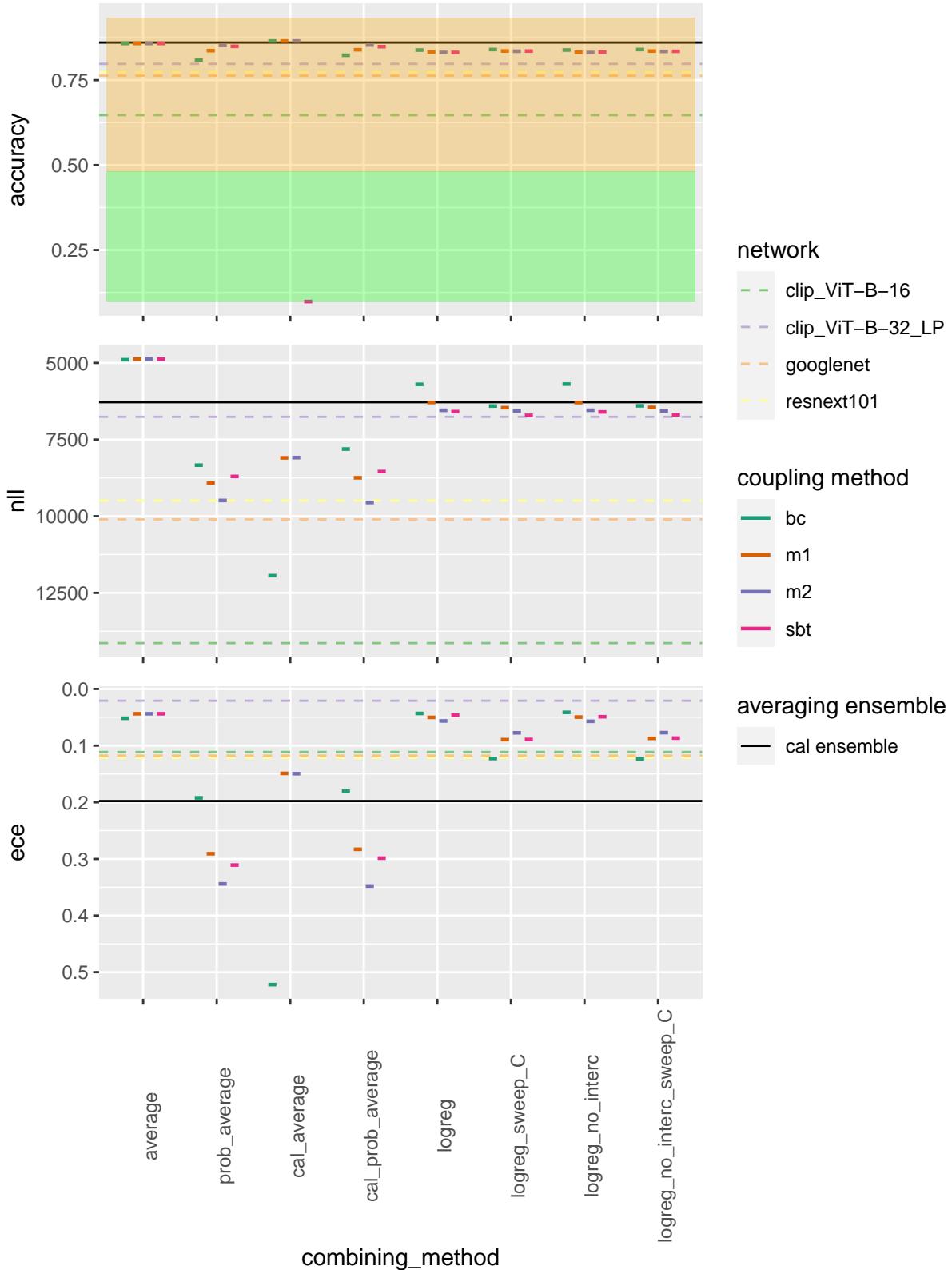
Ensemble metrics  
Error inconsistency 0.339899986982346



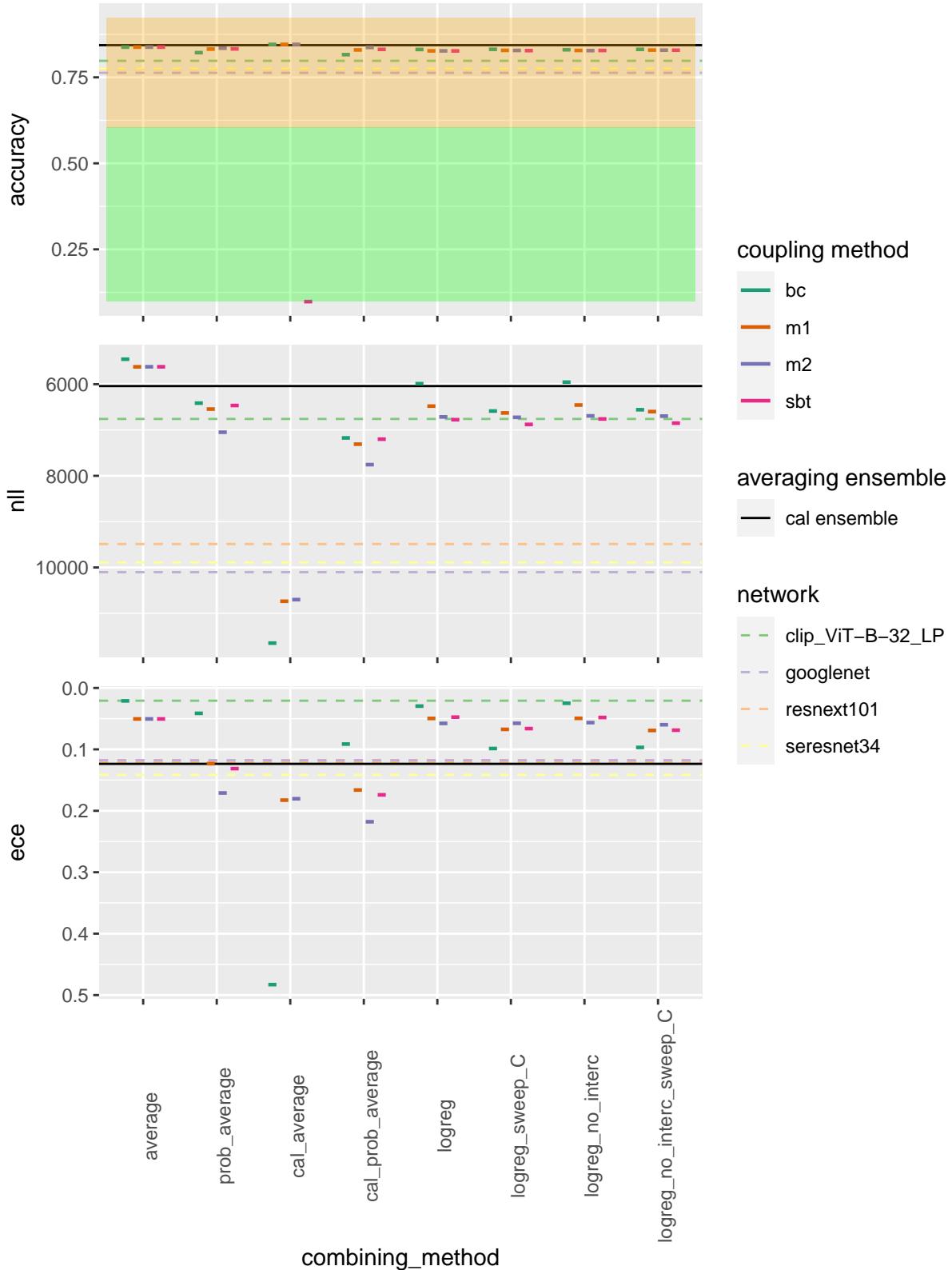
Ensemble metrics  
Error inconsistency 0.435499995946884



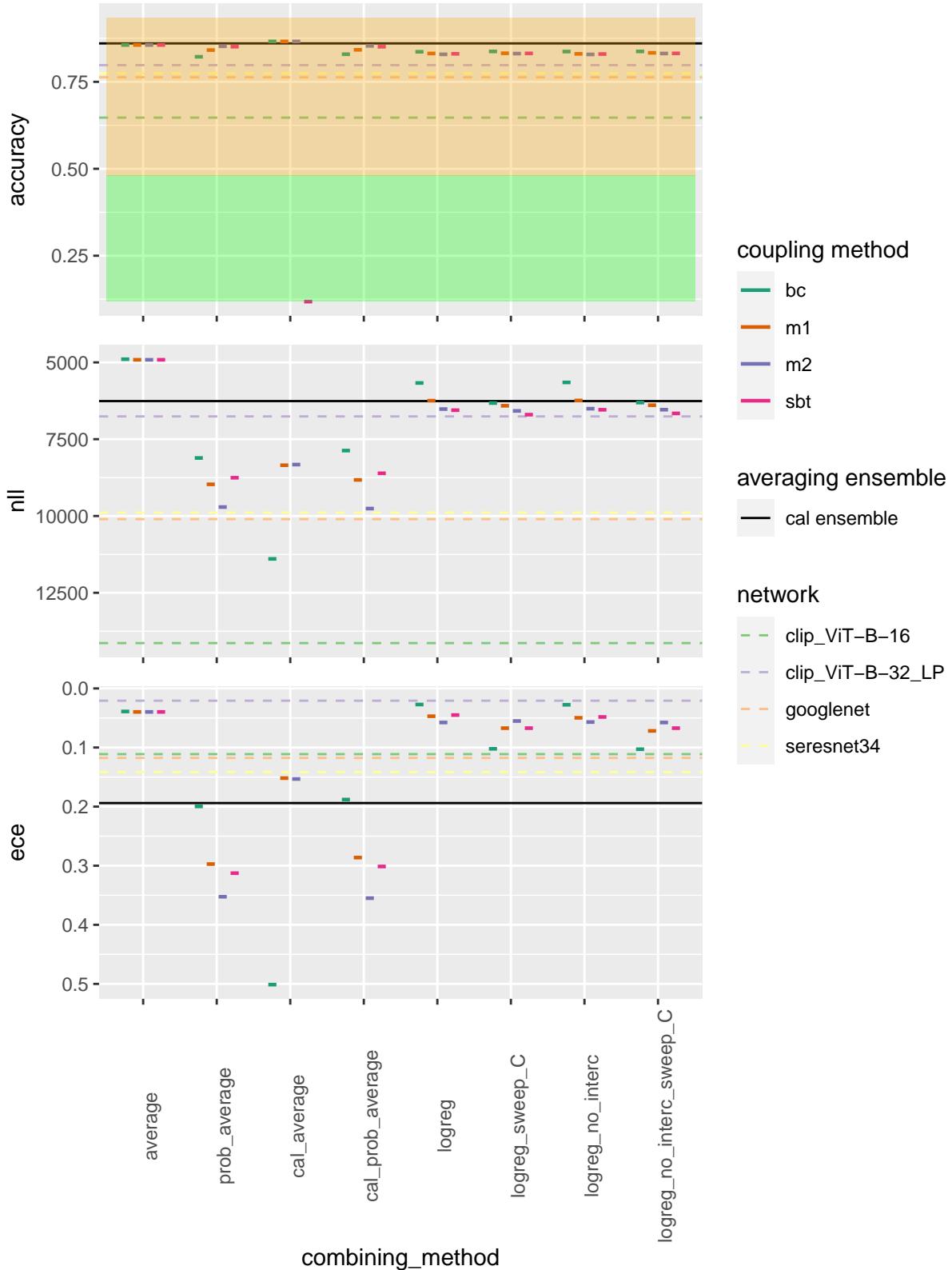
Ensemble metrics  
Error inconsistency 0.453199982643127



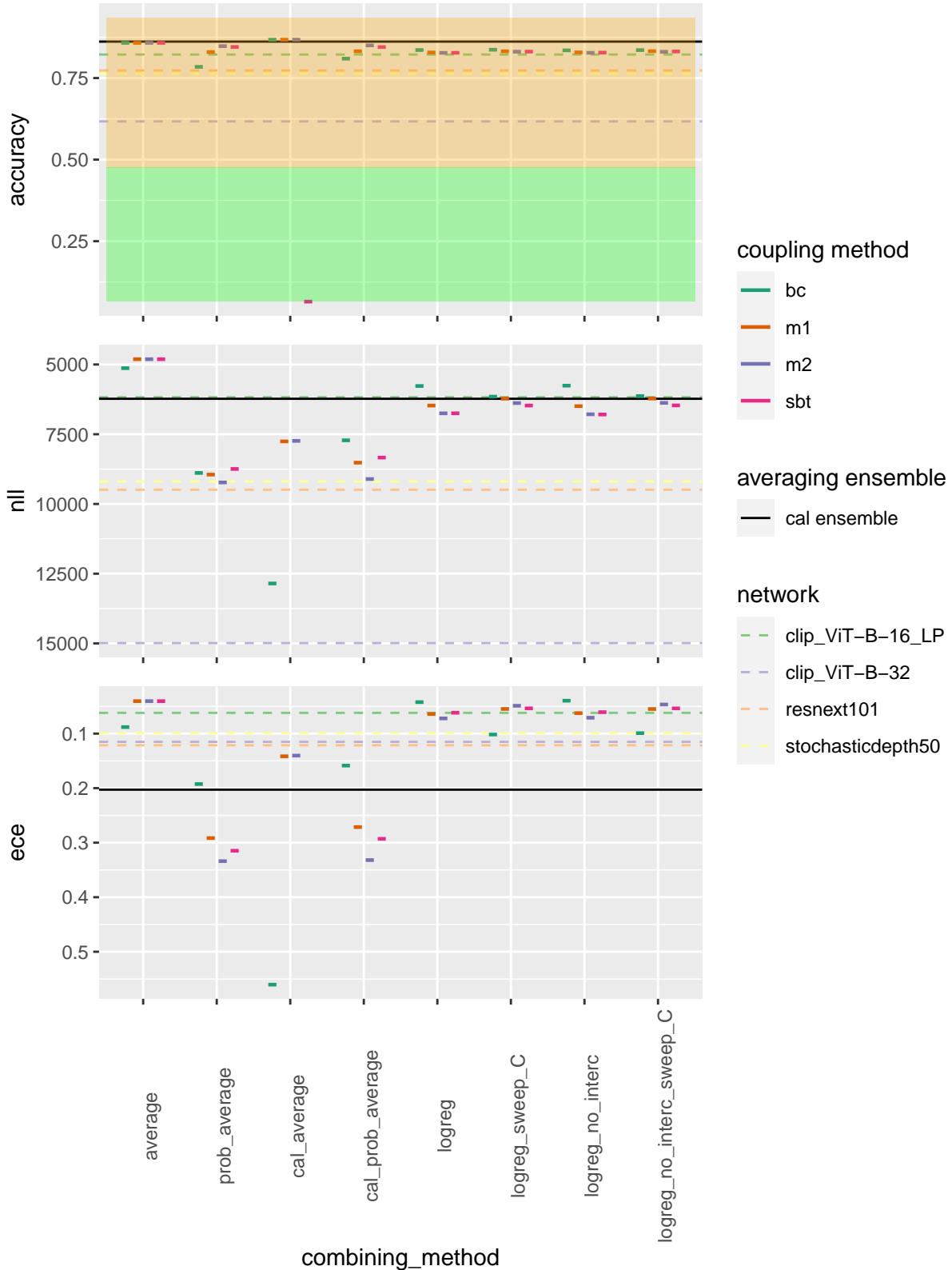
Ensemble metrics  
Error inconsistency 0.319399982690811



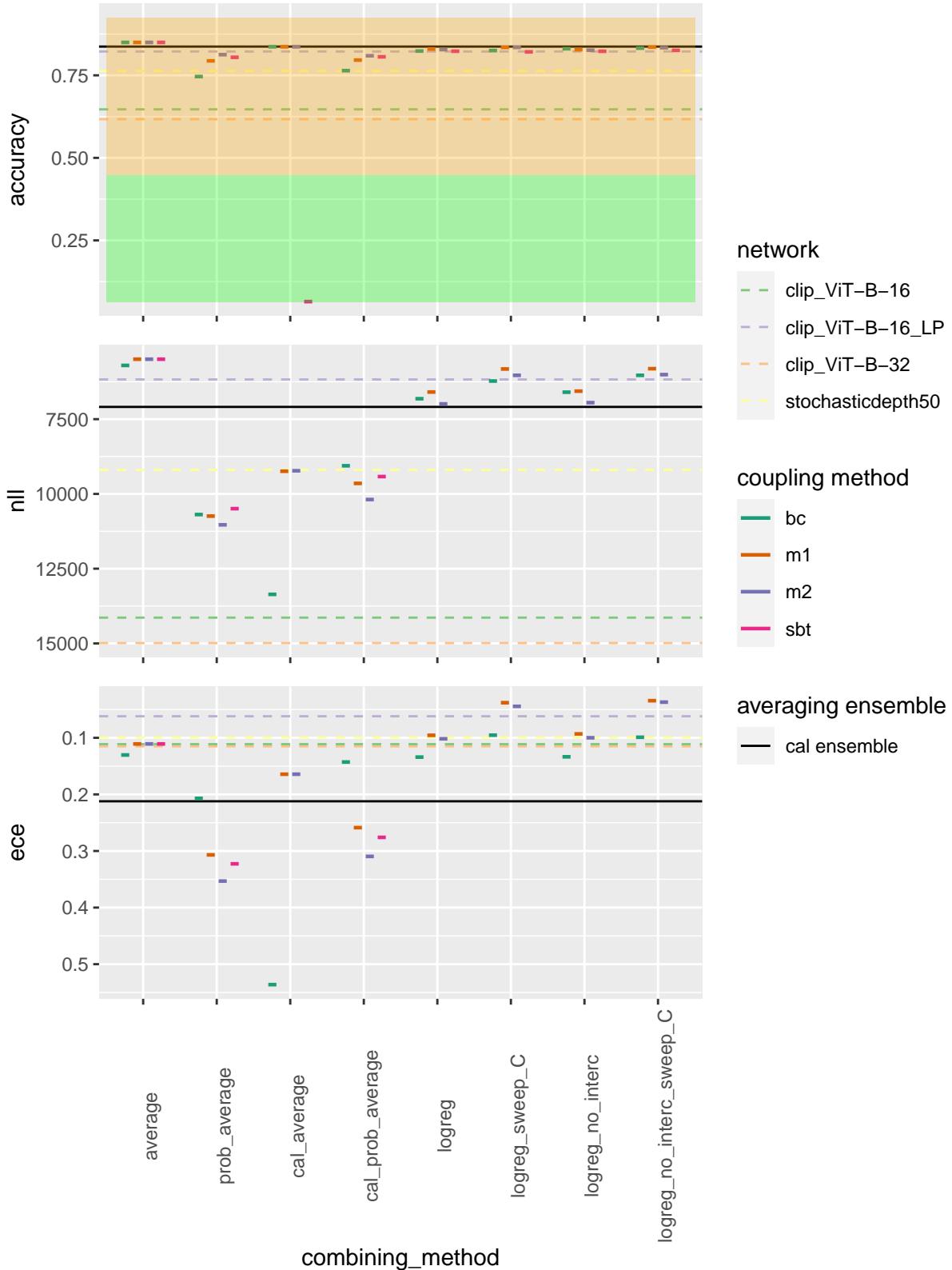
Ensemble metrics  
Error inconsistency 0.453799992799759



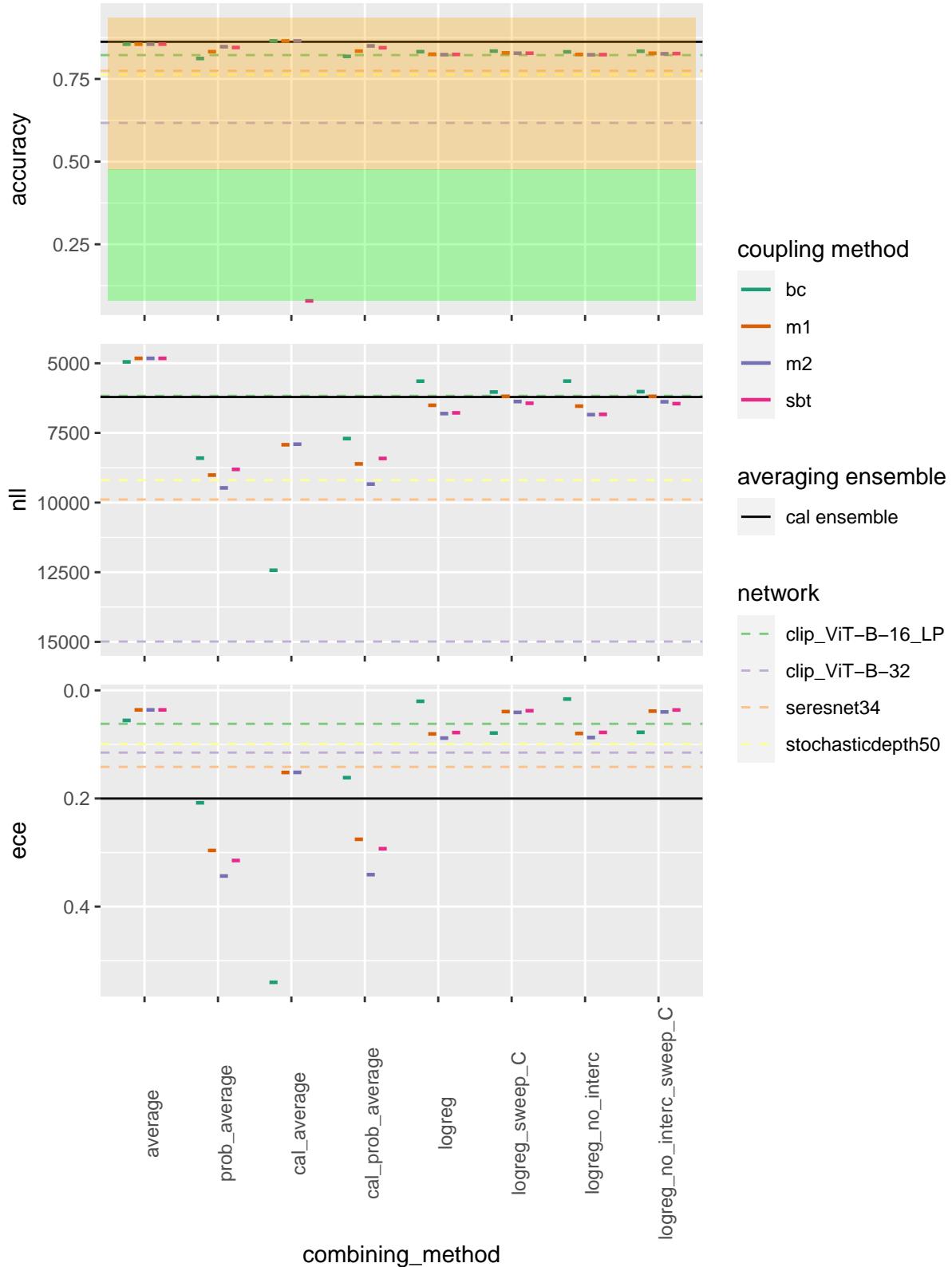
Ensemble metrics  
Error inconsistency 0.458700001239777



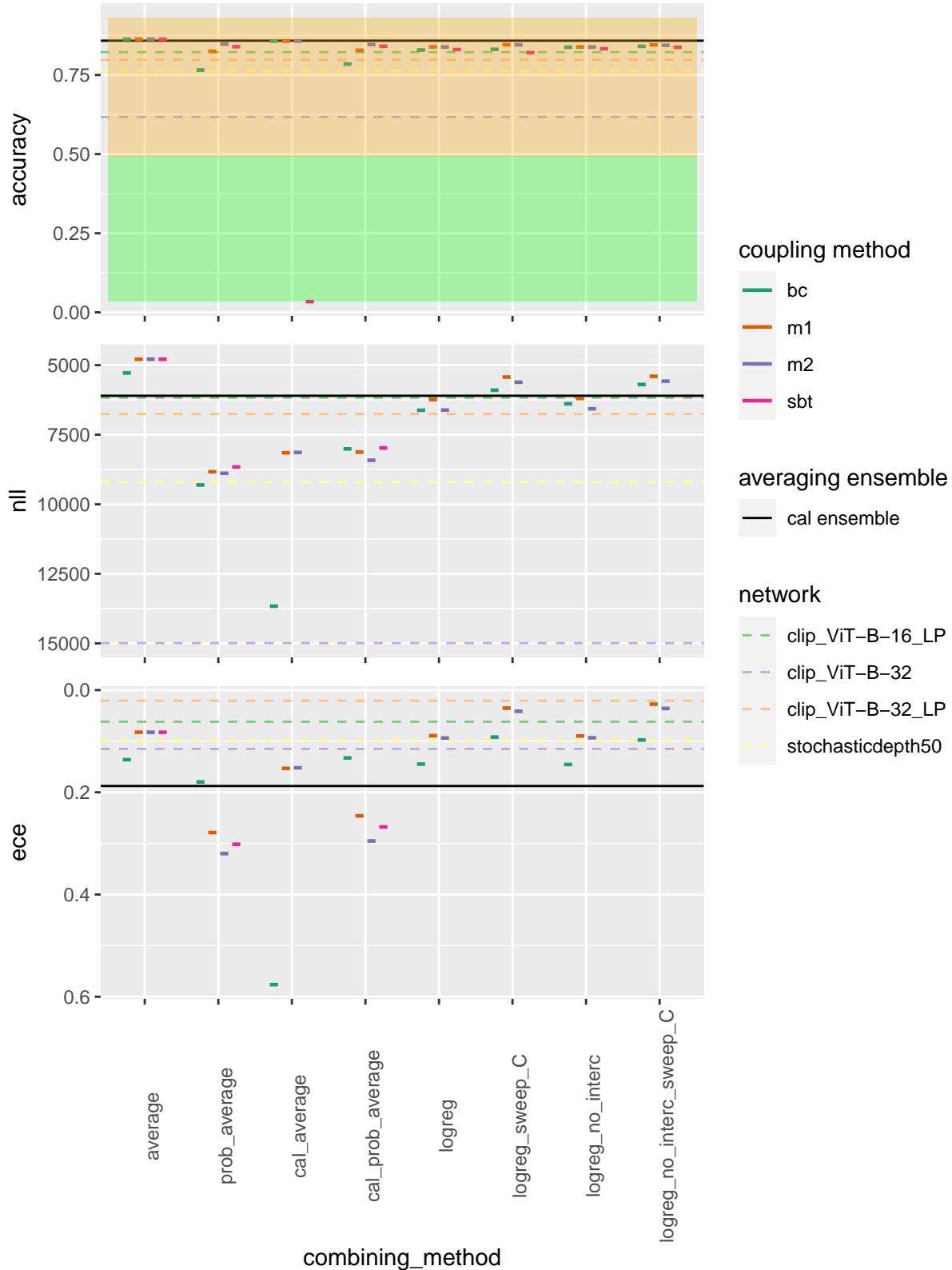
Ensemble metrics  
Error inconsistency 0.476899981498718



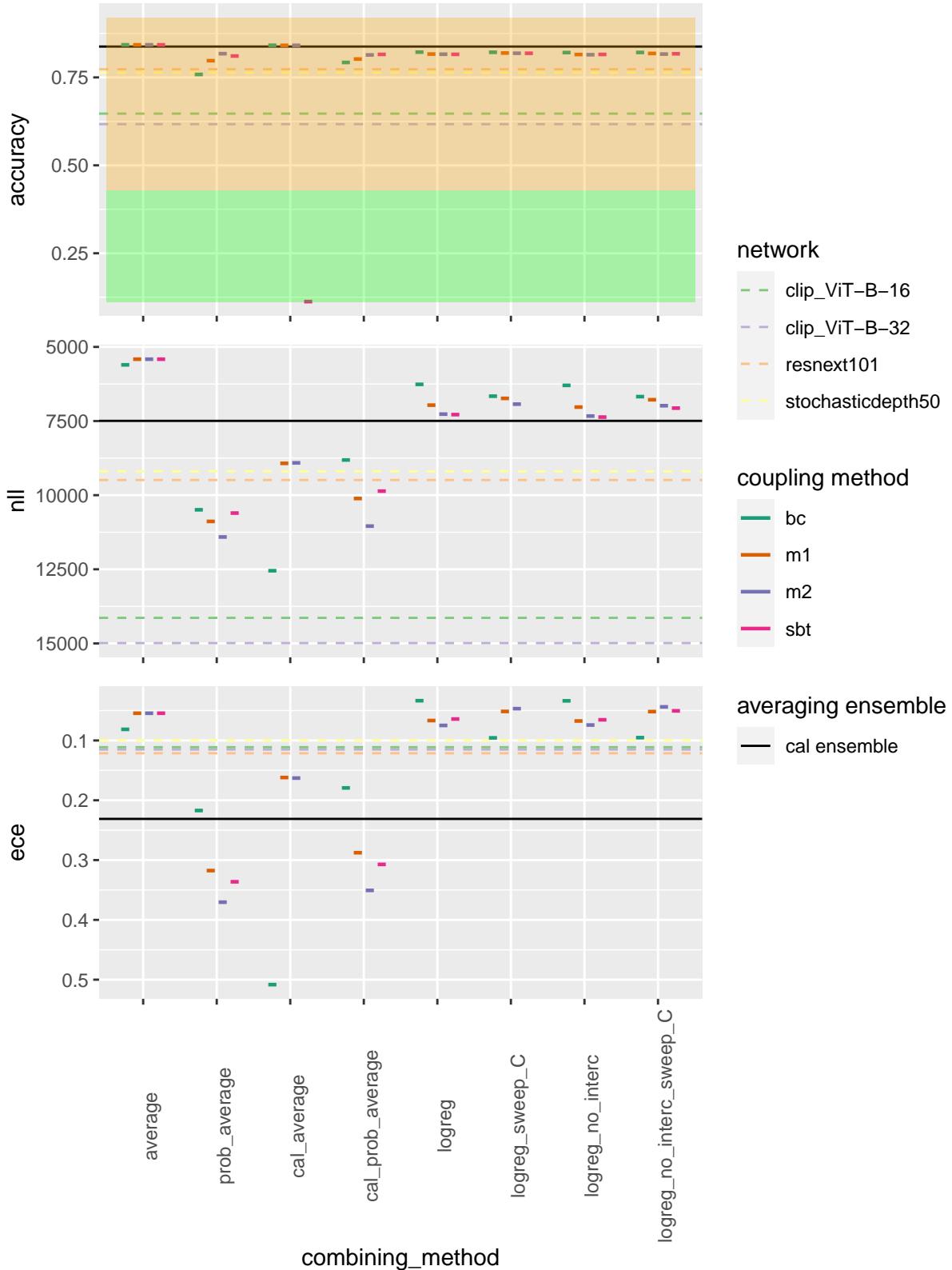
Ensemble metrics  
Error inconsistency 0.459199994802475



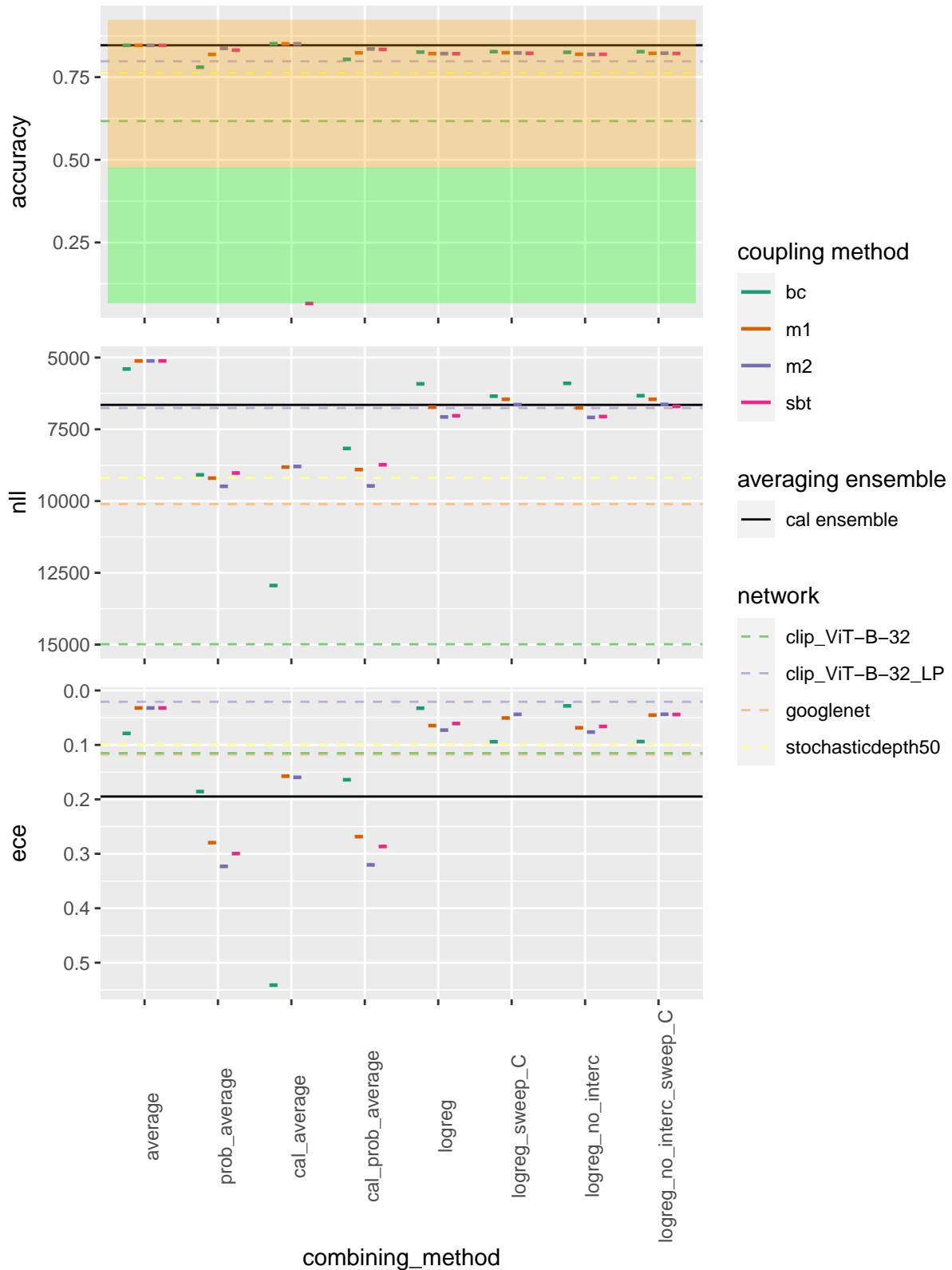
Ensemble metrics  
Error inconsistency 0.43860000371933



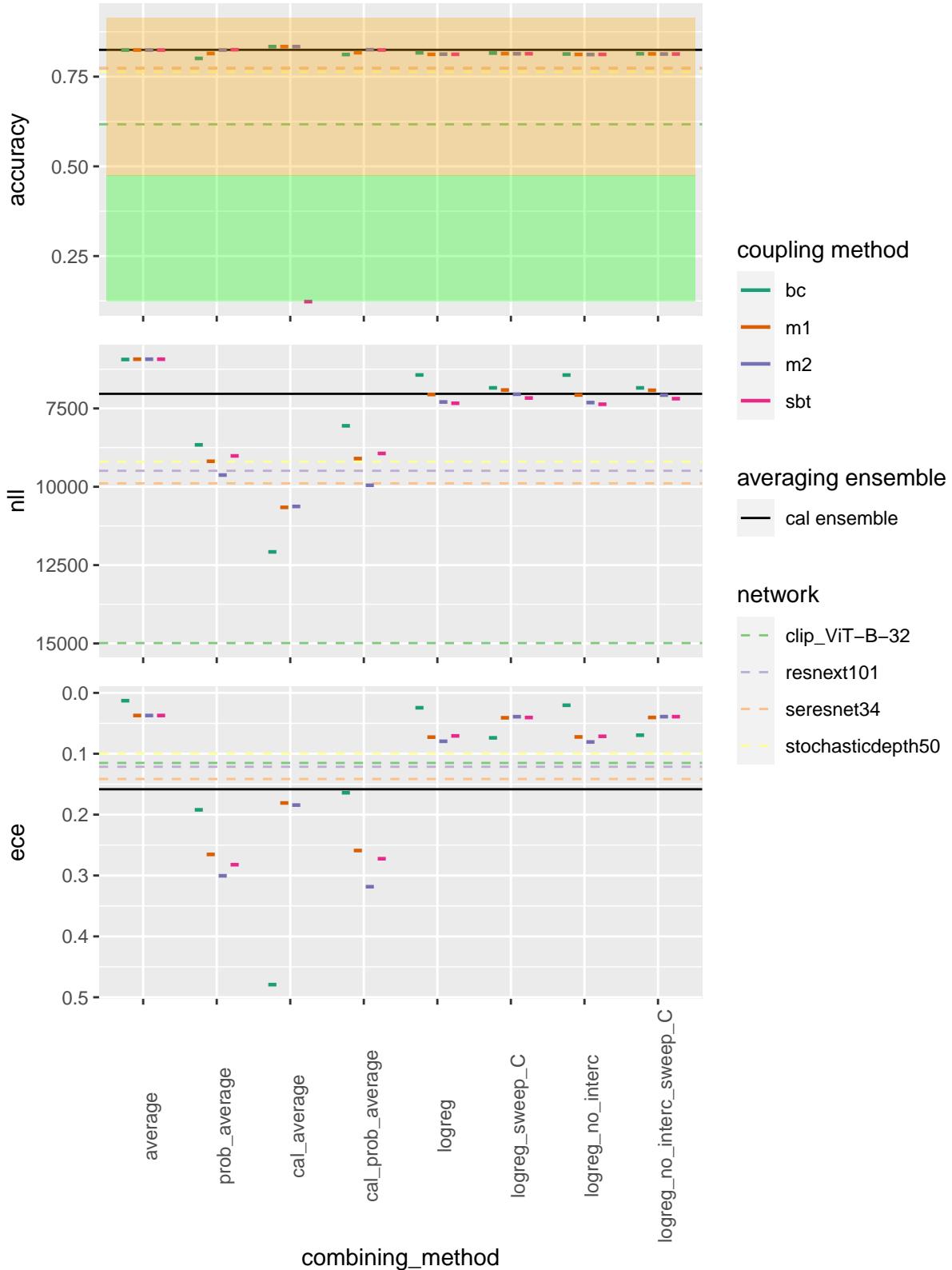
Ensemble metrics  
Error inconsistency 0.490699976682663



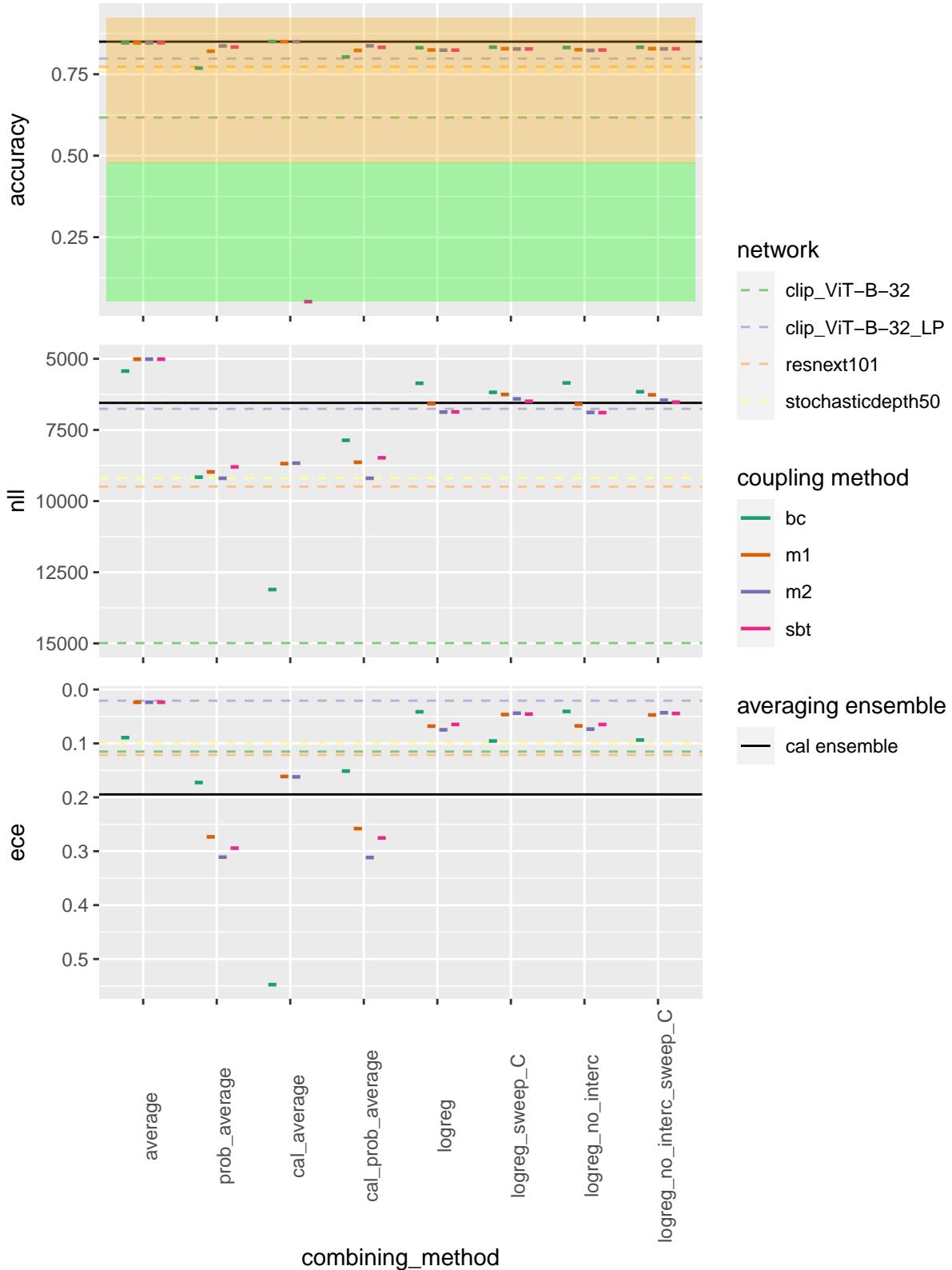
Ensemble metrics  
Error inconsistency 0.44829997420311



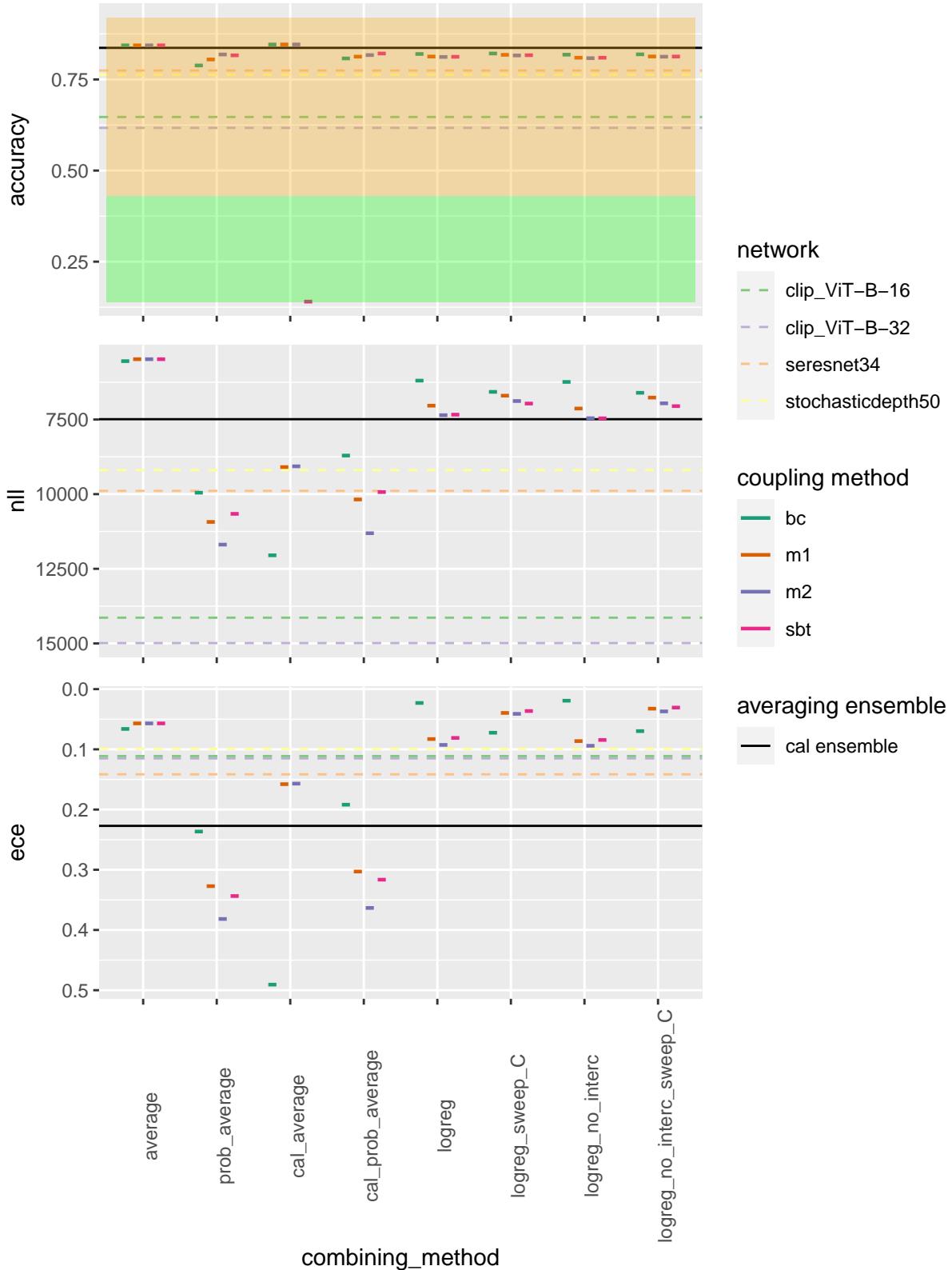
Ensemble metrics  
Error inconsistency 0.439899981021881



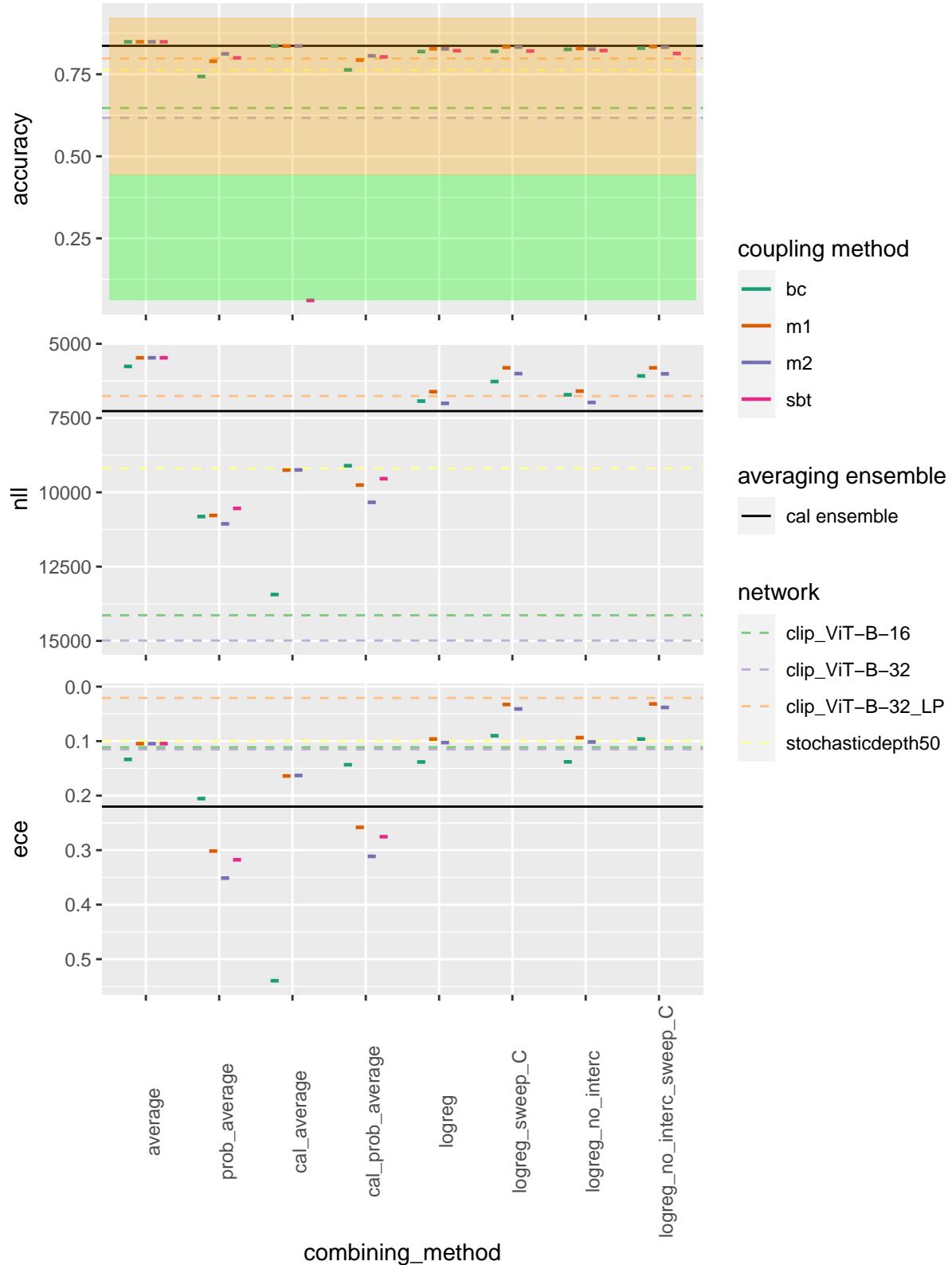
Ensemble metrics  
Error inconsistency 0.444700002670288



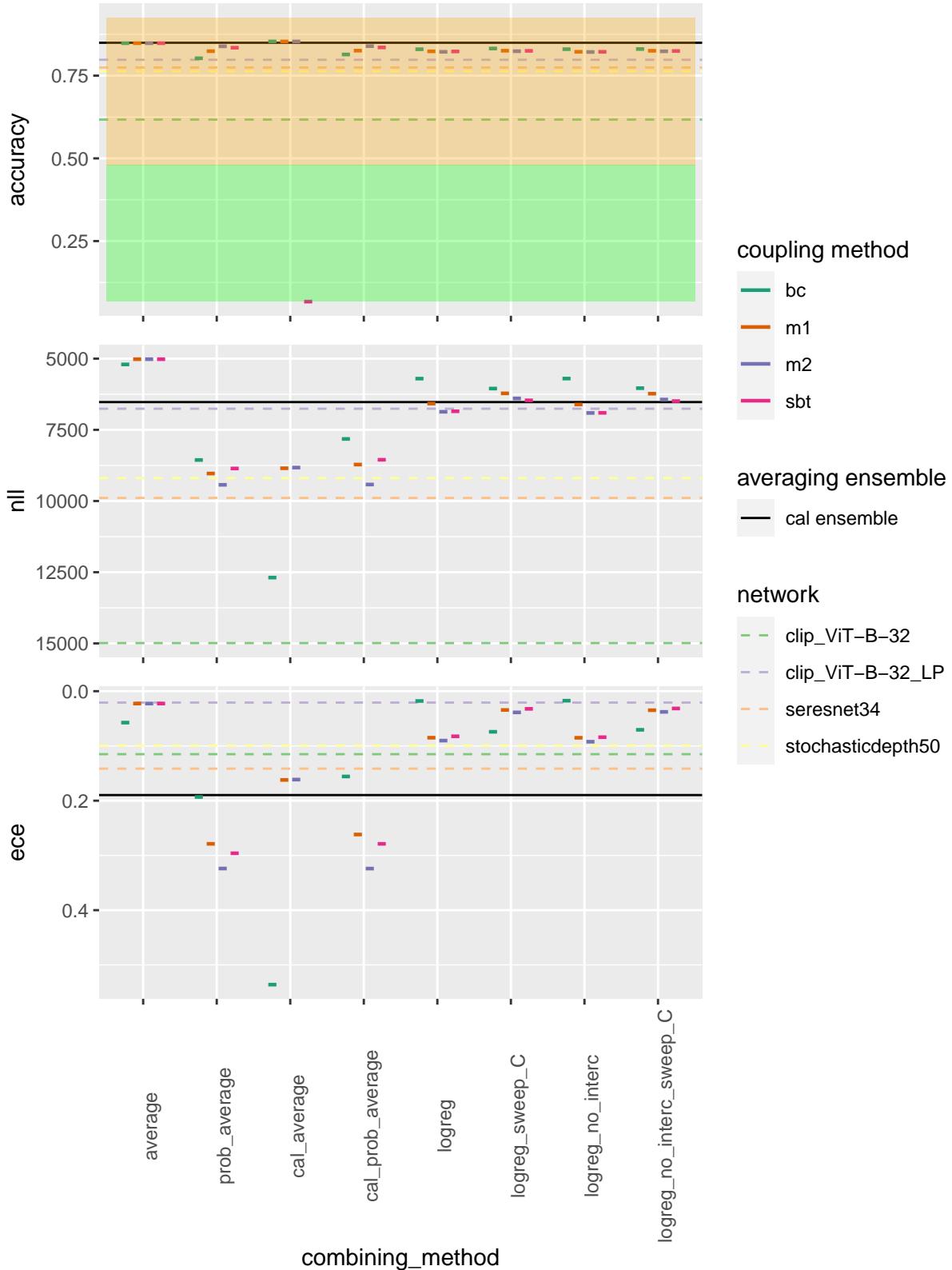
Ensemble metrics  
Error inconsistency 0.489099979400635



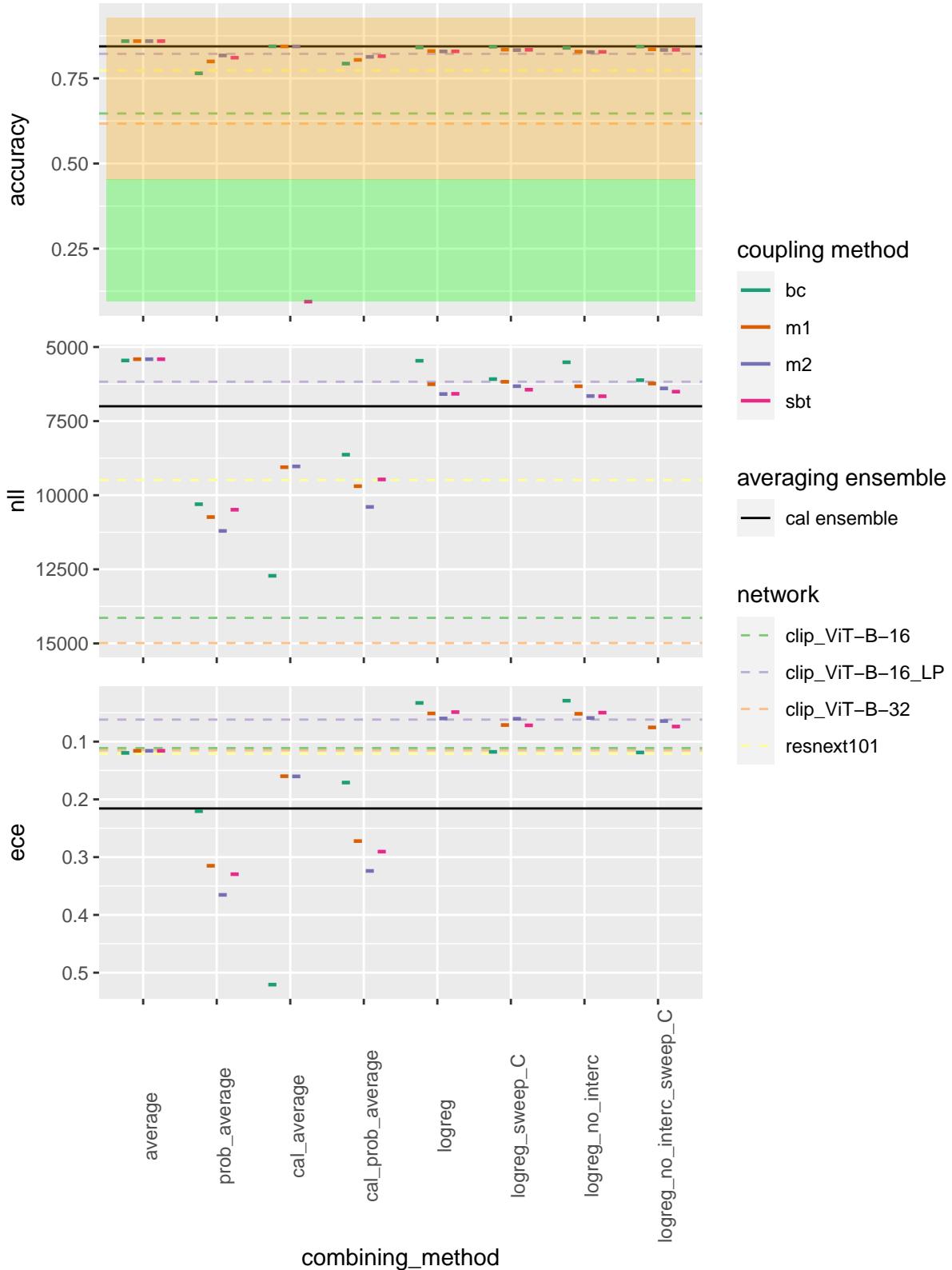
Ensemble metrics  
Error inconsistency 0.479299992322922



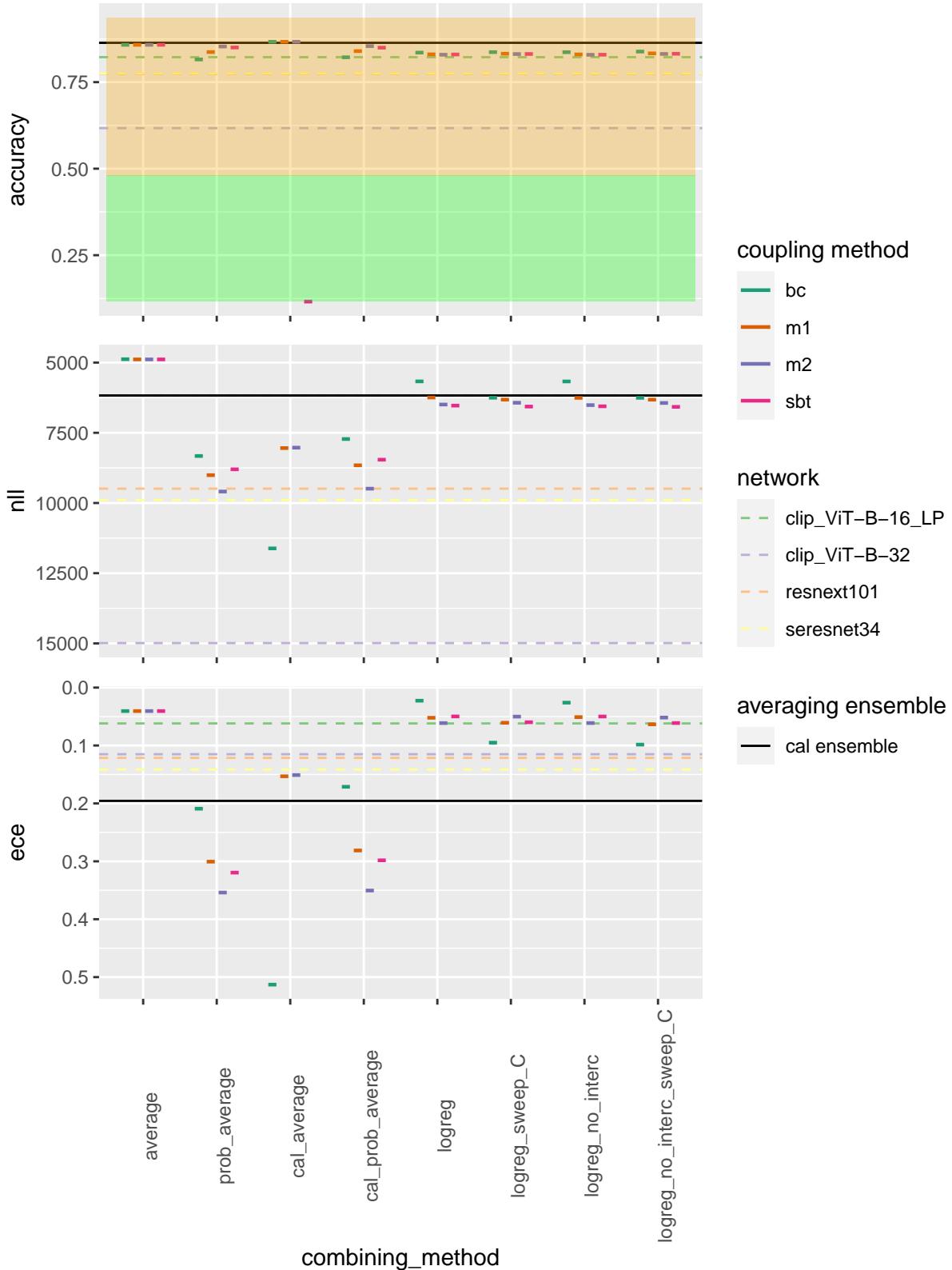
Ensemble metrics  
Error inconsistency 0.445499986410141



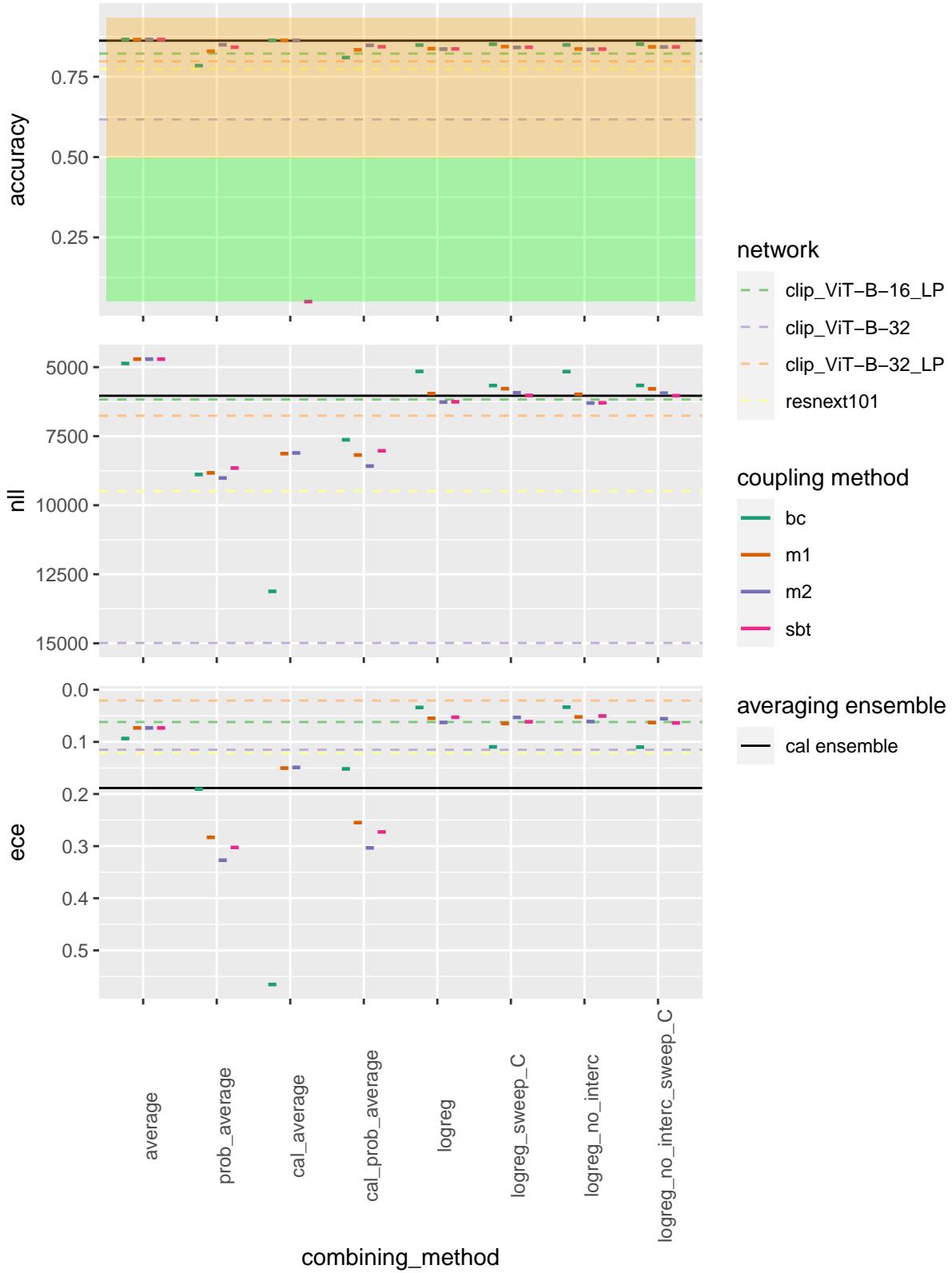
Ensemble metrics  
Error inconsistency 0.475499987602234



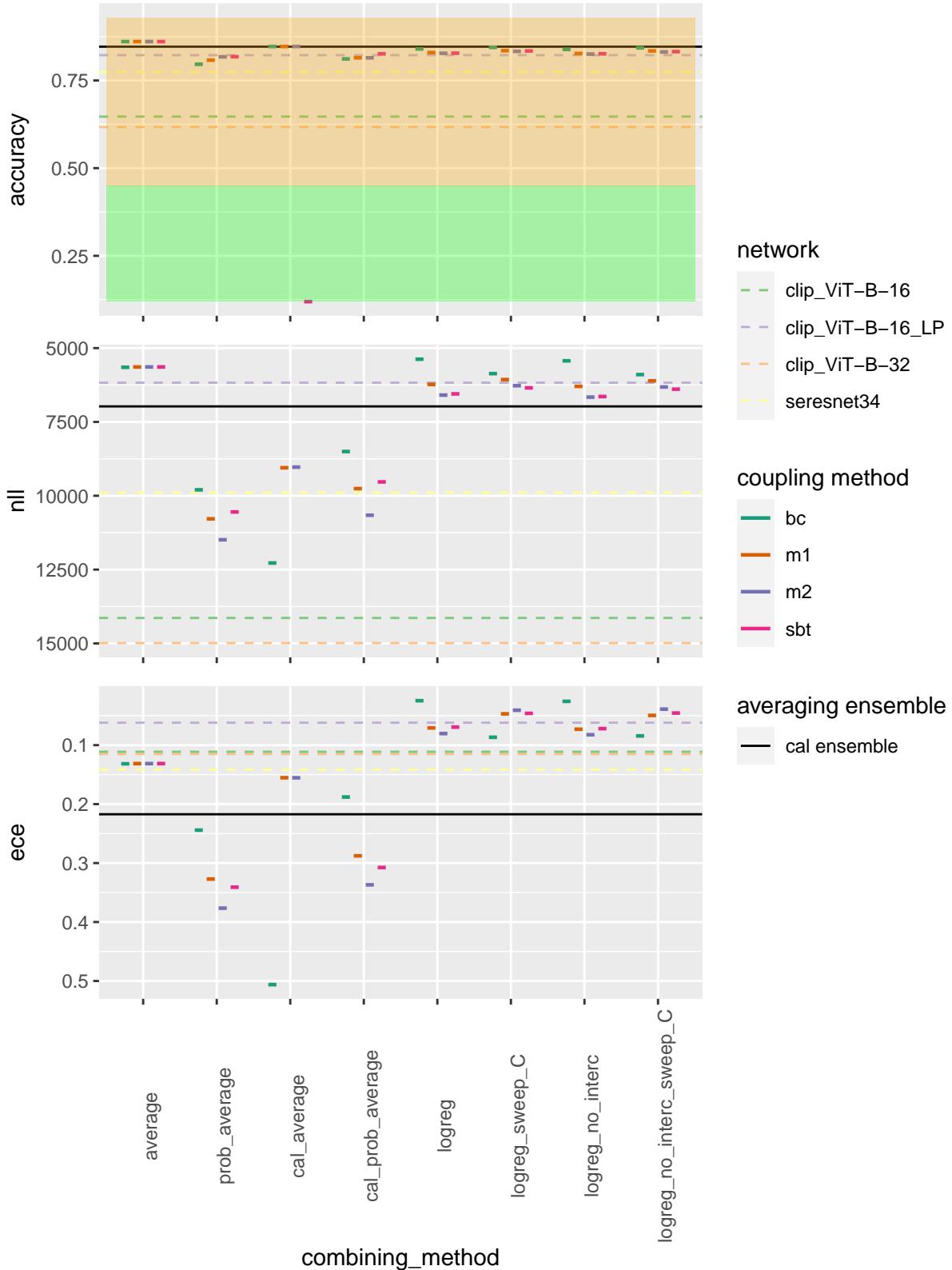
Ensemble metrics  
Error inconsistency 0.455499976873398



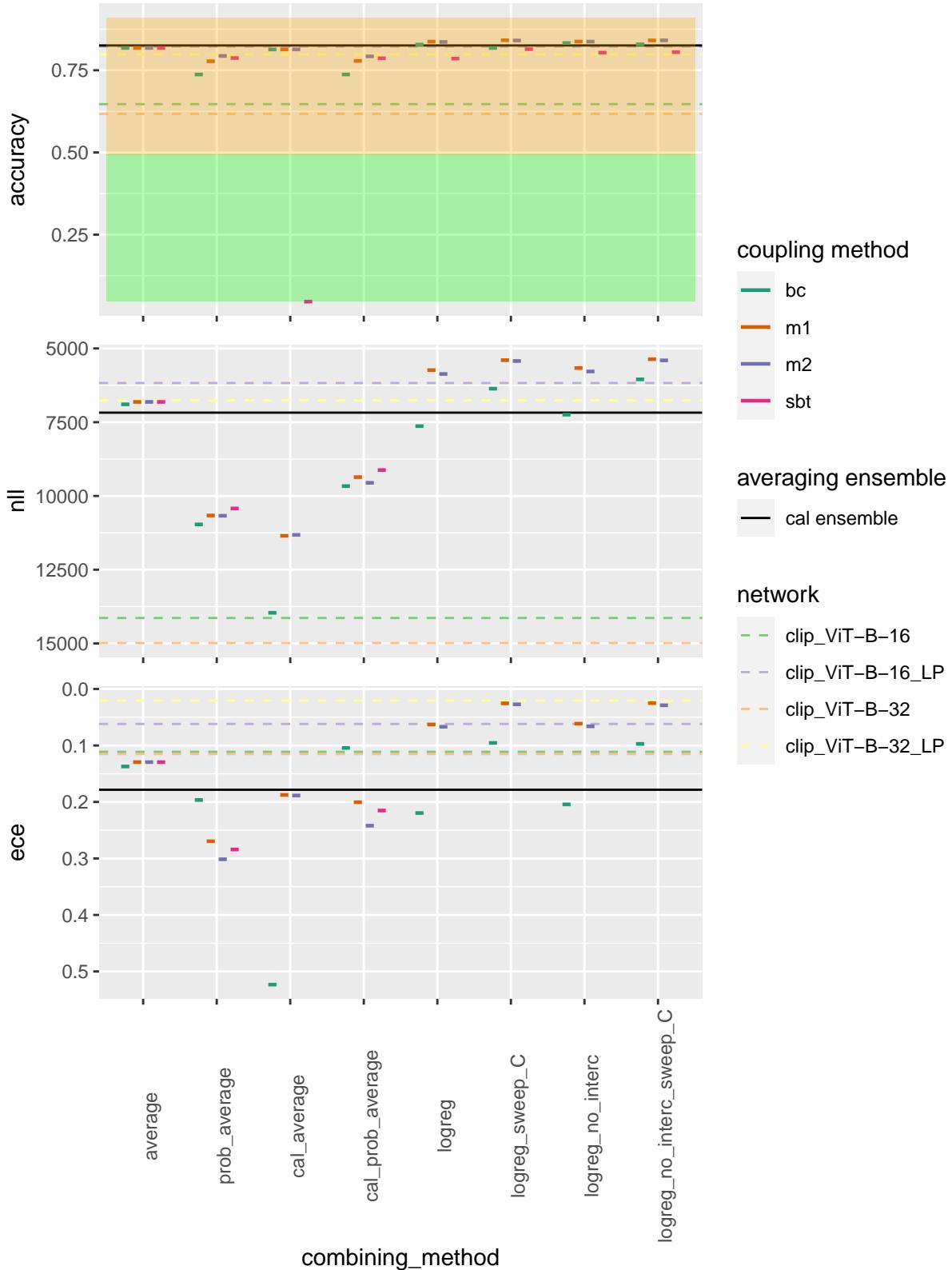
Ensemble metrics  
Error inconsistency 0.436199992895126



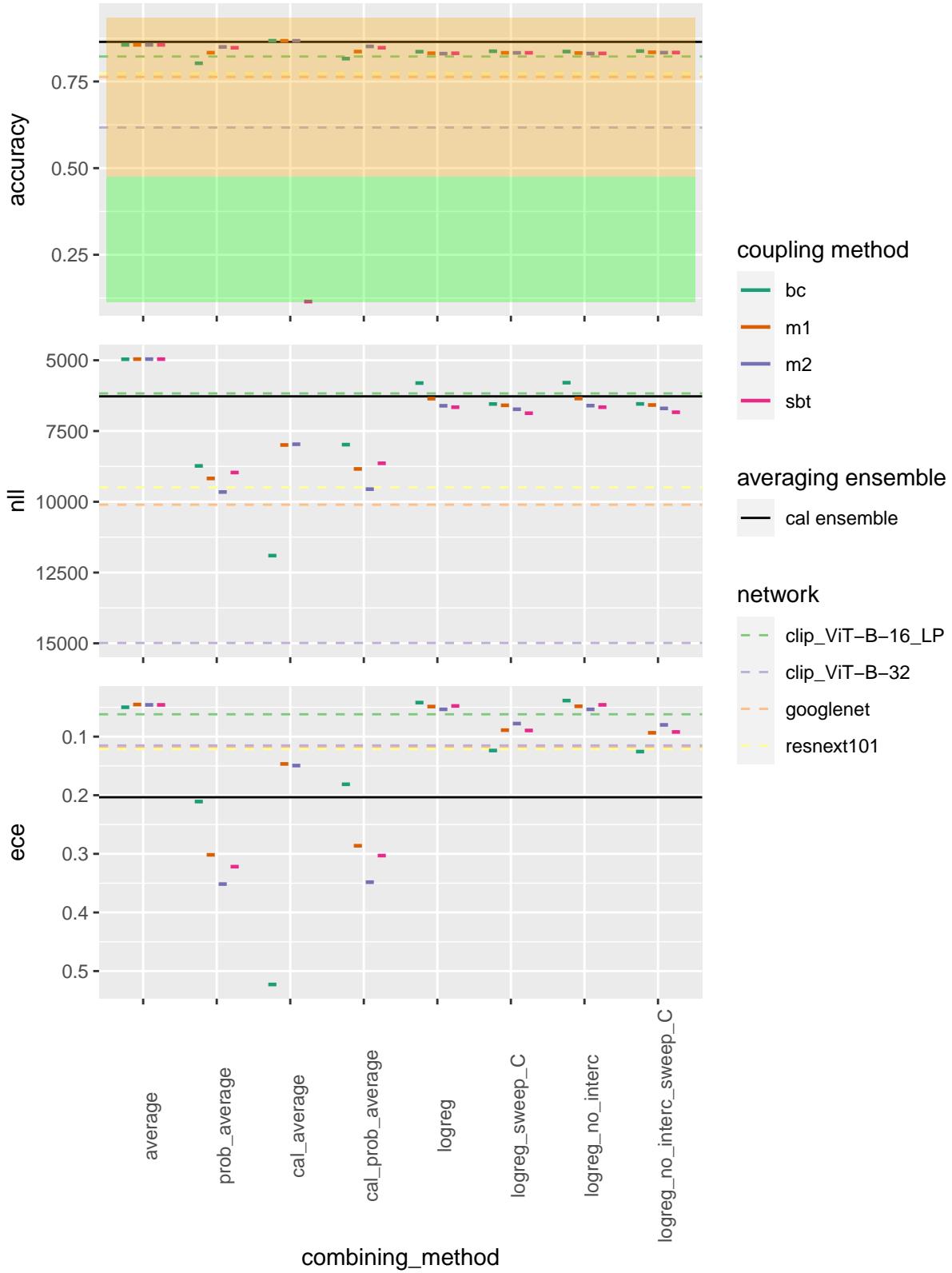
Ensemble metrics  
Error inconsistency 0.479099988937378



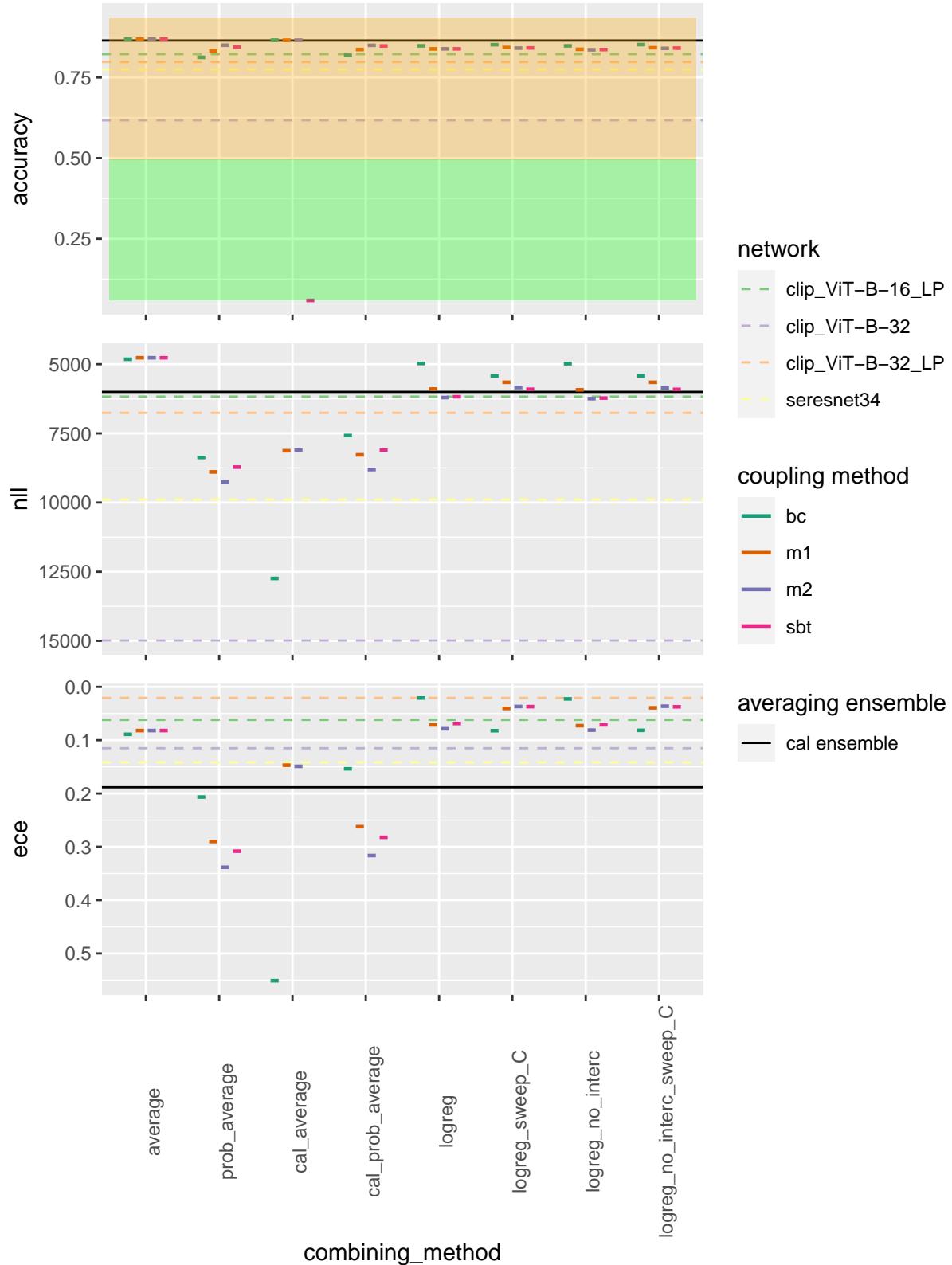
Ensemble metrics  
Error inconsistency 0.416699975728989



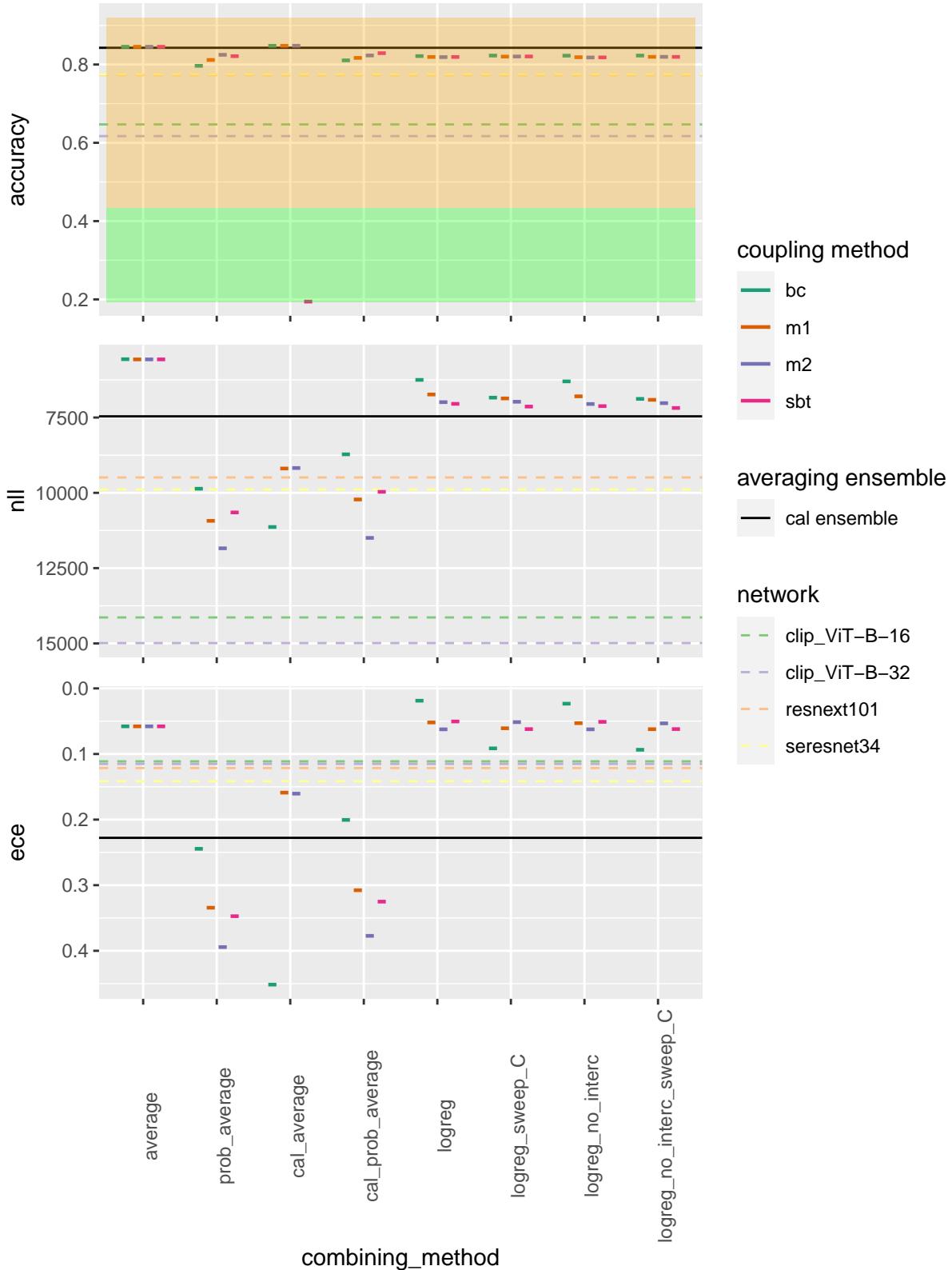
Ensemble metrics  
Error inconsistency 0.458999991416931



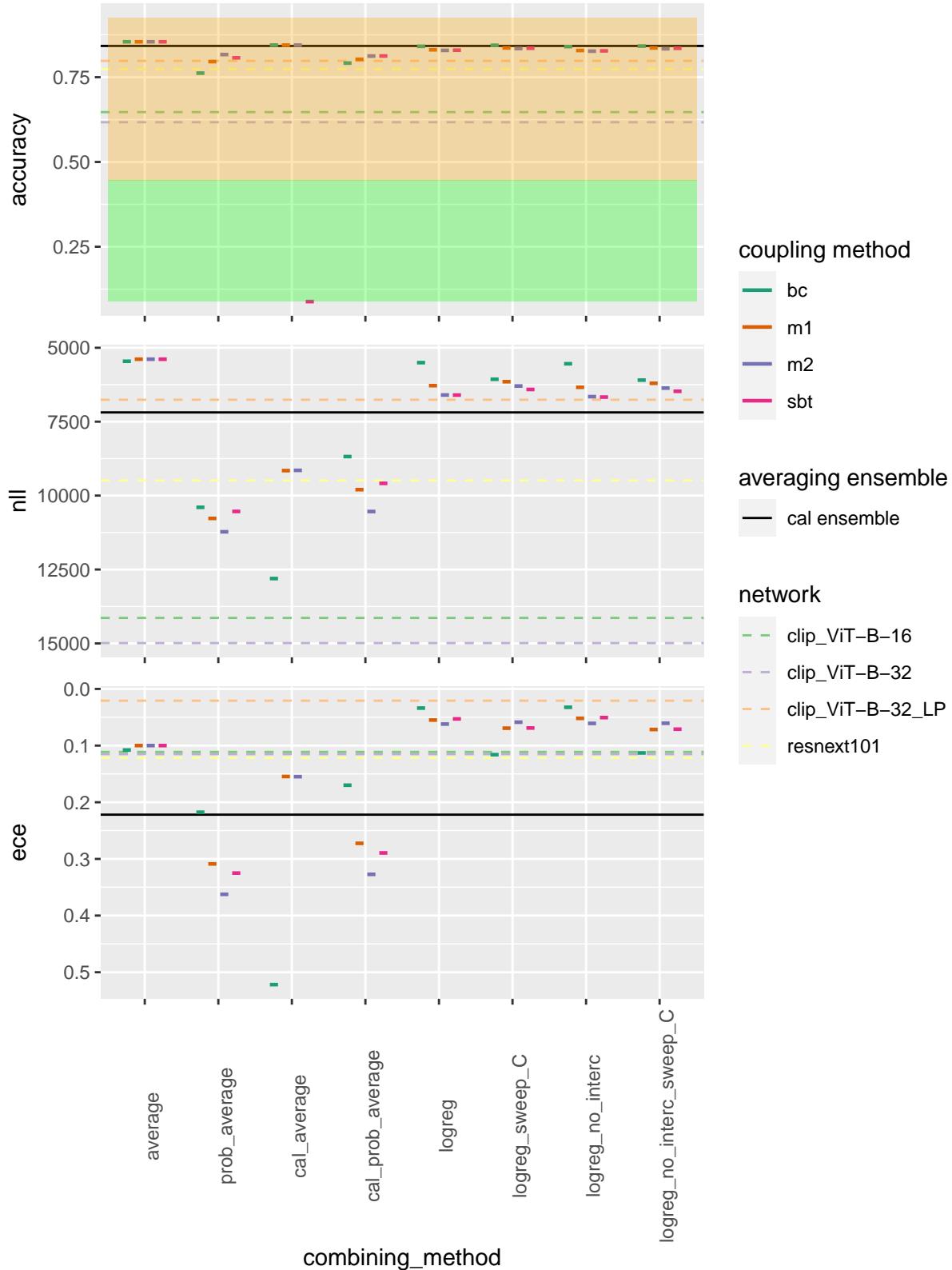
Ensemble metrics  
Error inconsistency 0.440999984741211



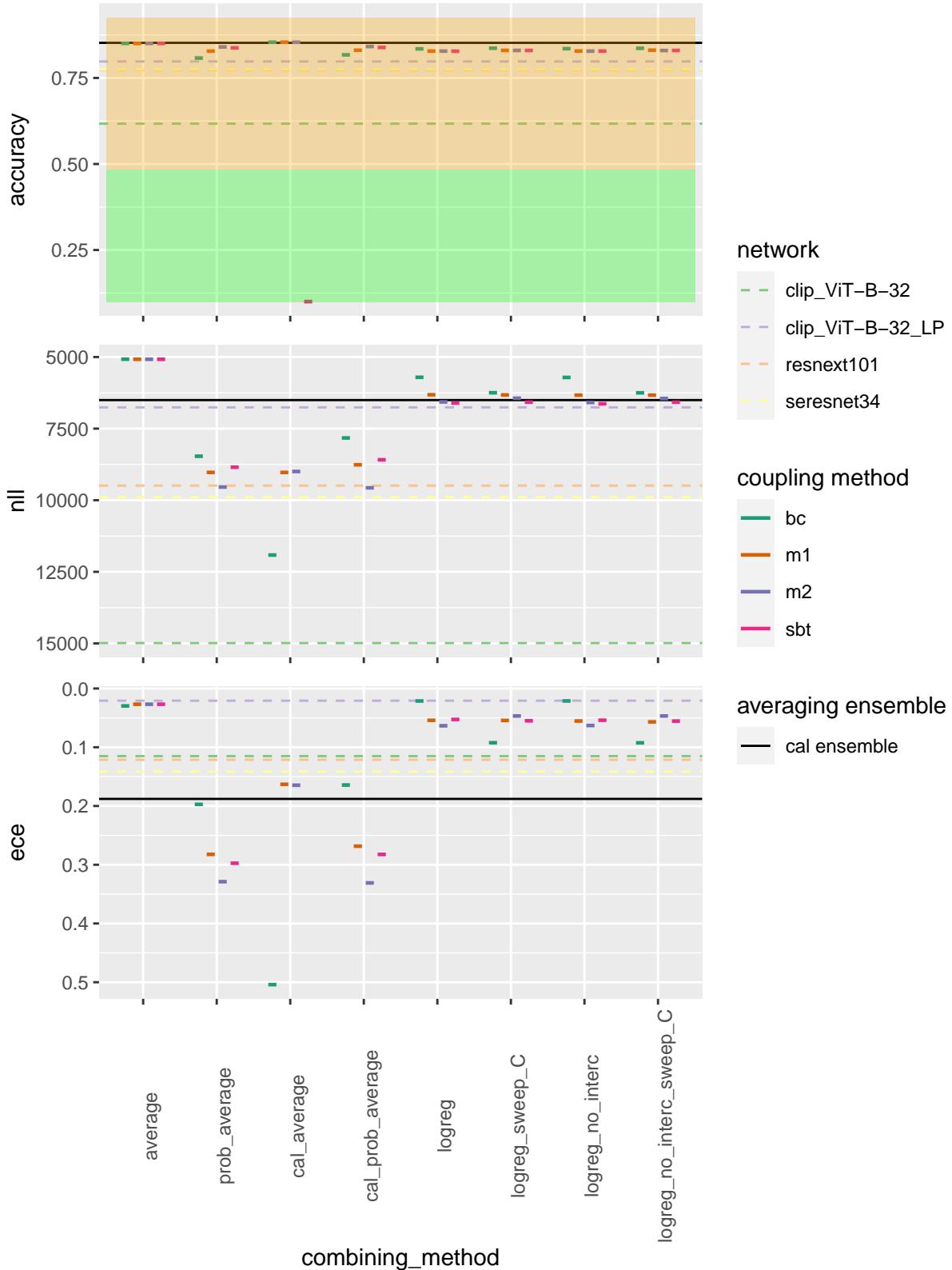
Ensemble metrics  
Error inconsistency 0.486099988222122



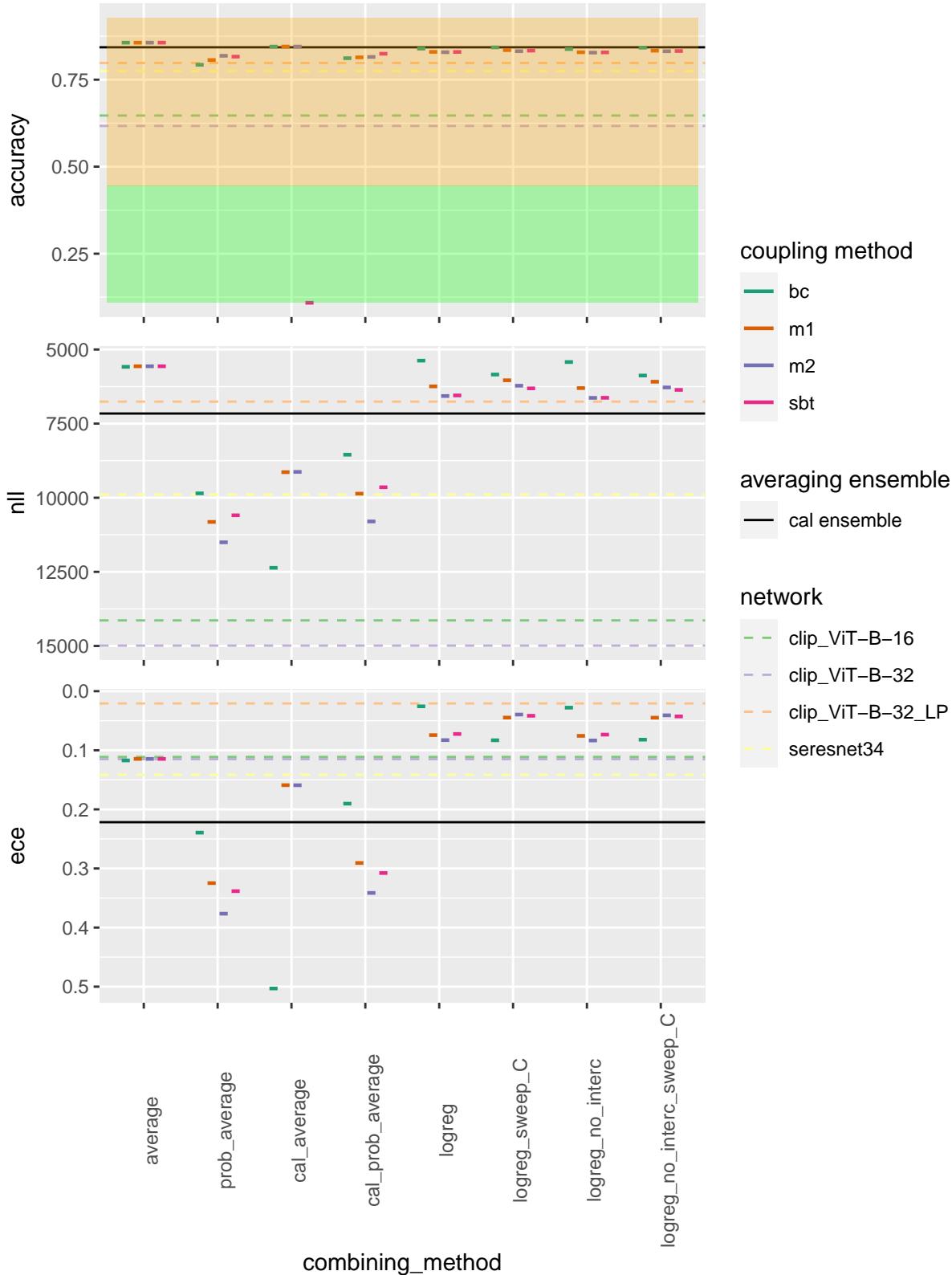
Ensemble metrics  
Error inconsistency 0.47869998216629



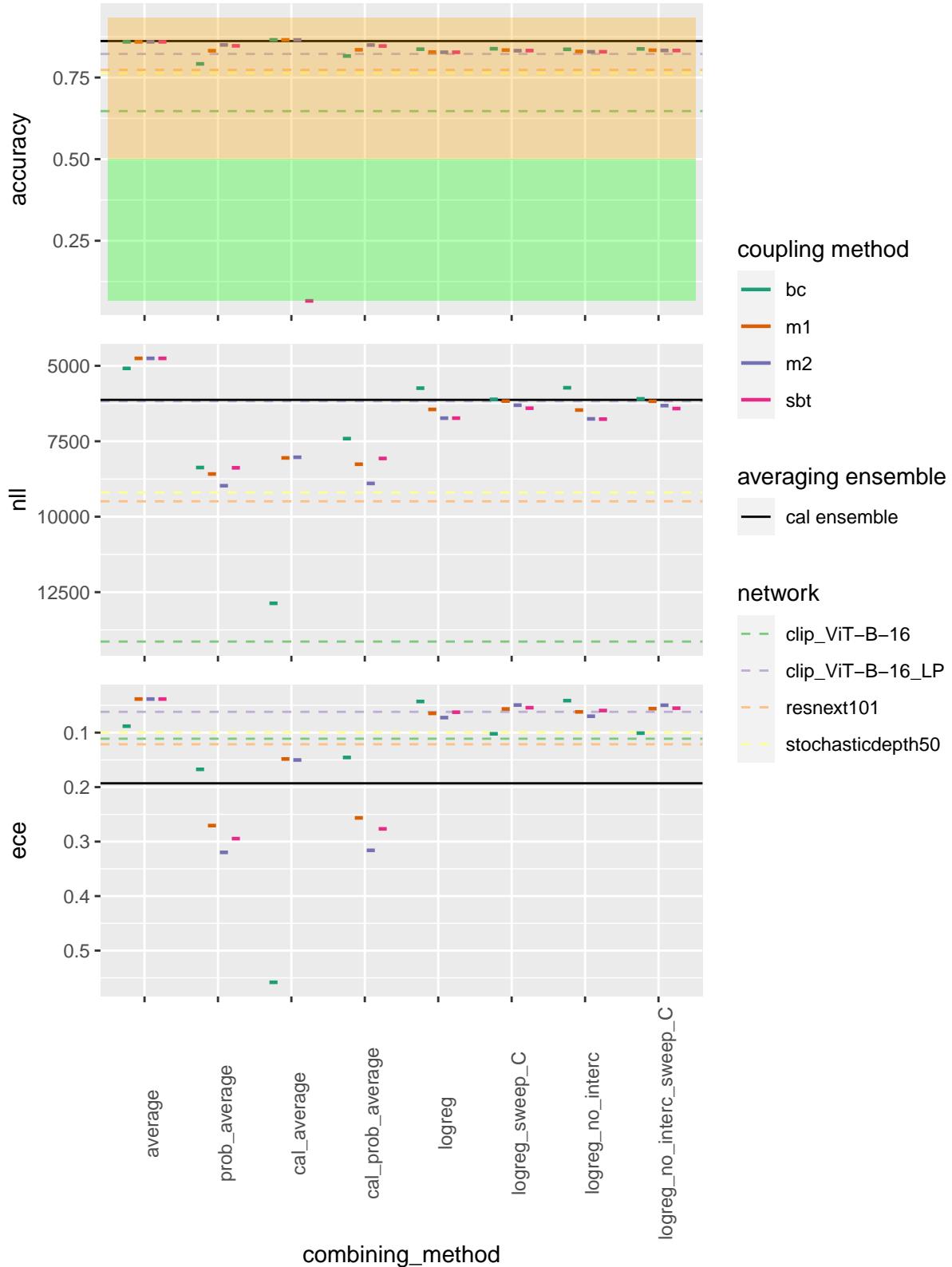
Ensemble metrics  
Error inconsistency 0.441199988126755



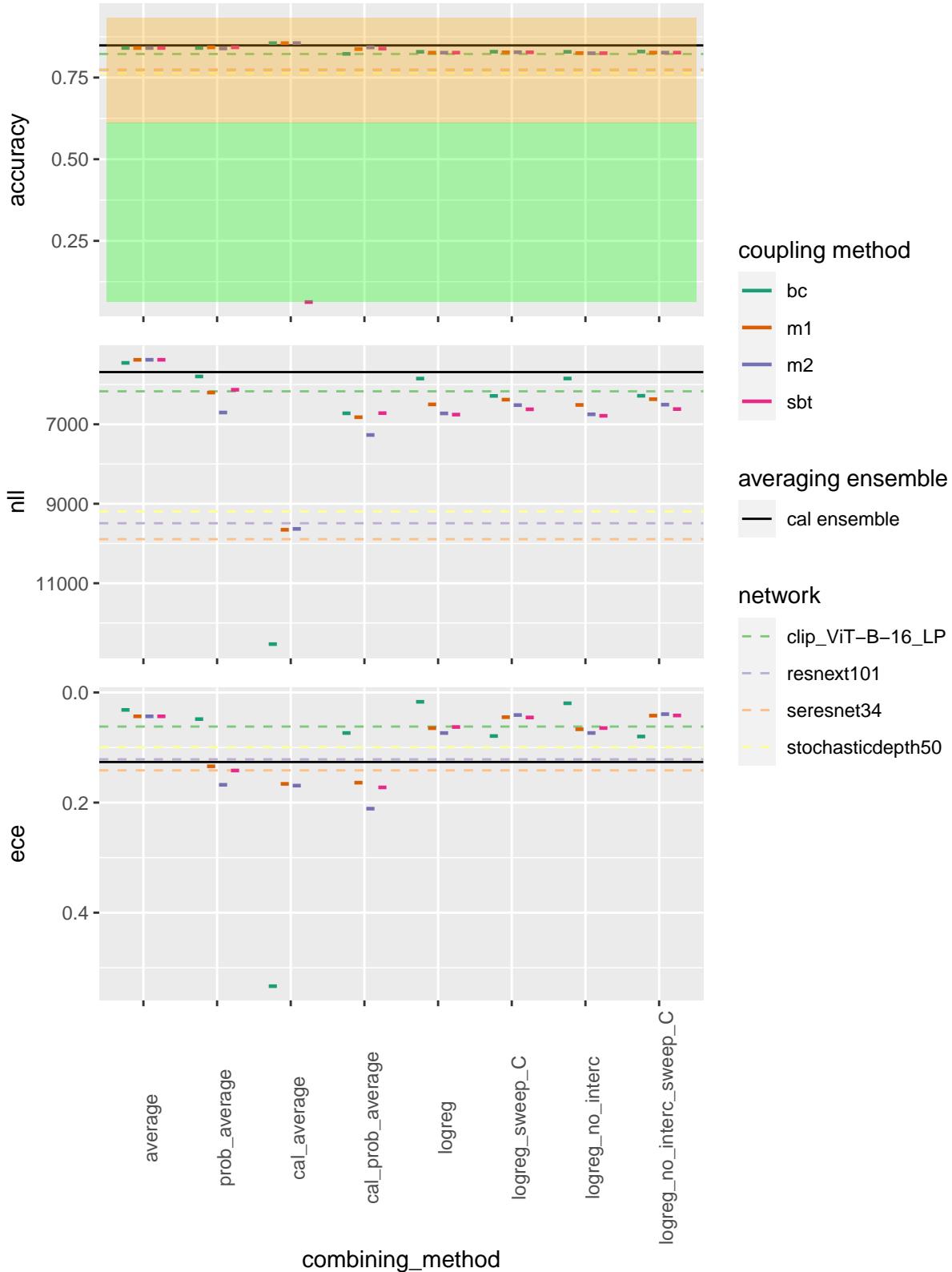
Ensemble metrics  
Error inconsistency 0.482899993658066



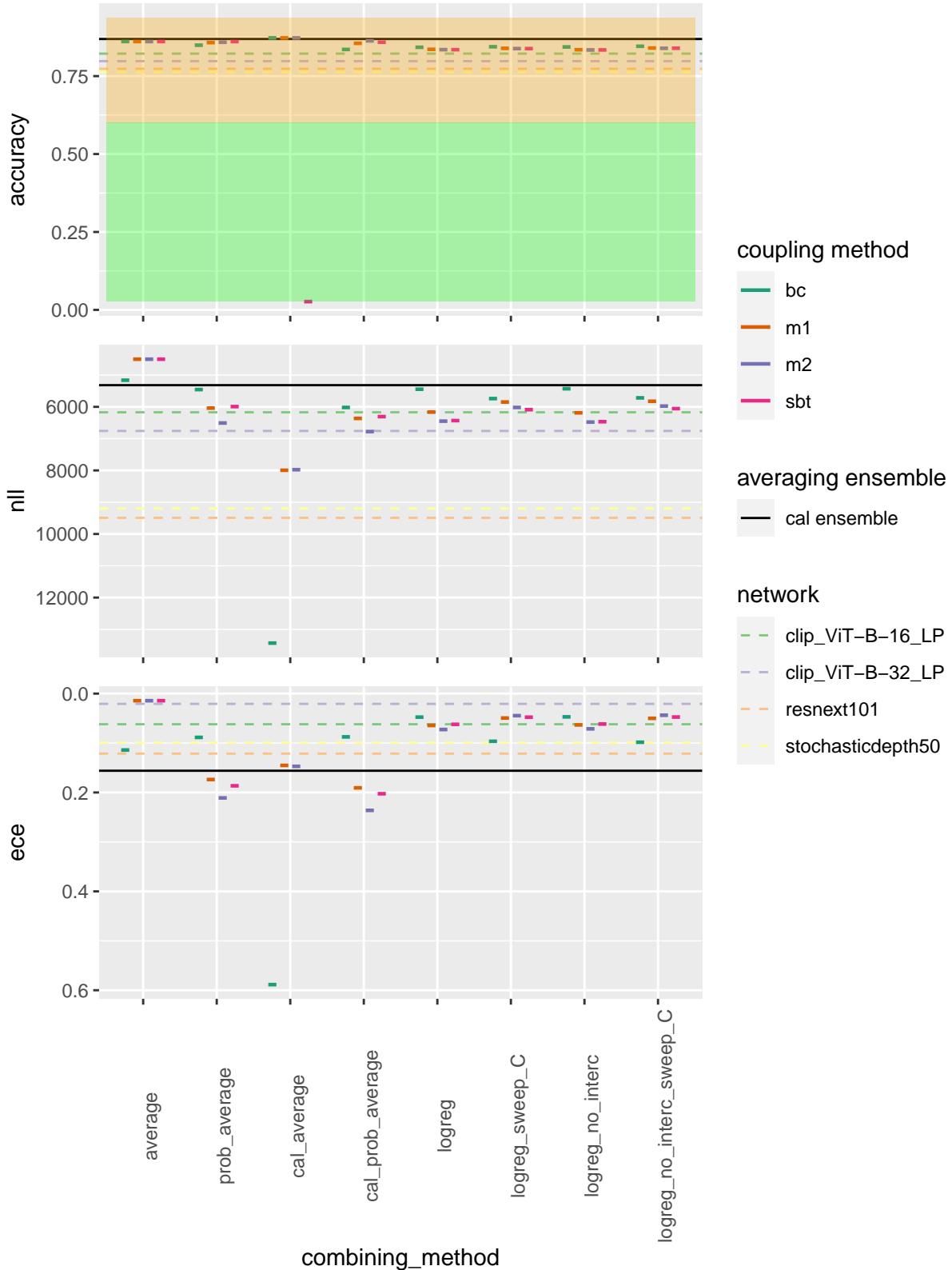
Ensemble metrics  
Error inconsistency 0.433699995279312



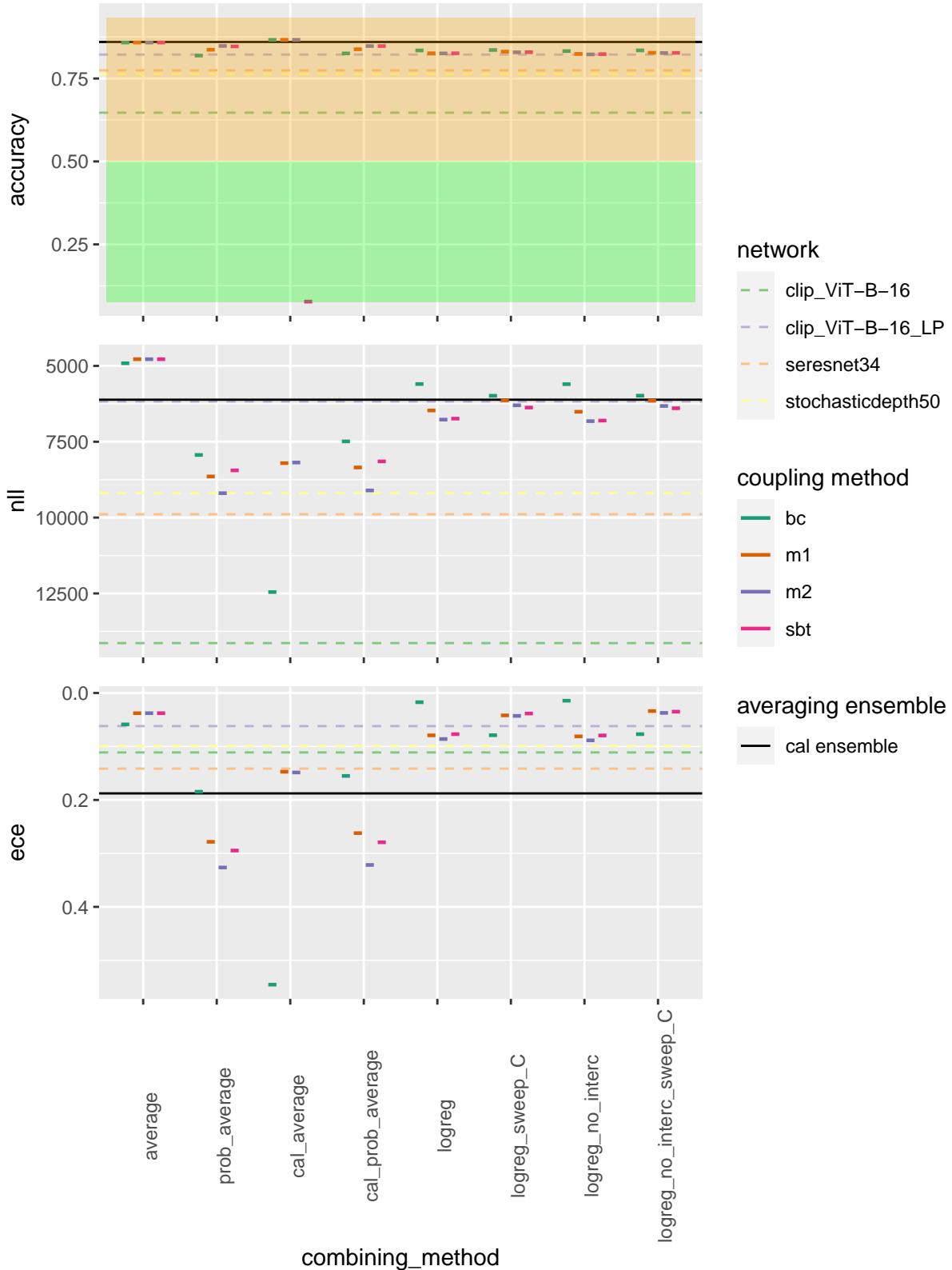
Ensemble metrics  
Error inconsistency 0.321399986743927



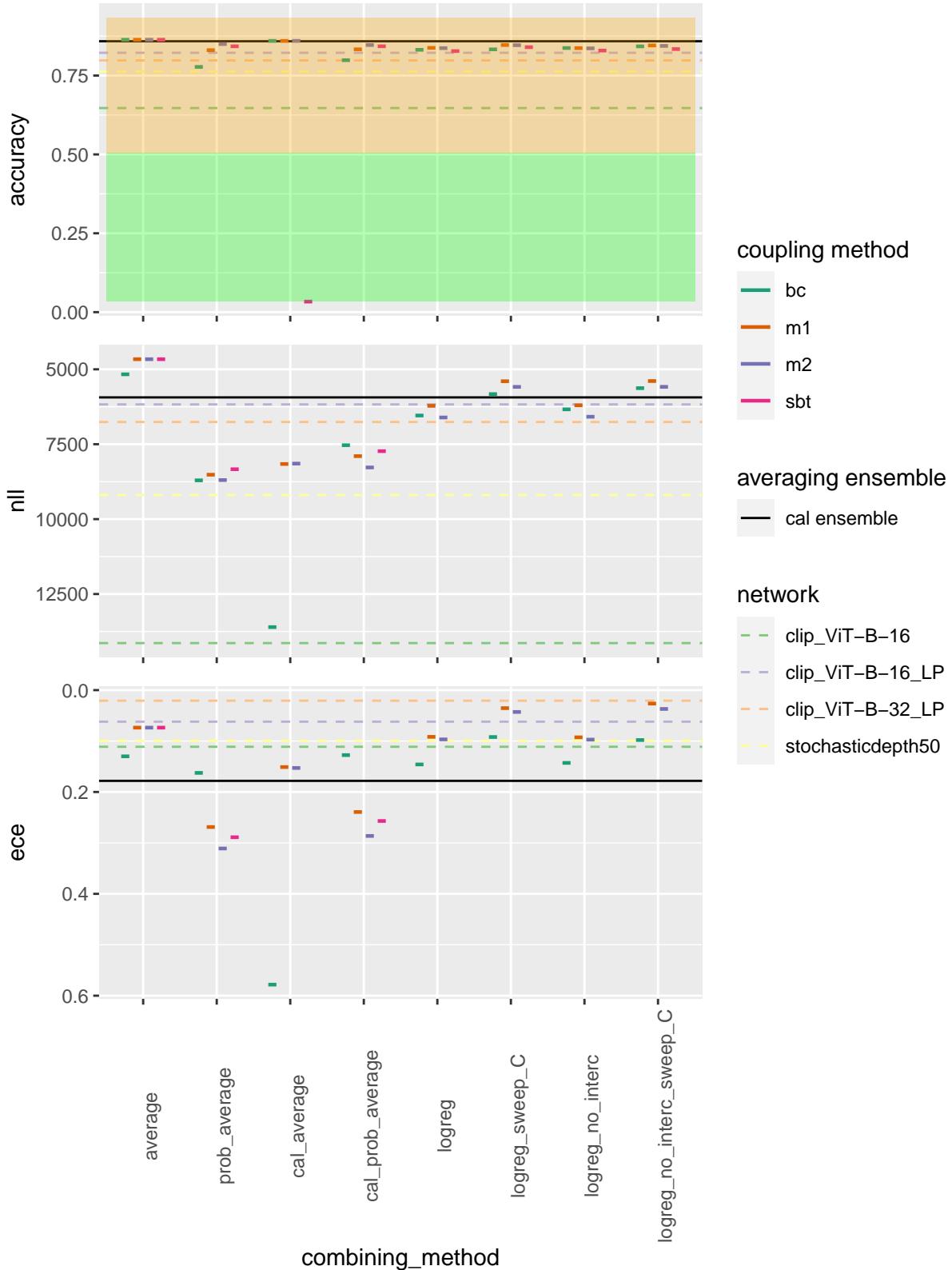
Ensemble metrics  
Error inconsistency 0.337999999523163



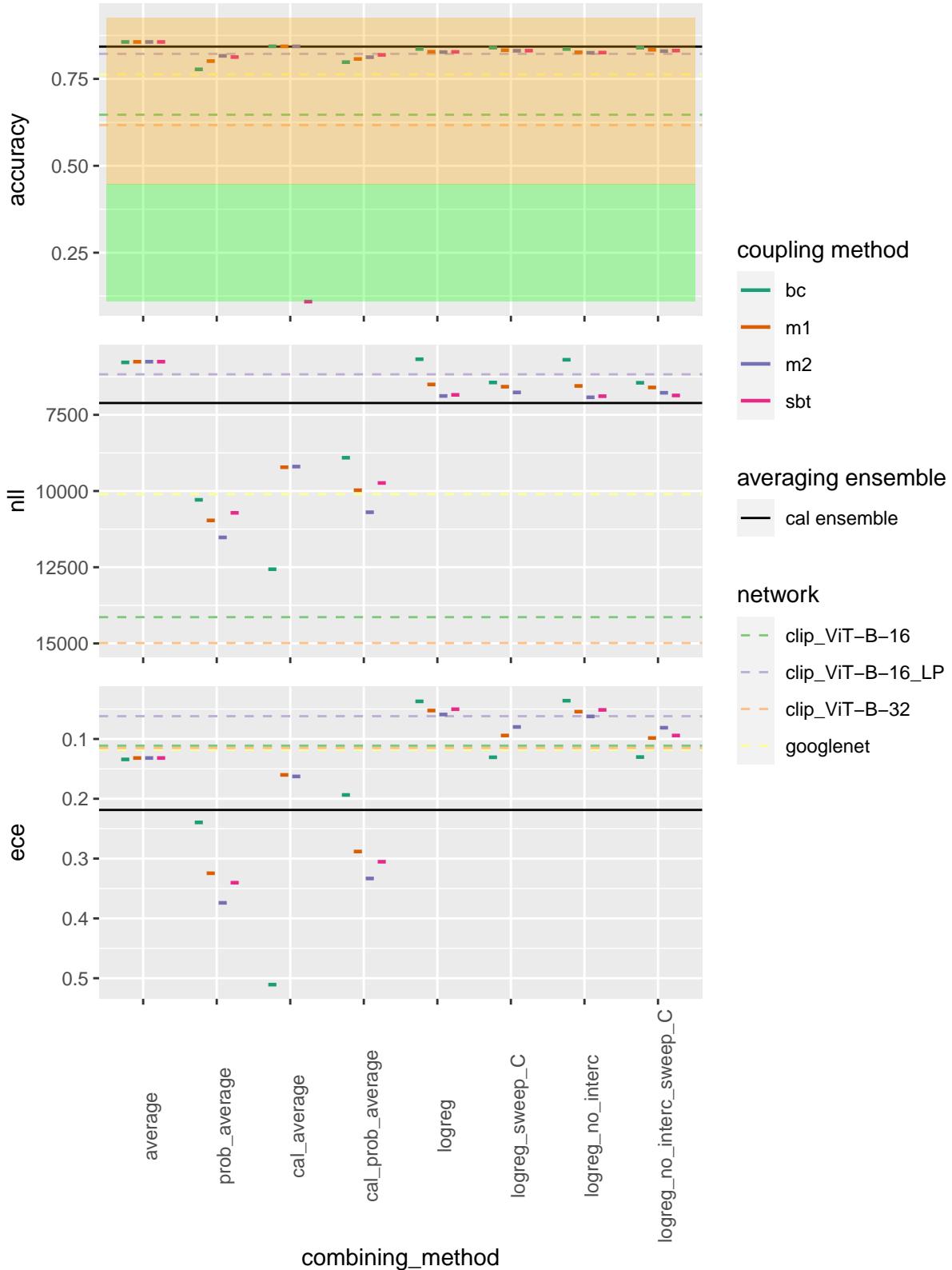
Ensemble metrics  
Error inconsistency 0.433699995279312



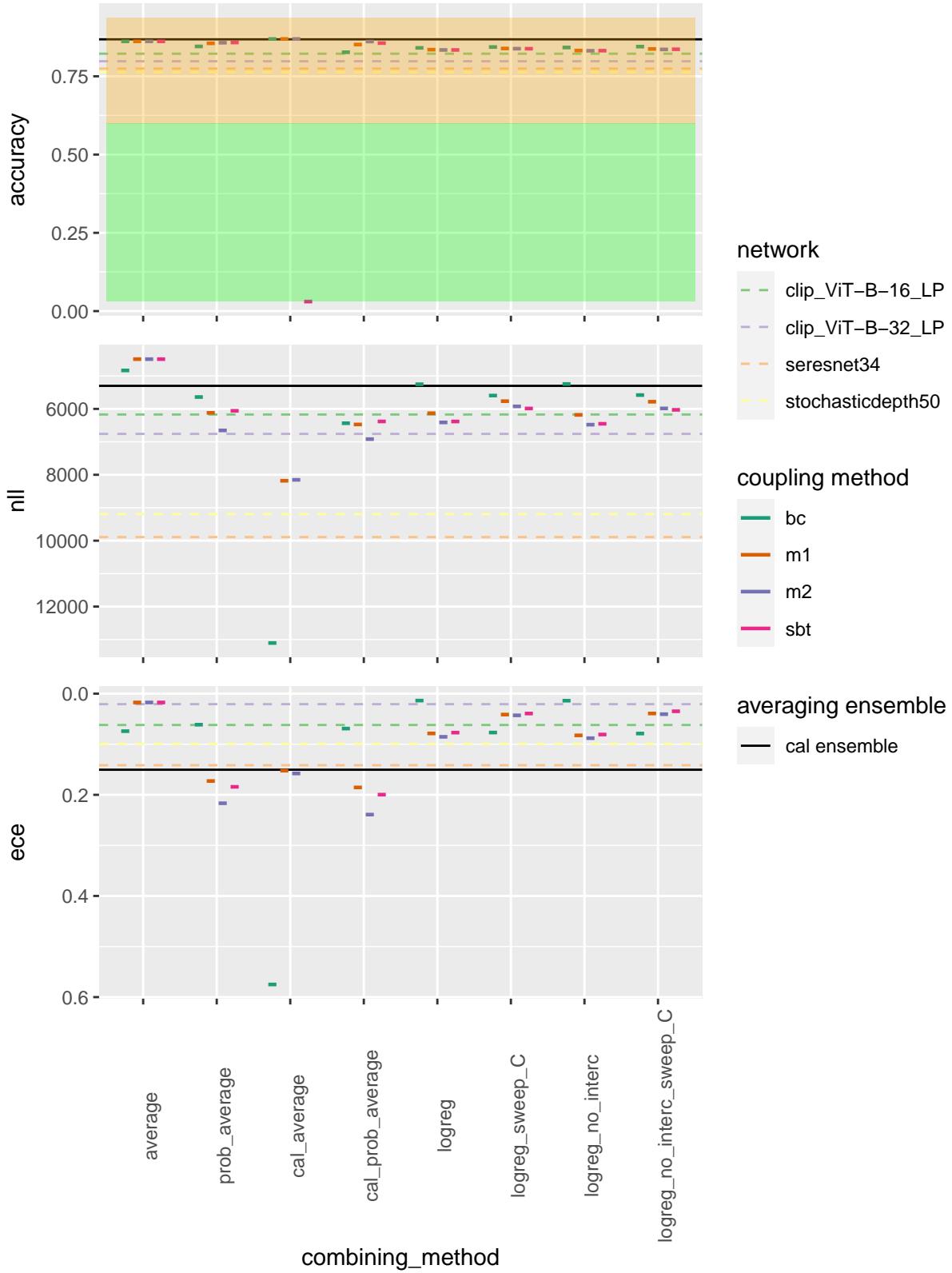
Ensemble metrics  
Error inconsistency 0.429499983787537



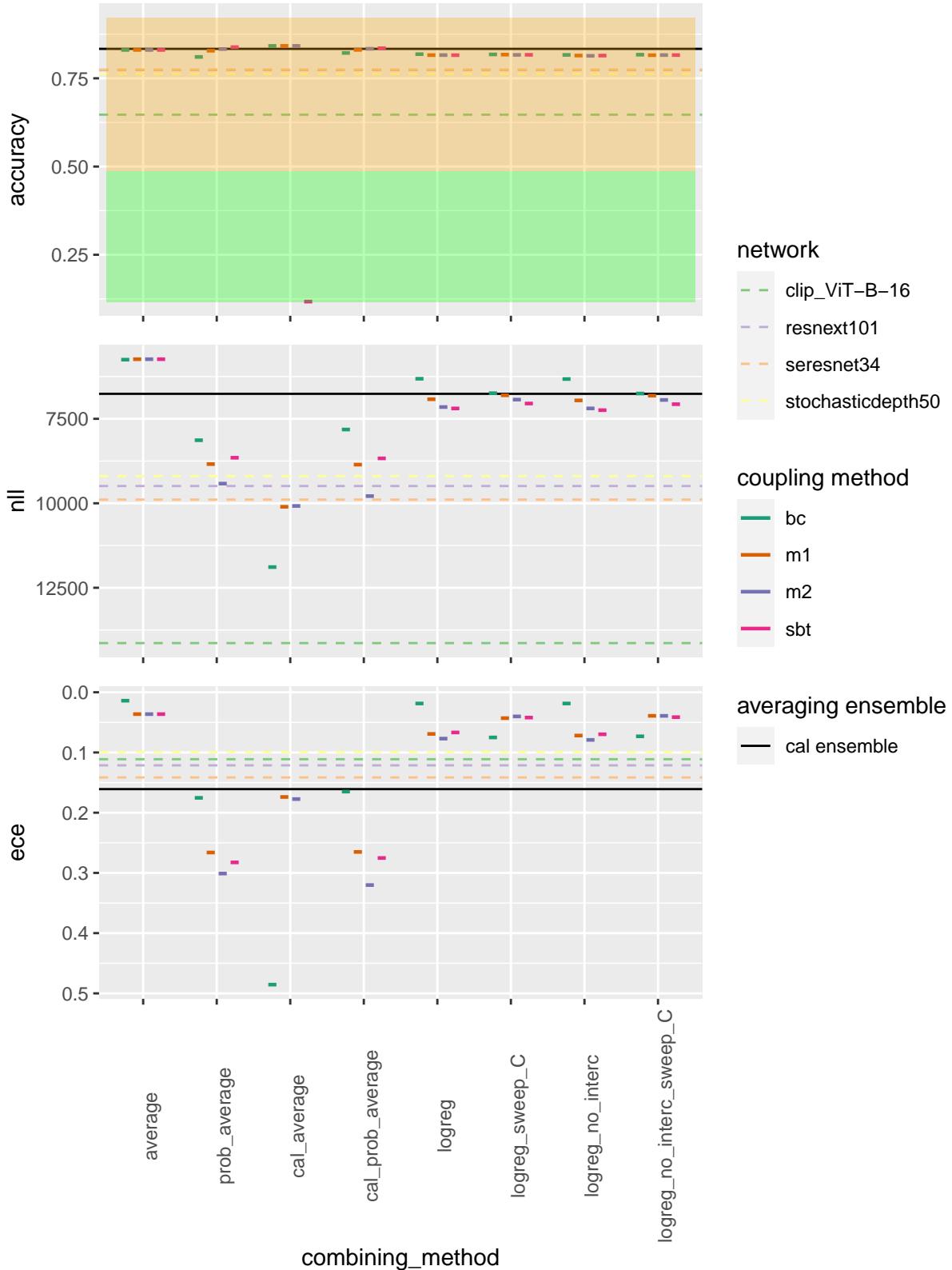
Ensemble metrics  
Error inconsistency 0.479000002145767



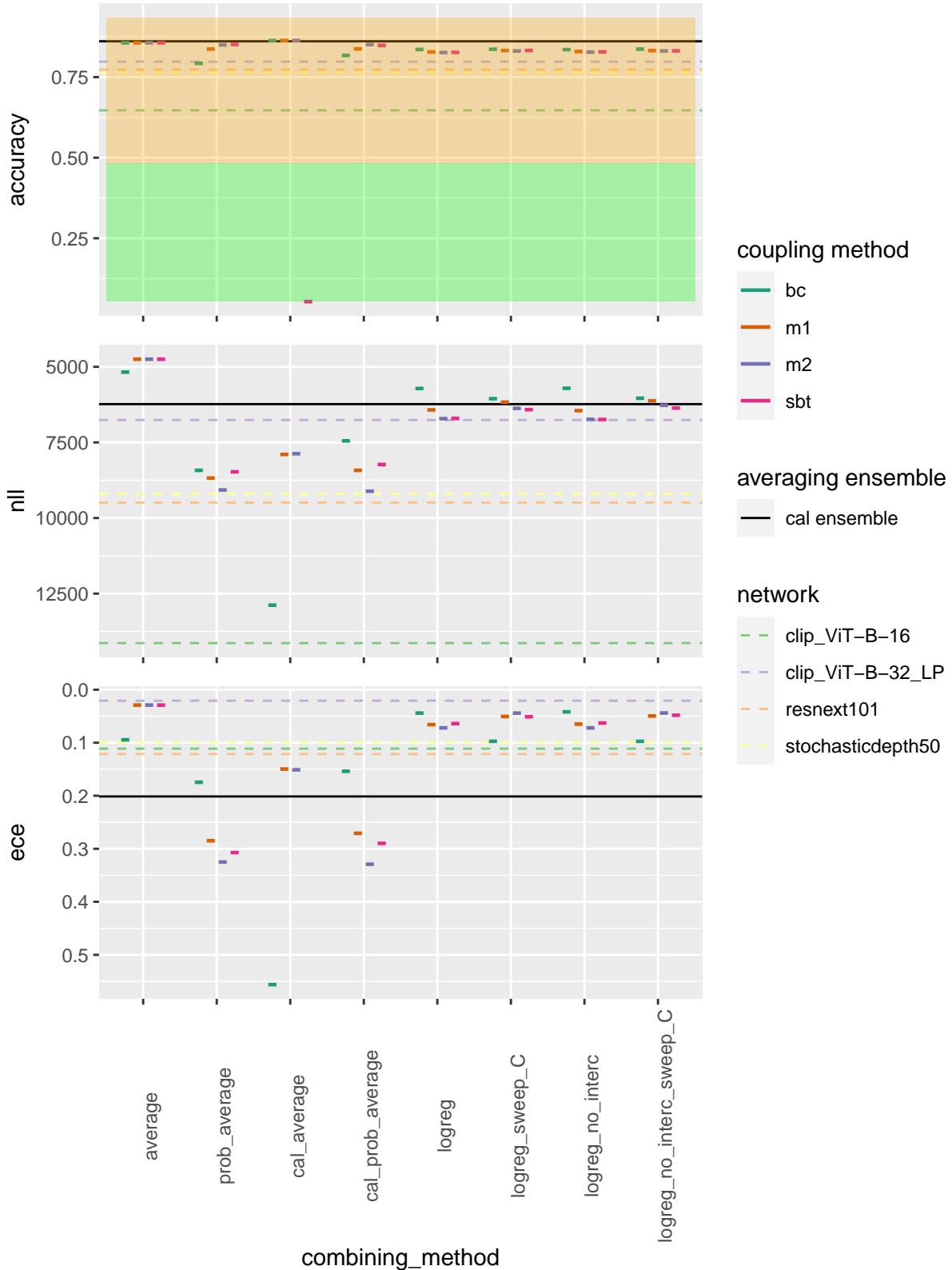
Ensemble metrics  
Error inconsistency 0.338499993085861



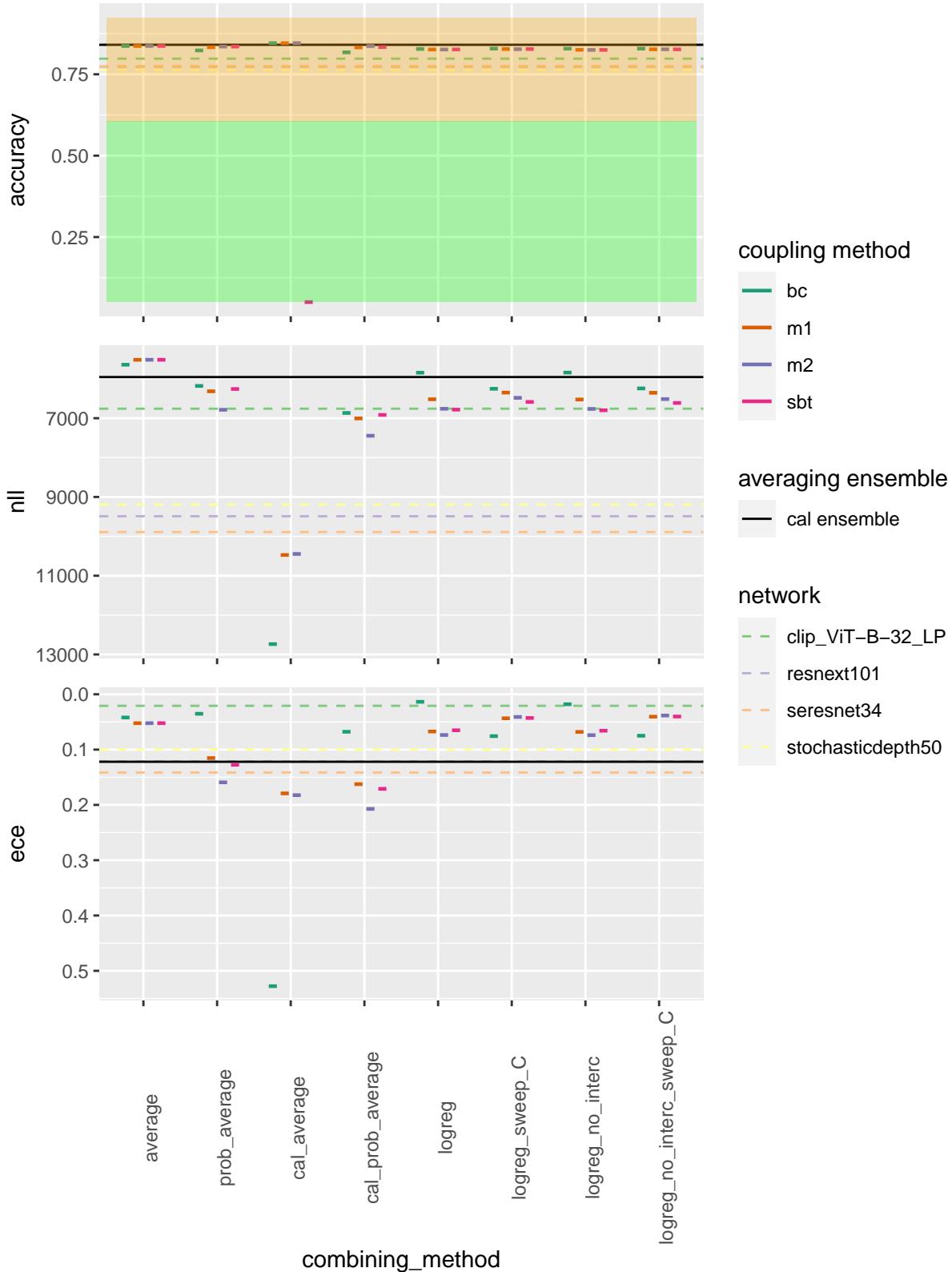
Ensemble metrics  
Error inconsistency 0.435099989175797



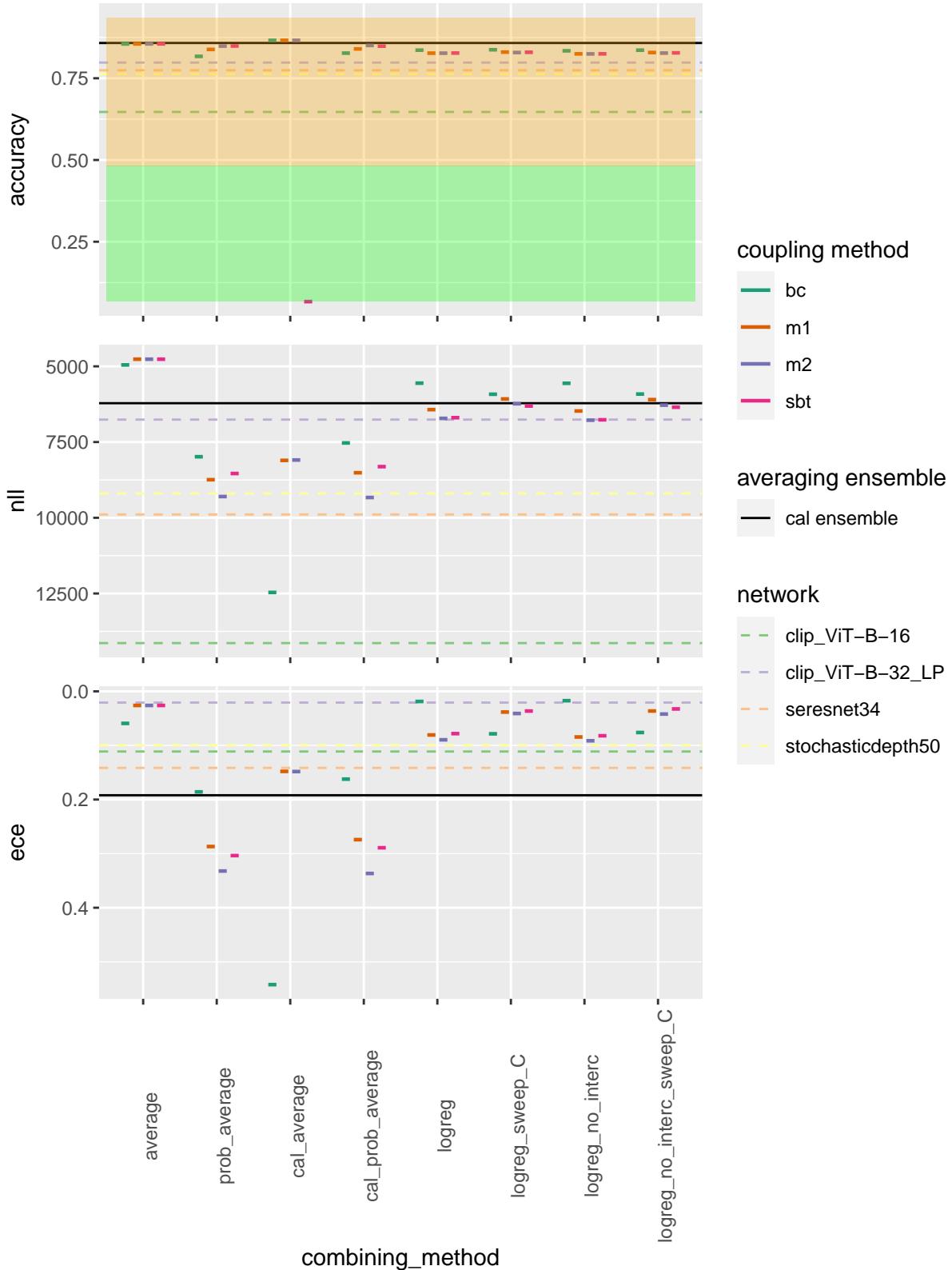
Ensemble metrics  
Error inconsistency 0.452899992465973



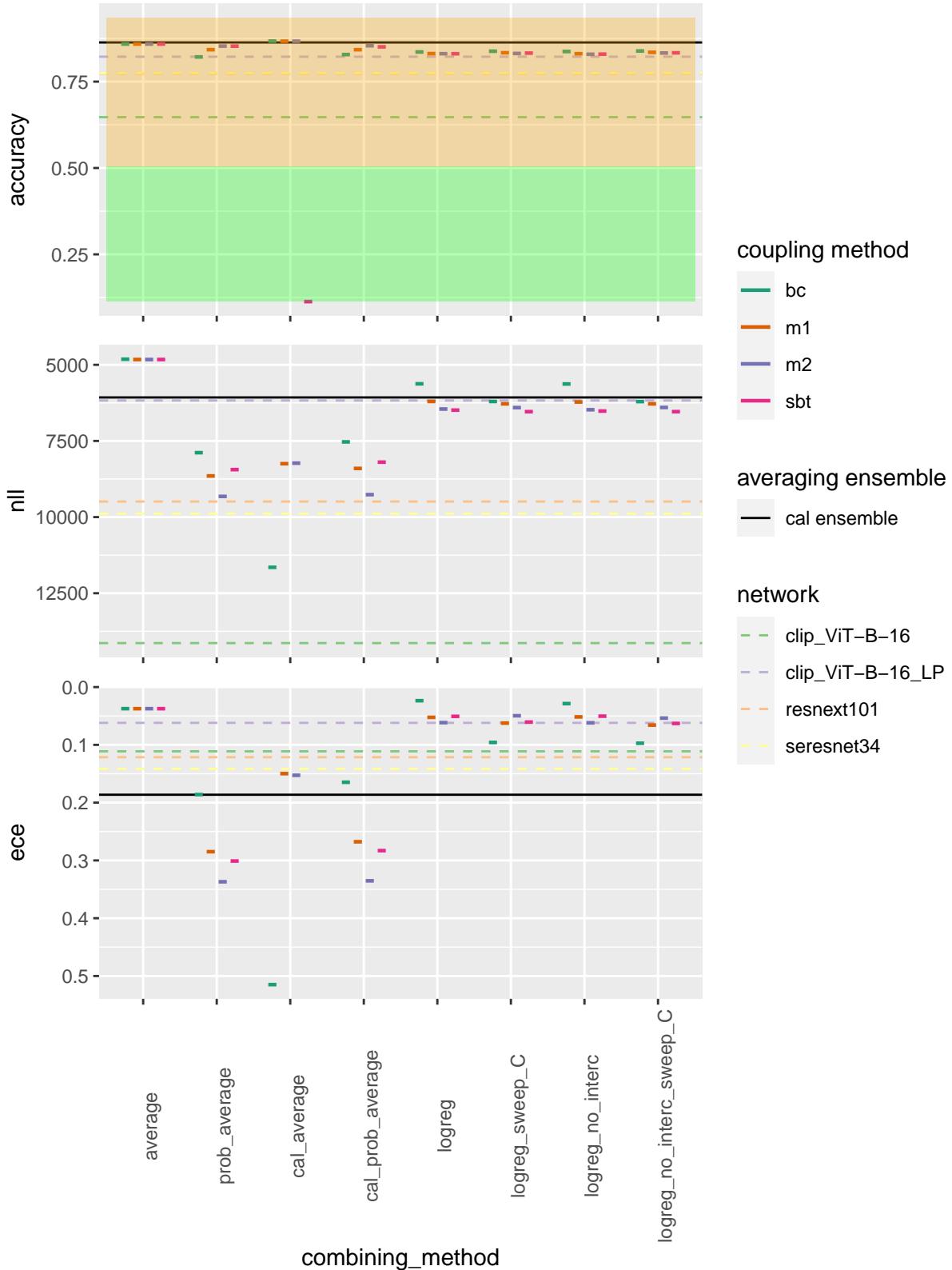
Ensemble metrics  
Error inconsistency 0.319199979305267



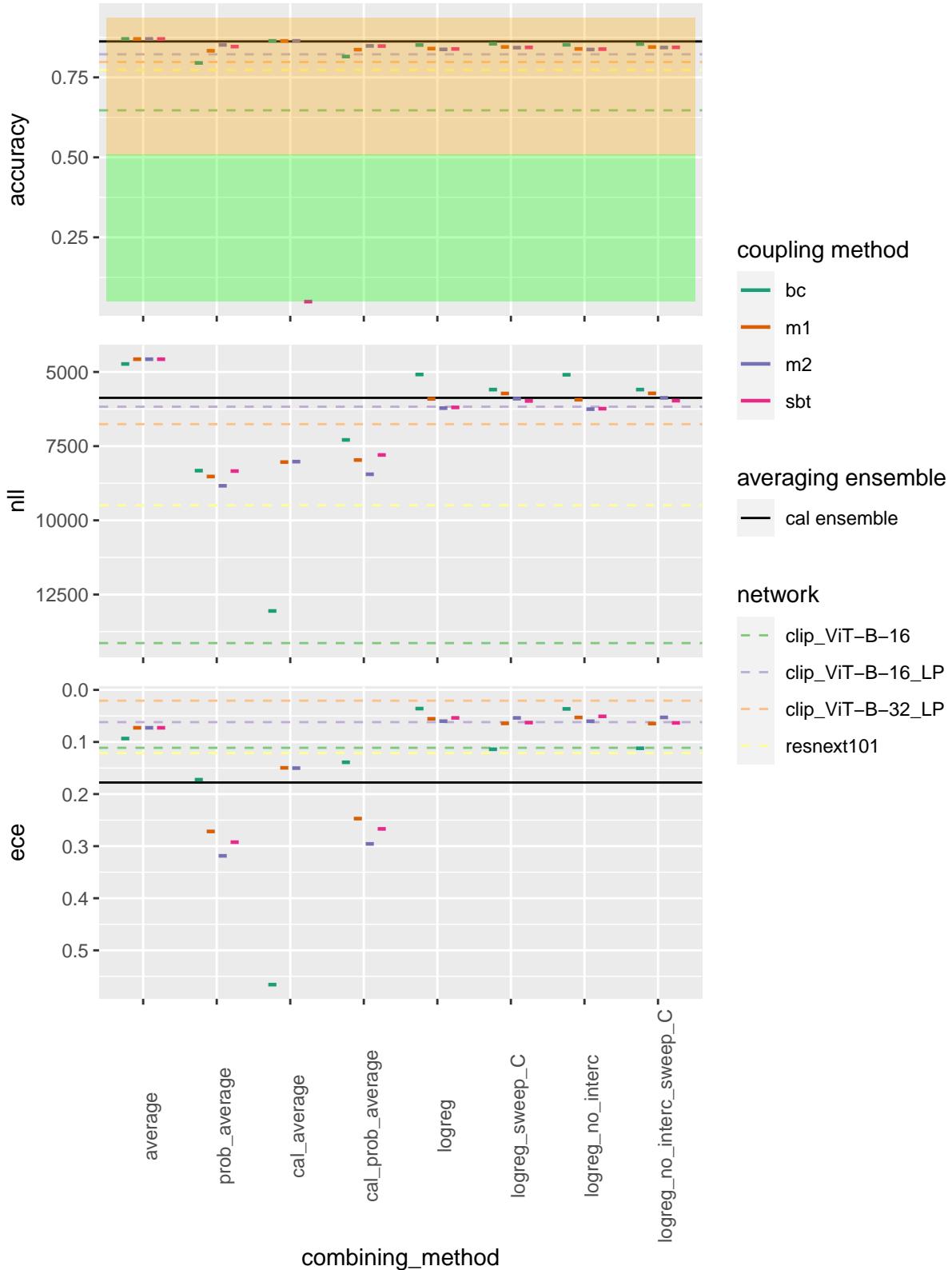
Ensemble metrics  
Error inconsistency 0.453500002622604



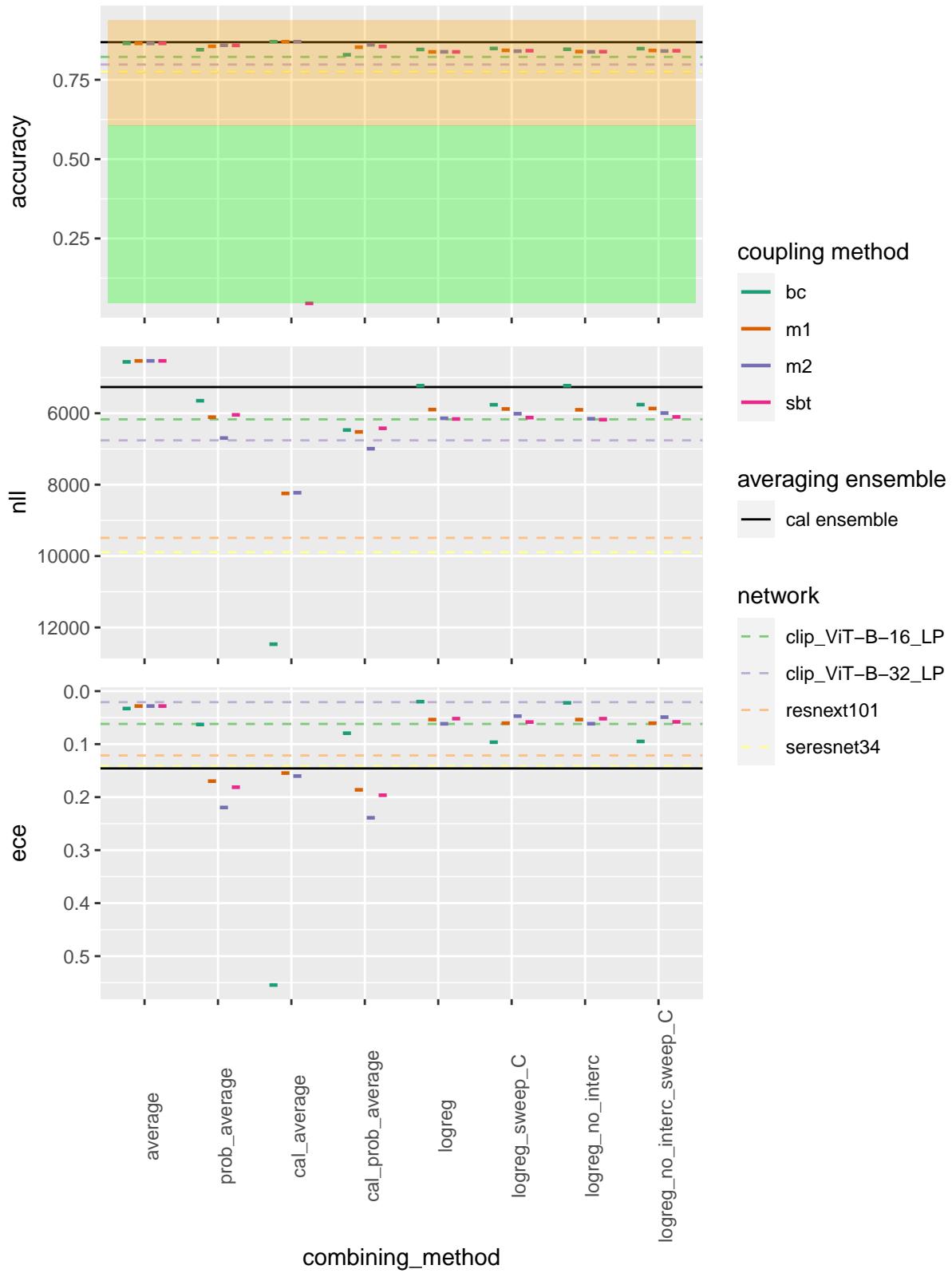
Ensemble metrics  
Error inconsistency 0.431299984455109



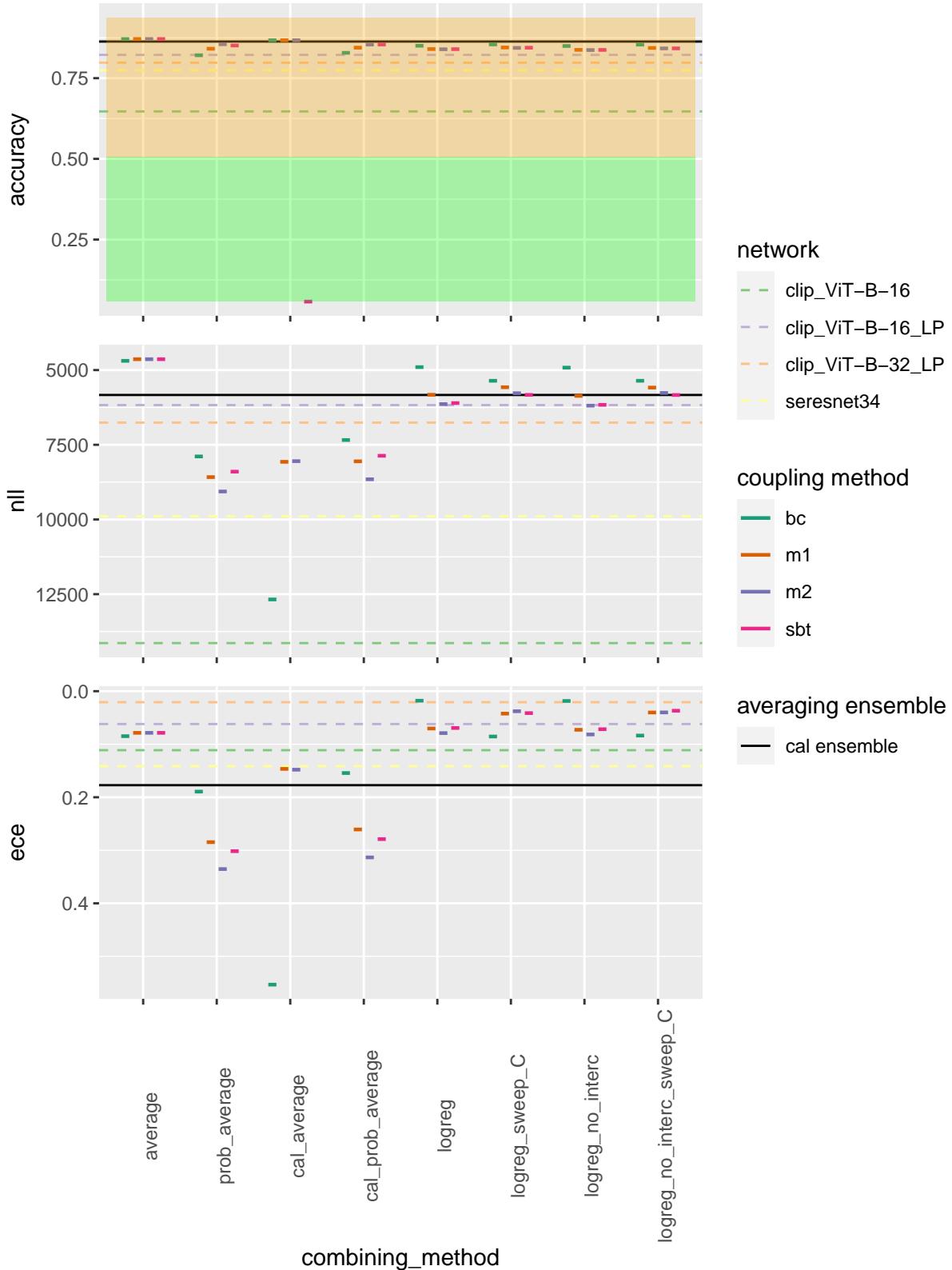
Ensemble metrics  
Error inconsistency 0.427499979734421



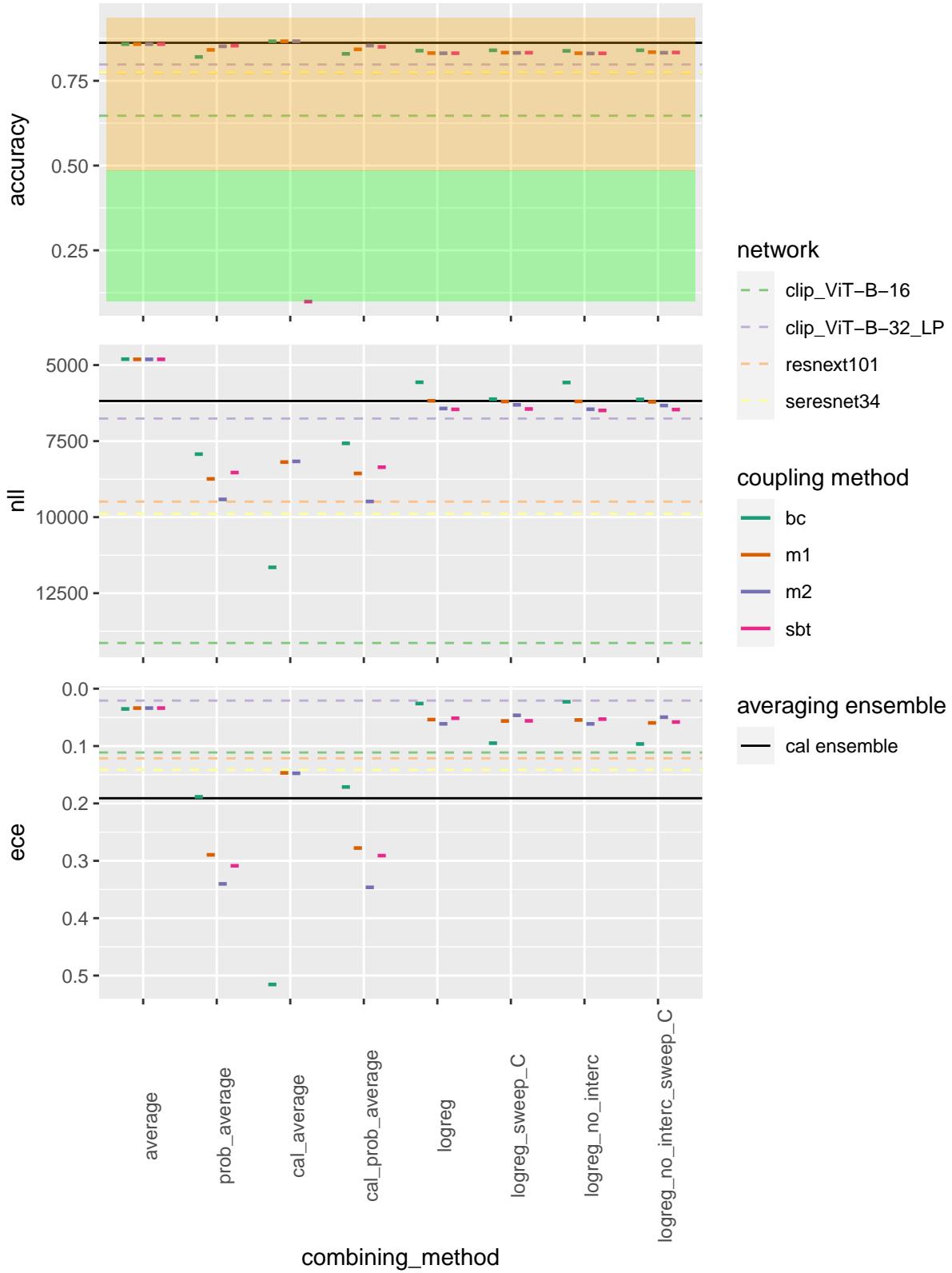
Ensemble metrics  
Error inconsistency 0.333099991083145



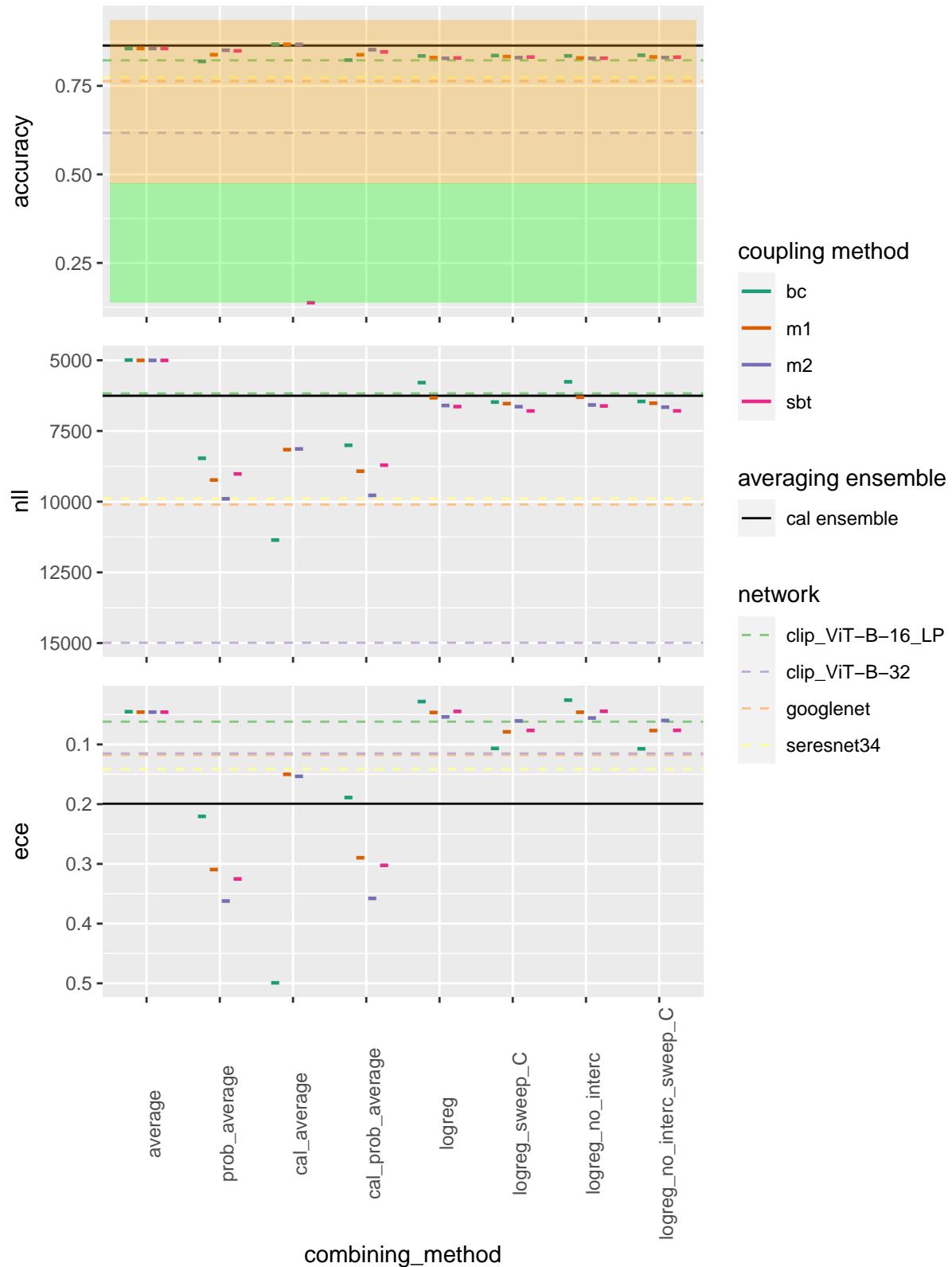
Ensemble metrics  
Error inconsistency 0.431499987840652



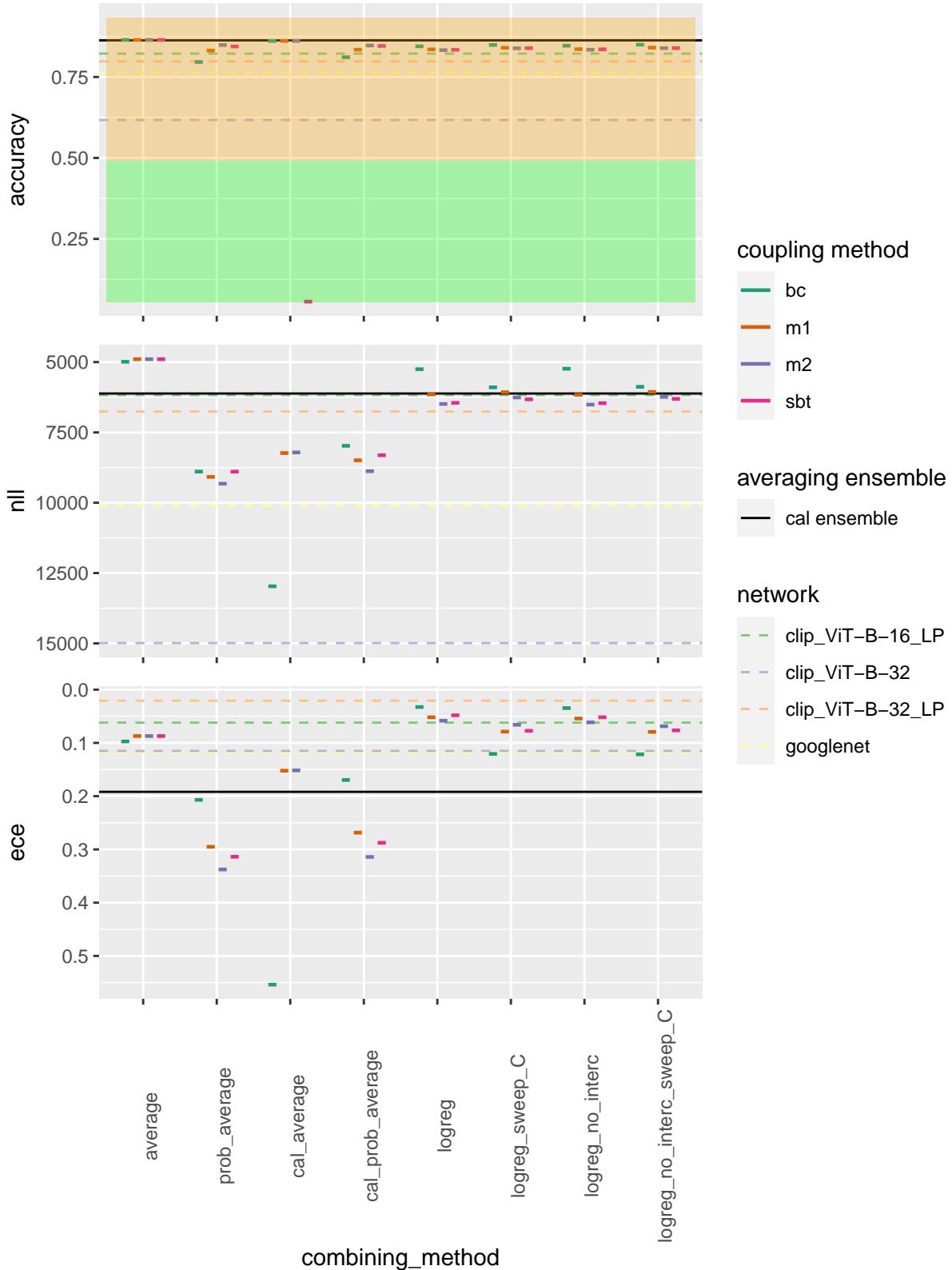
Ensemble metrics  
Error inconsistency 0.451599985361099



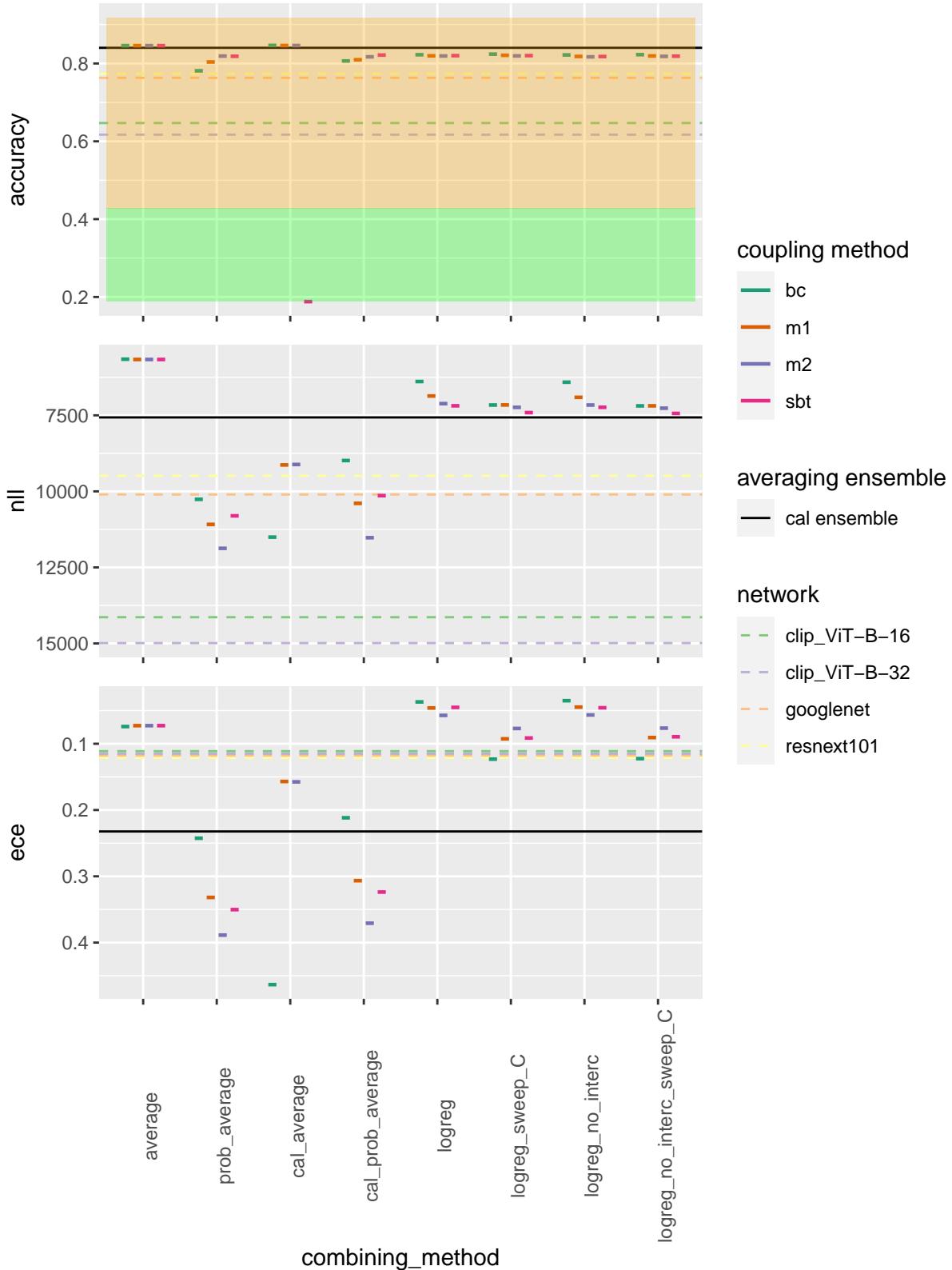
Ensemble metrics  
Error inconsistency 0.461099982261658



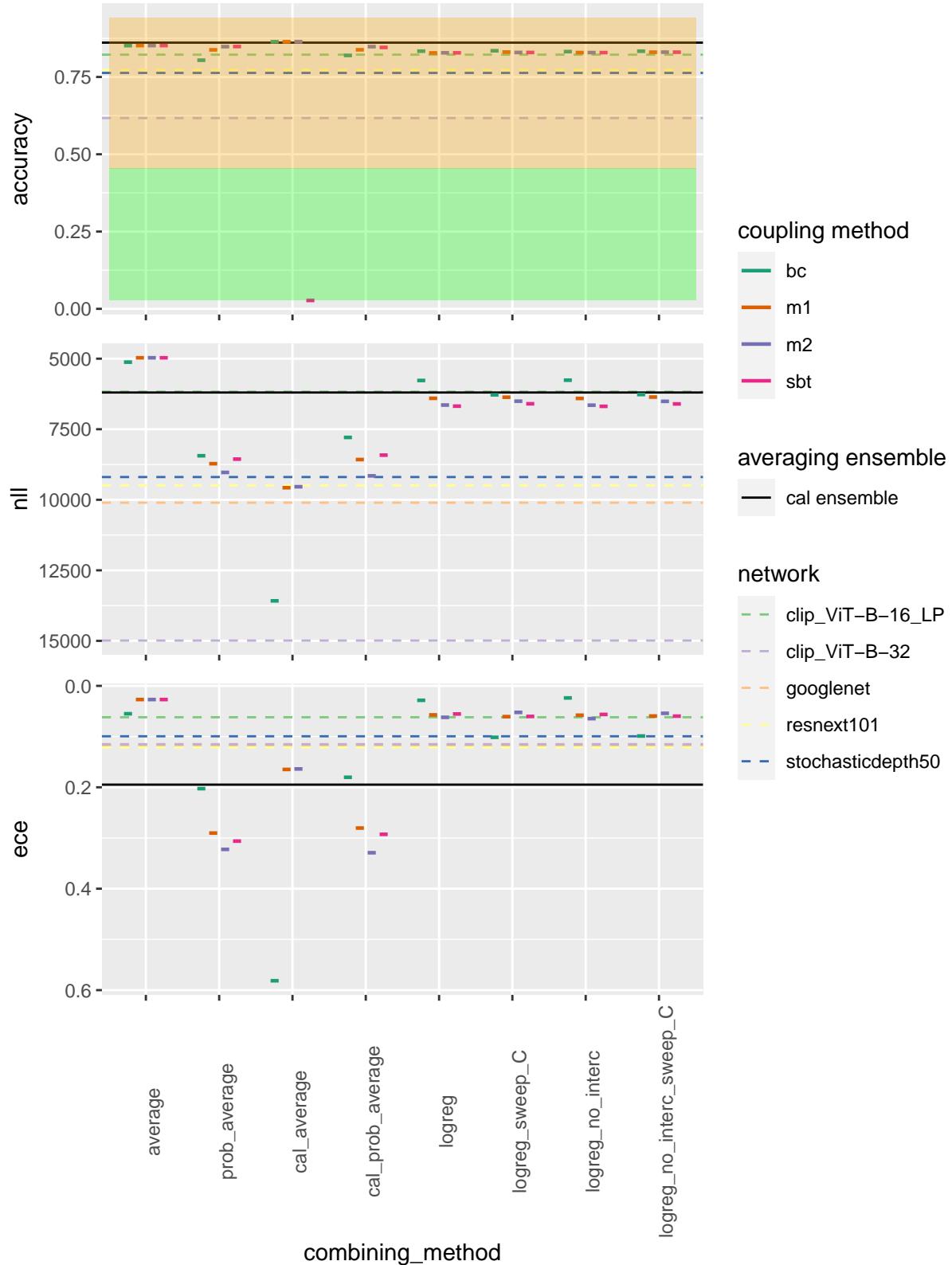
Ensemble metrics  
Error inconsistency 0.441399991512299



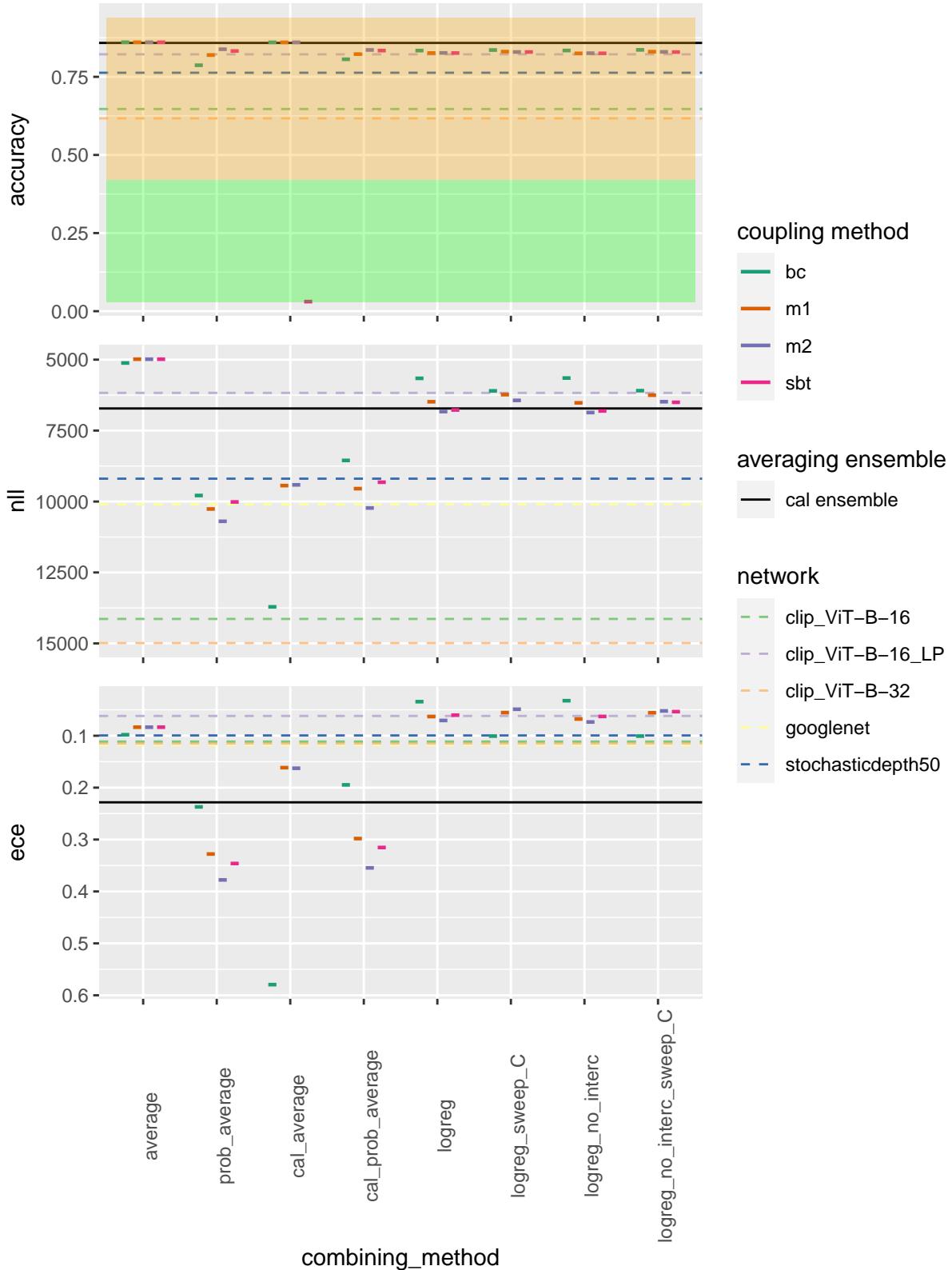
Ensemble metrics  
Error inconsistency 0.489199995994568



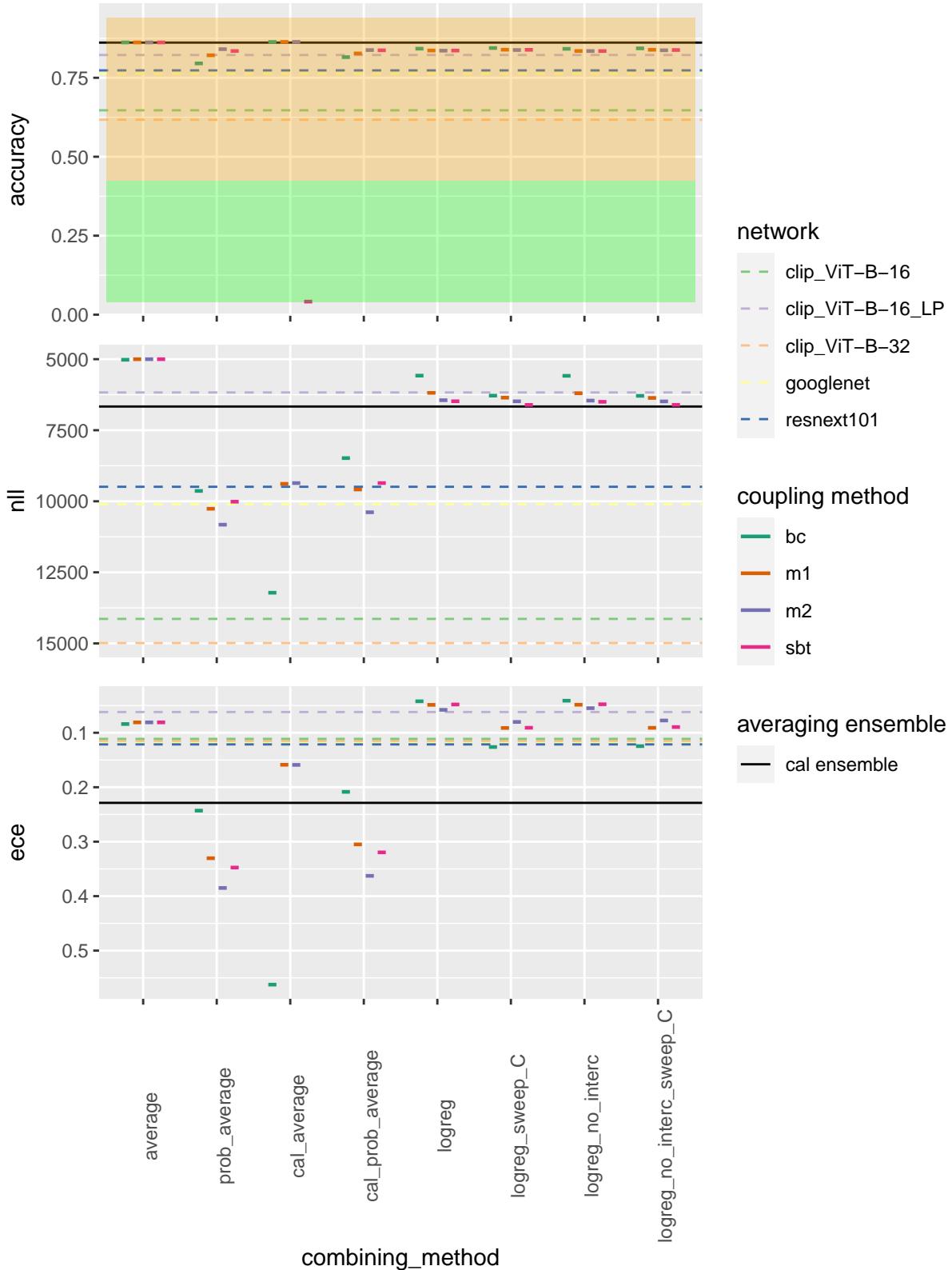
Ensemble metrics  
Error inconsistency 0.488099992275238



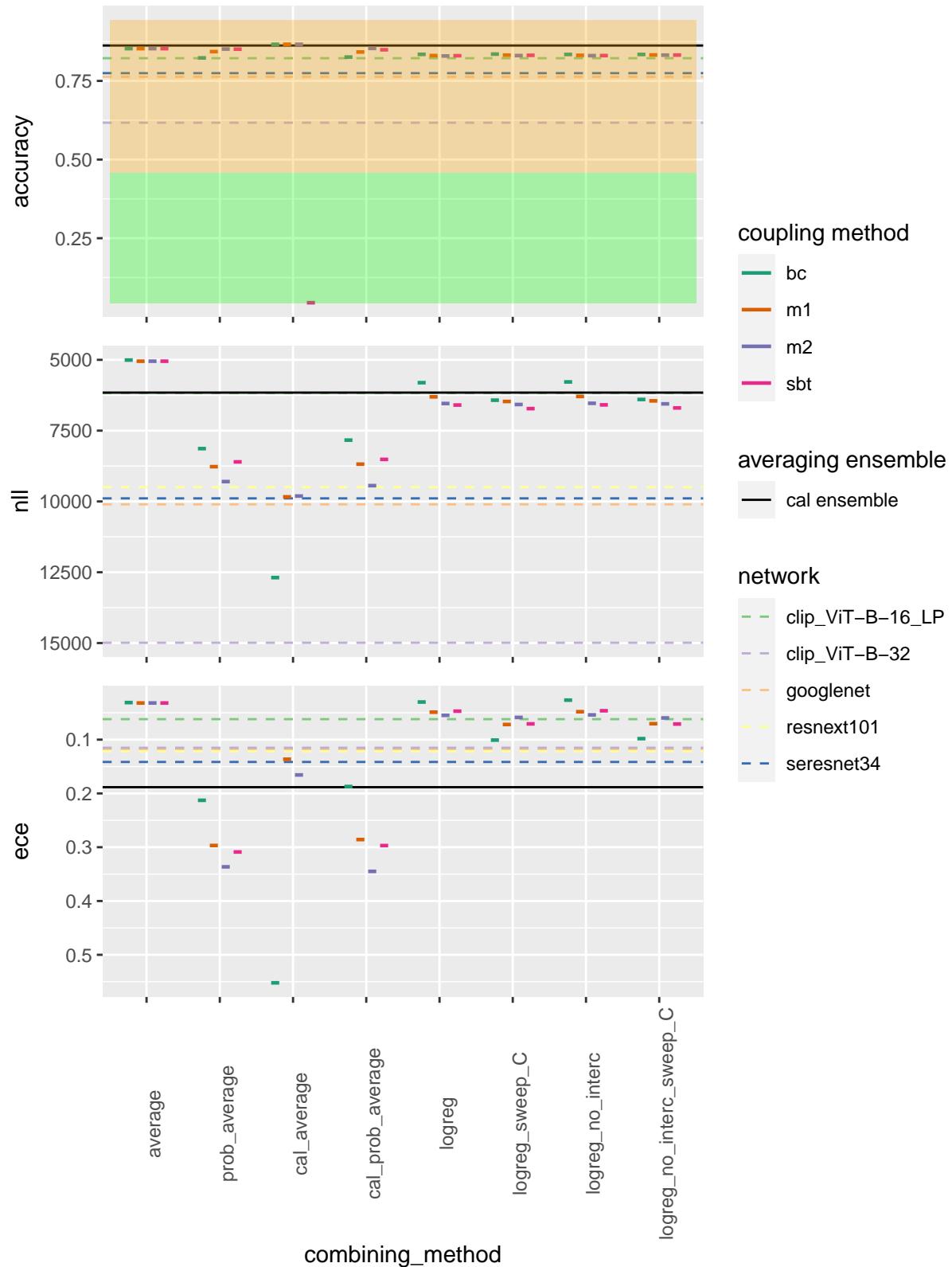
Ensemble metrics  
Error inconsistency 0.519400000572205



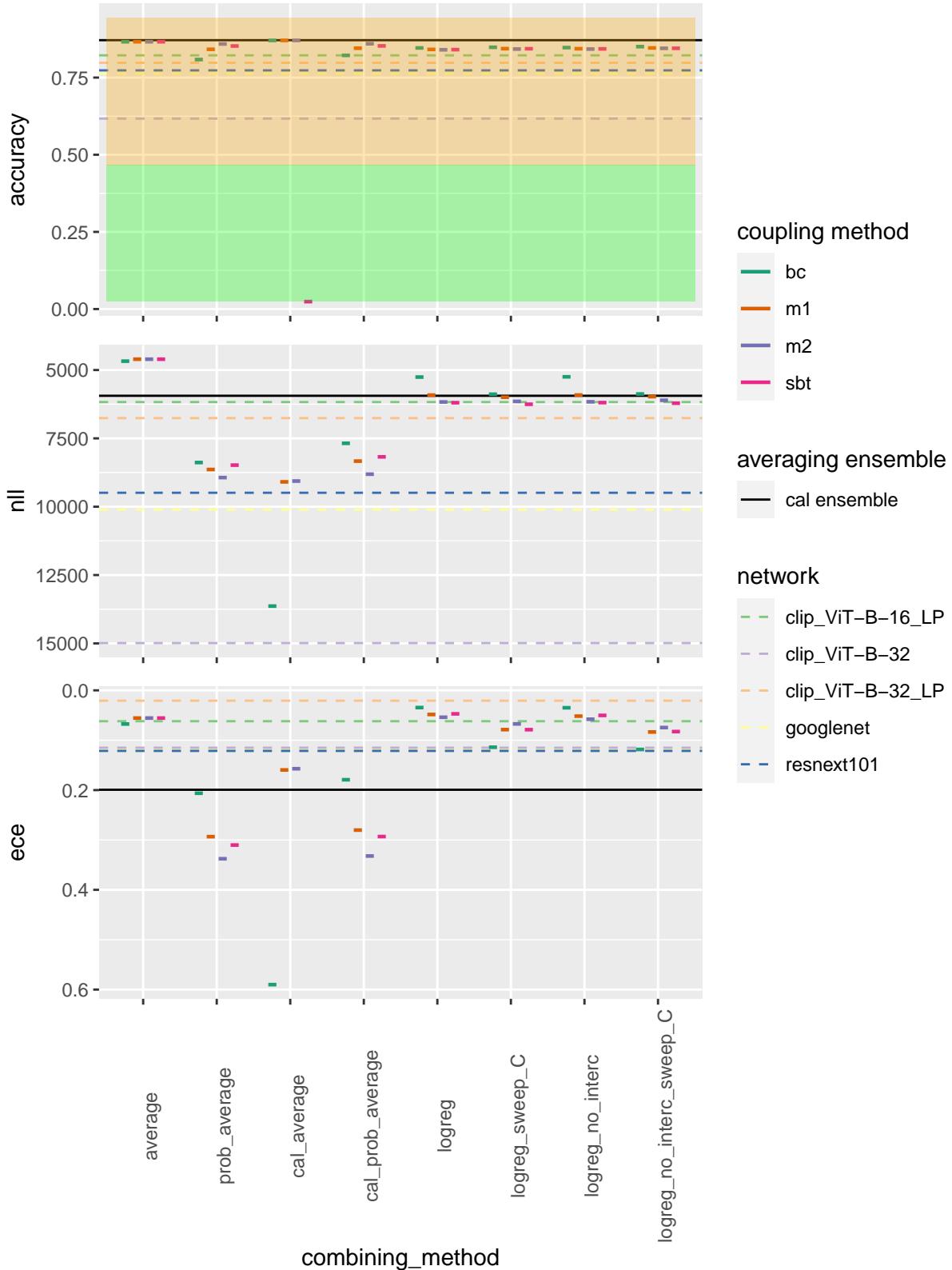
Ensemble metrics  
Error inconsistency 0.515999972820282



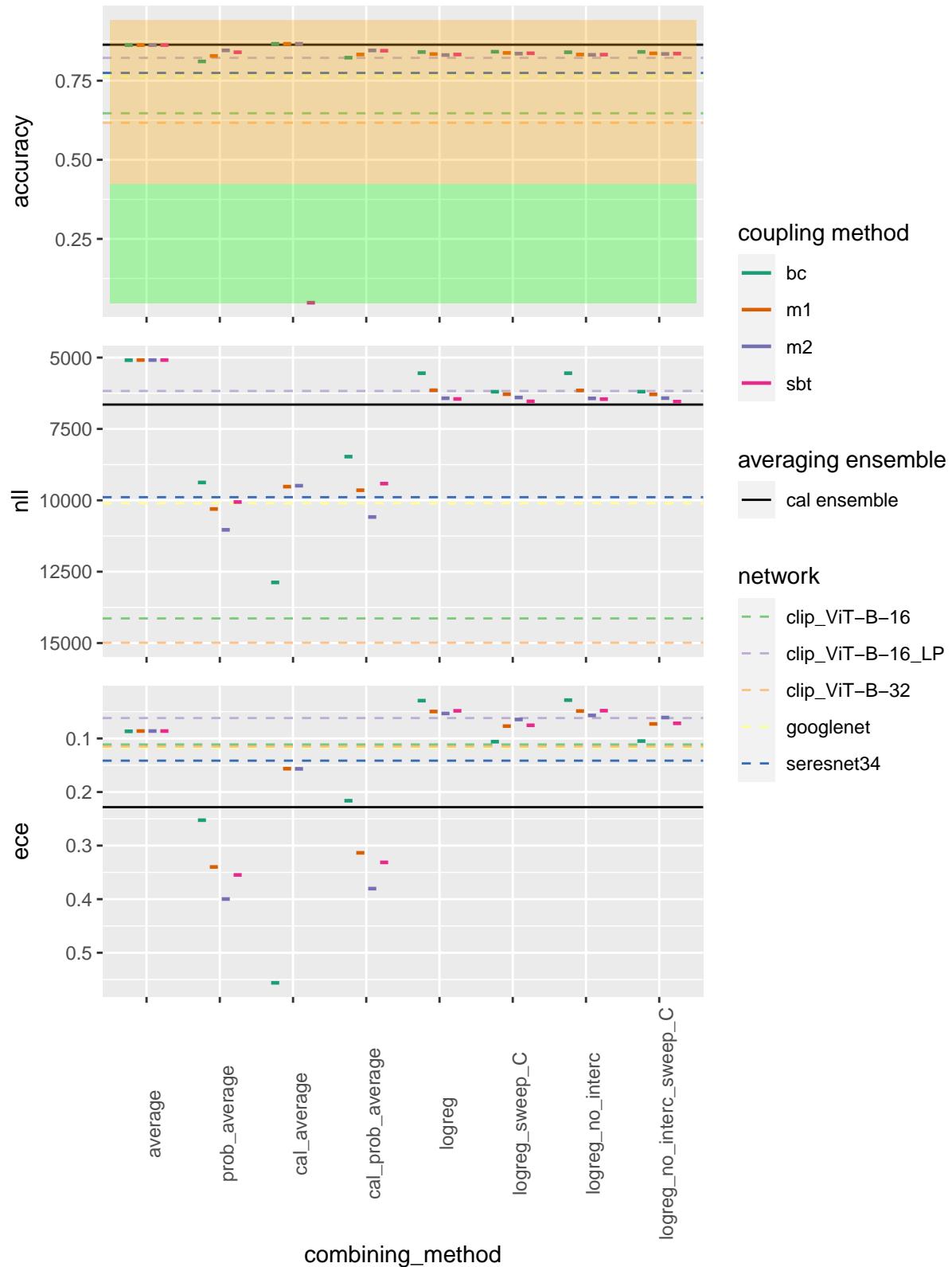
Ensemble metrics  
Error inconsistency 0.485699981451034



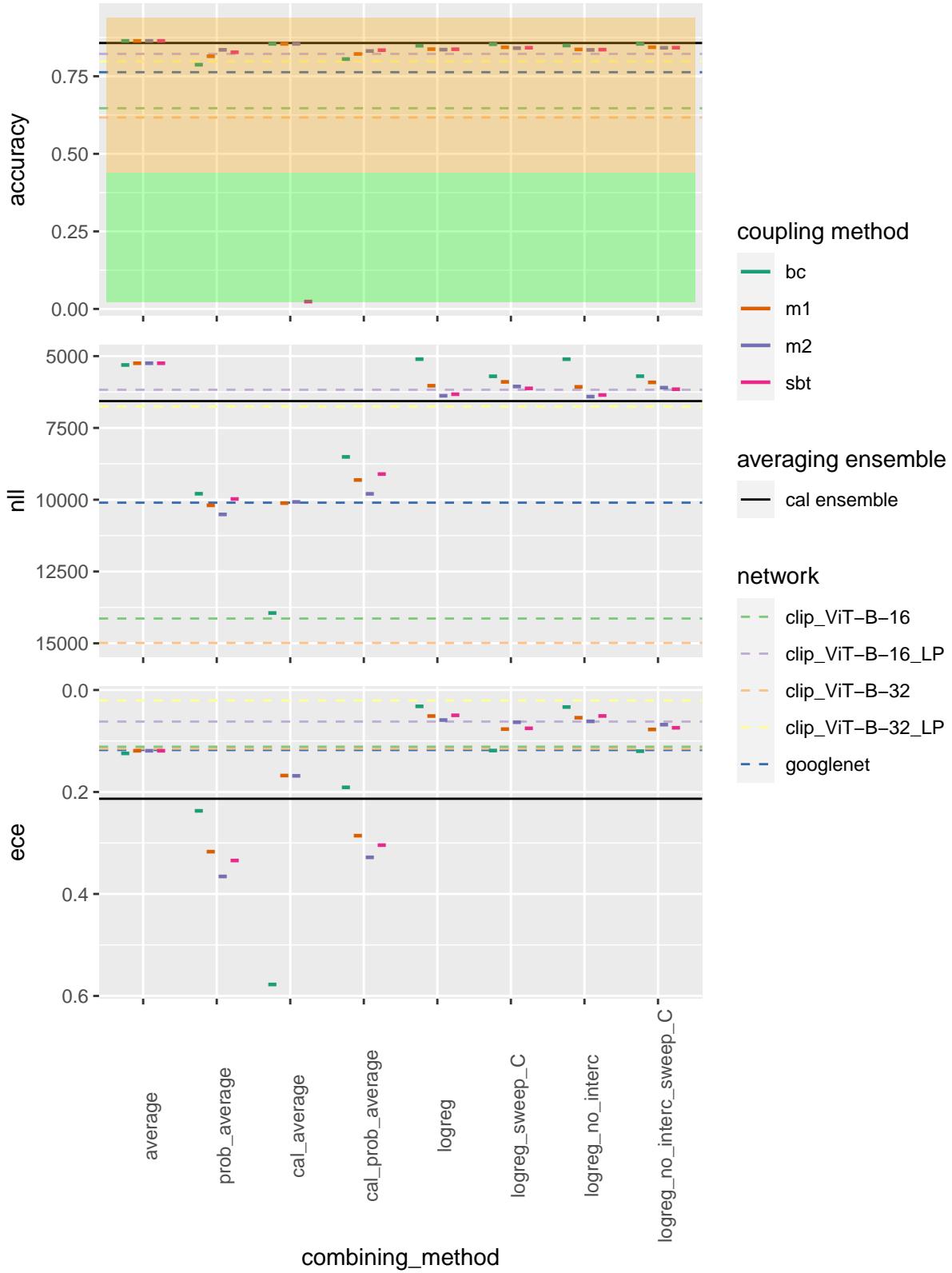
Ensemble metrics  
Error inconsistency 0.477399975061417



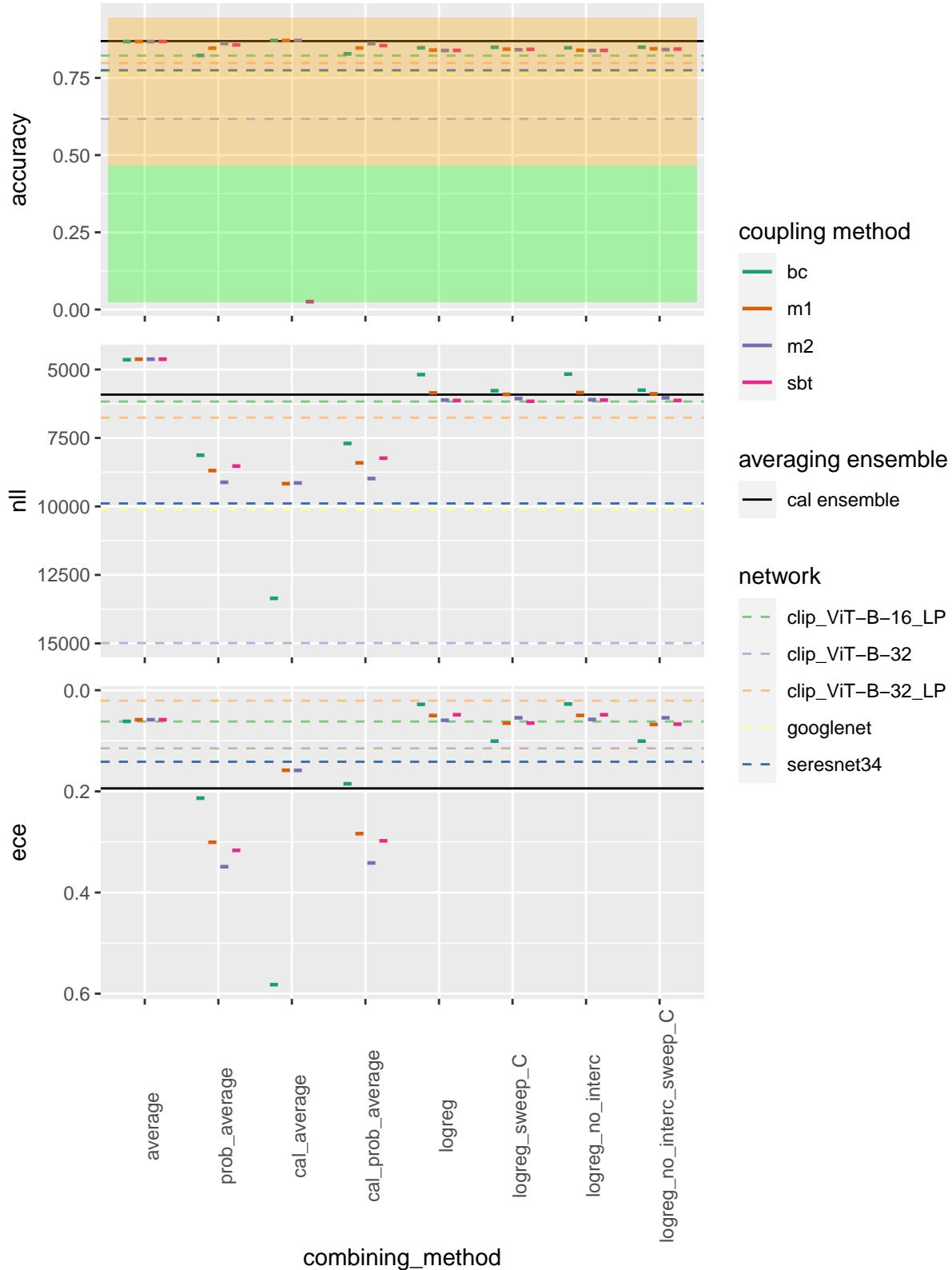
Ensemble metrics  
Error inconsistency 0.518199980258942



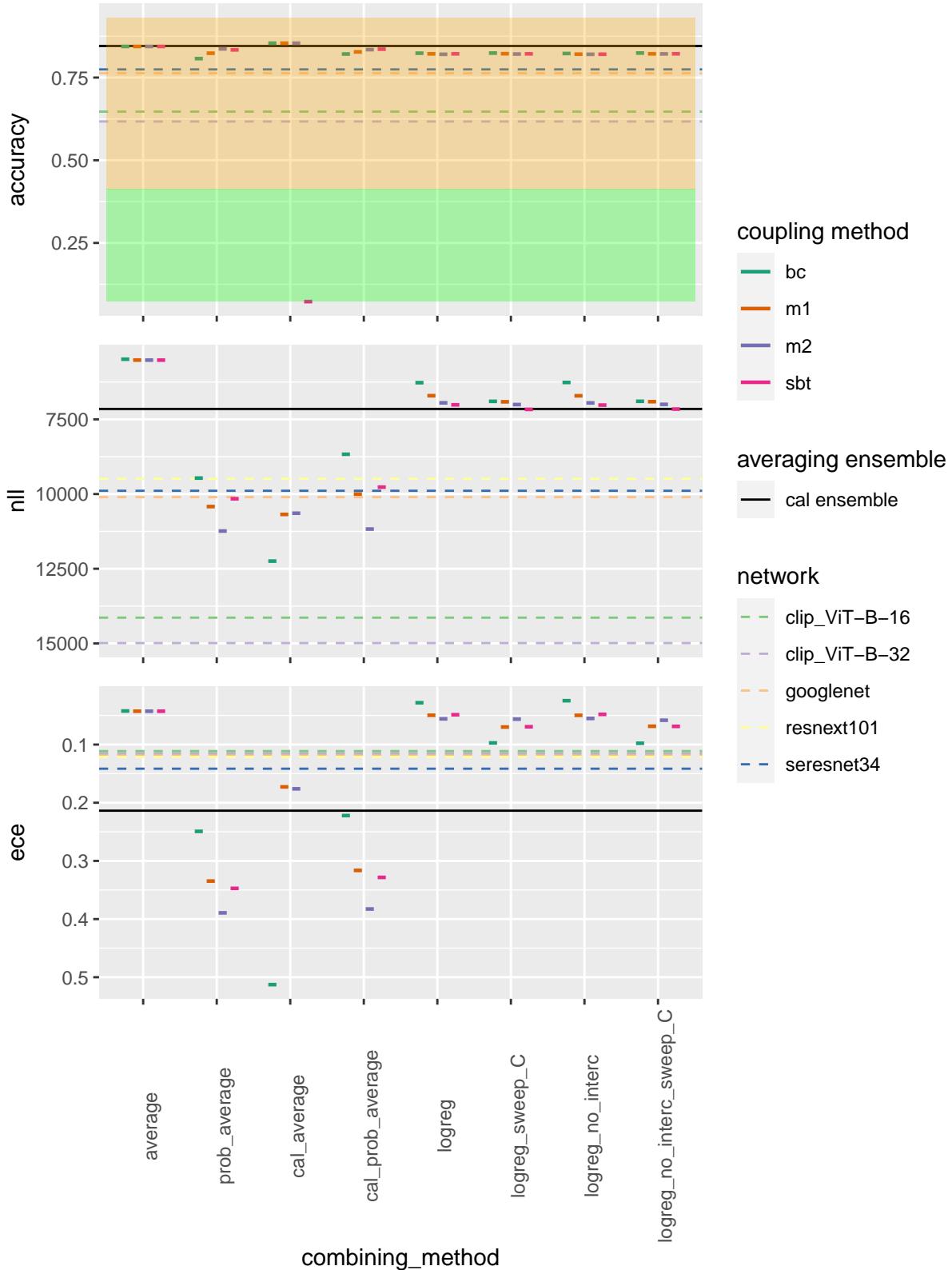
Ensemble metrics  
Error inconsistency 0.5



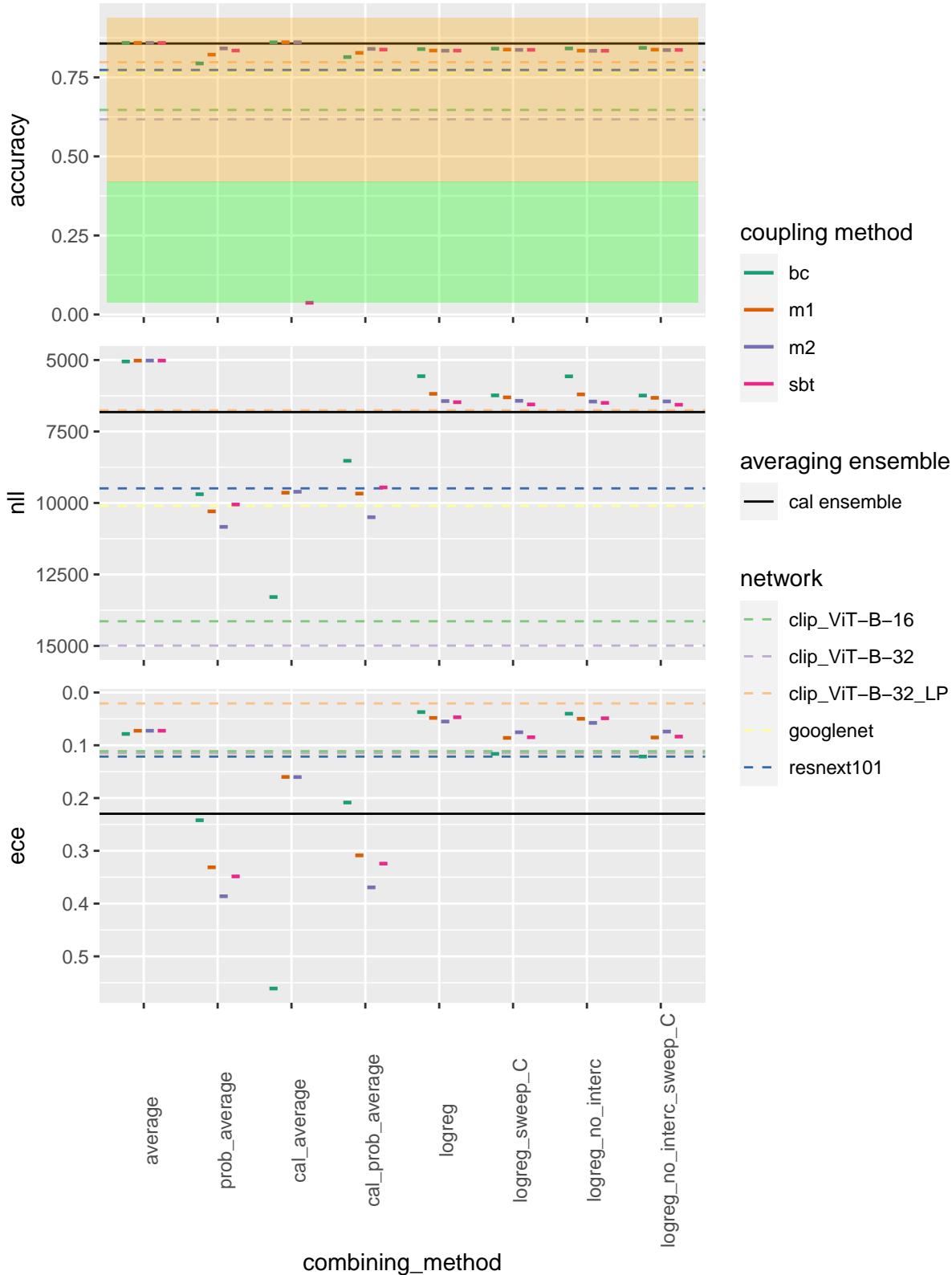
Ensemble metrics  
Error inconsistency 0.47859999537468



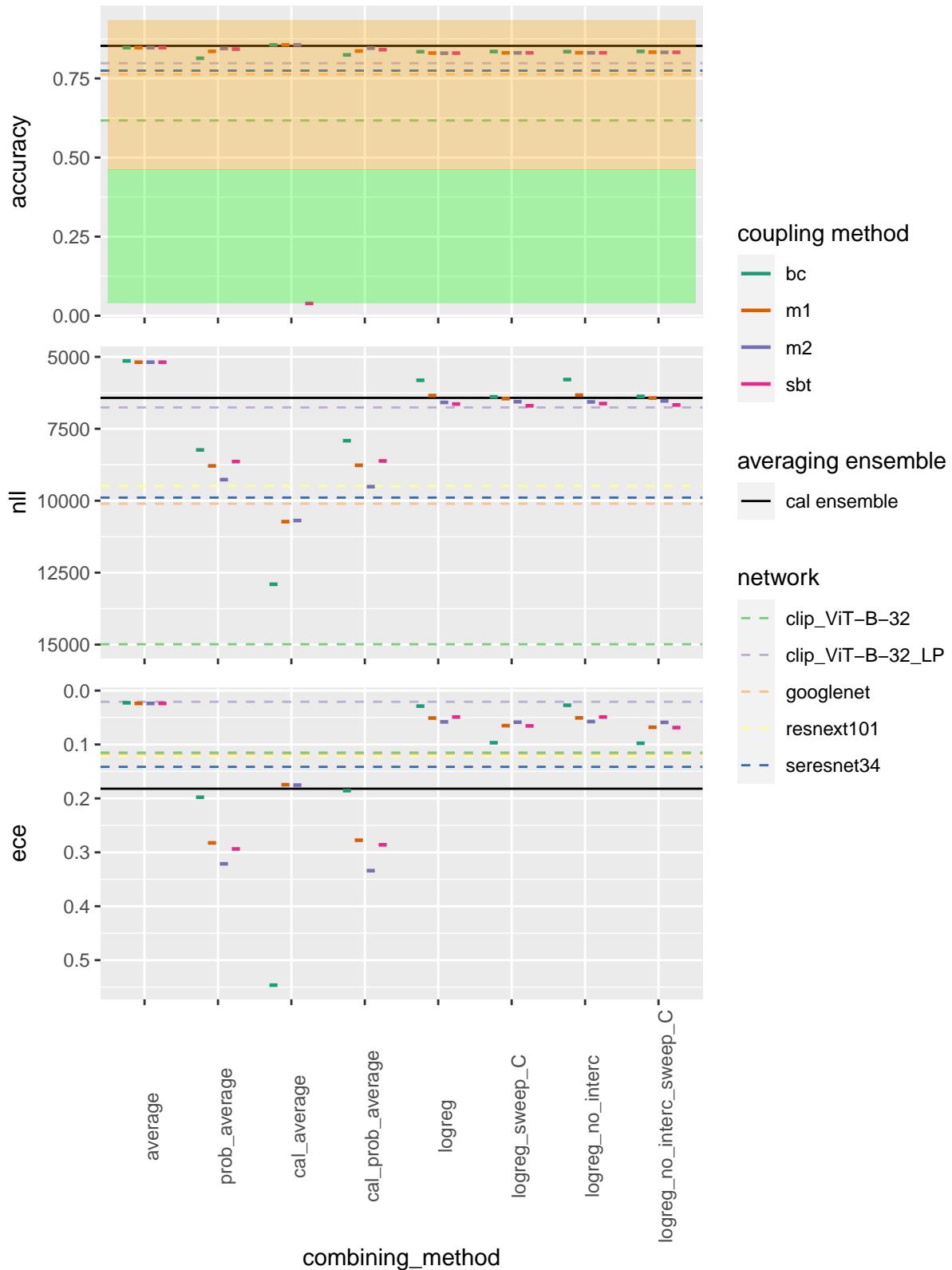
Ensemble metrics  
Error inconsistency 0.517799973487854



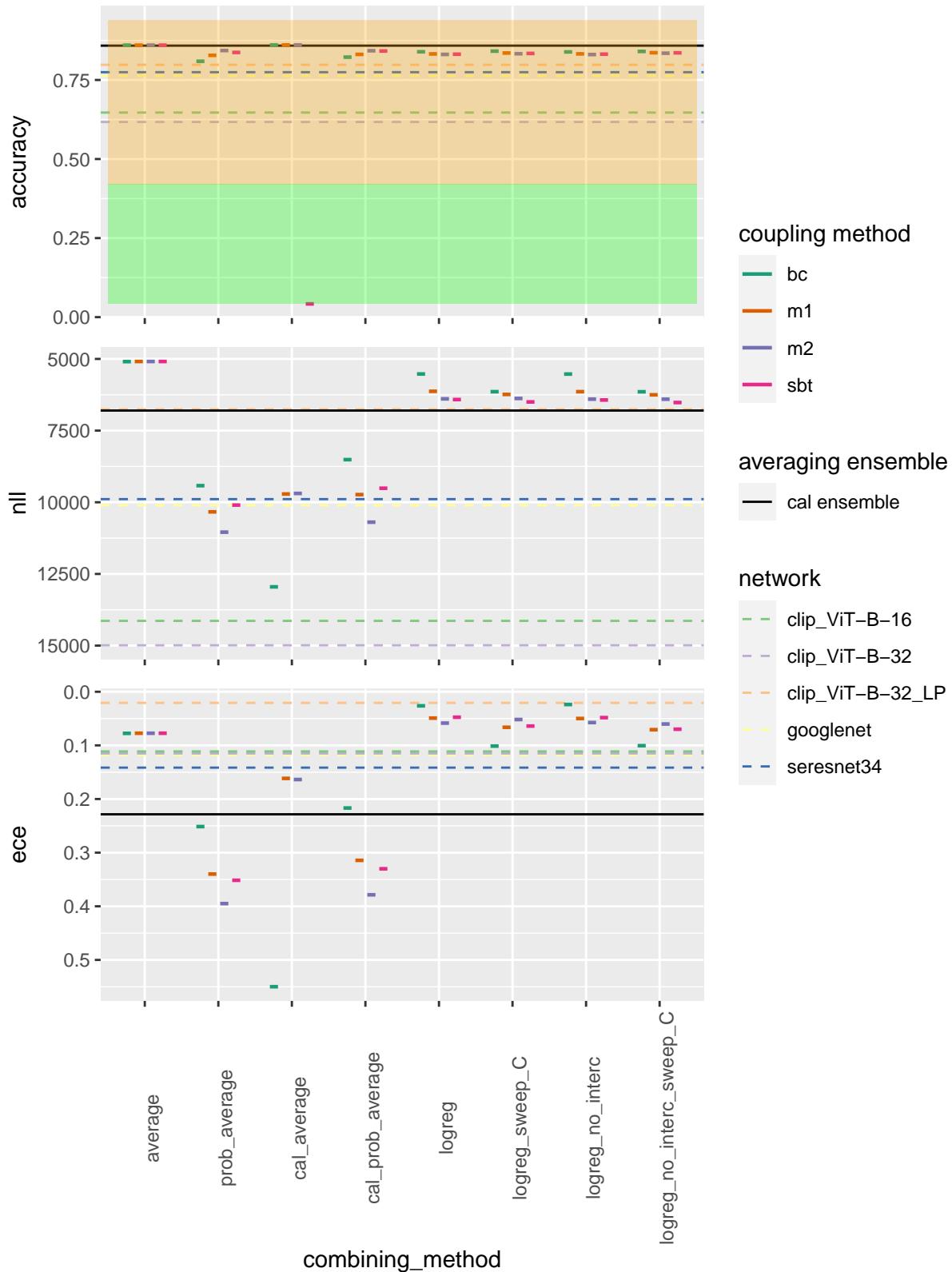
Ensemble metrics  
Error inconsistency 0.518400013446808



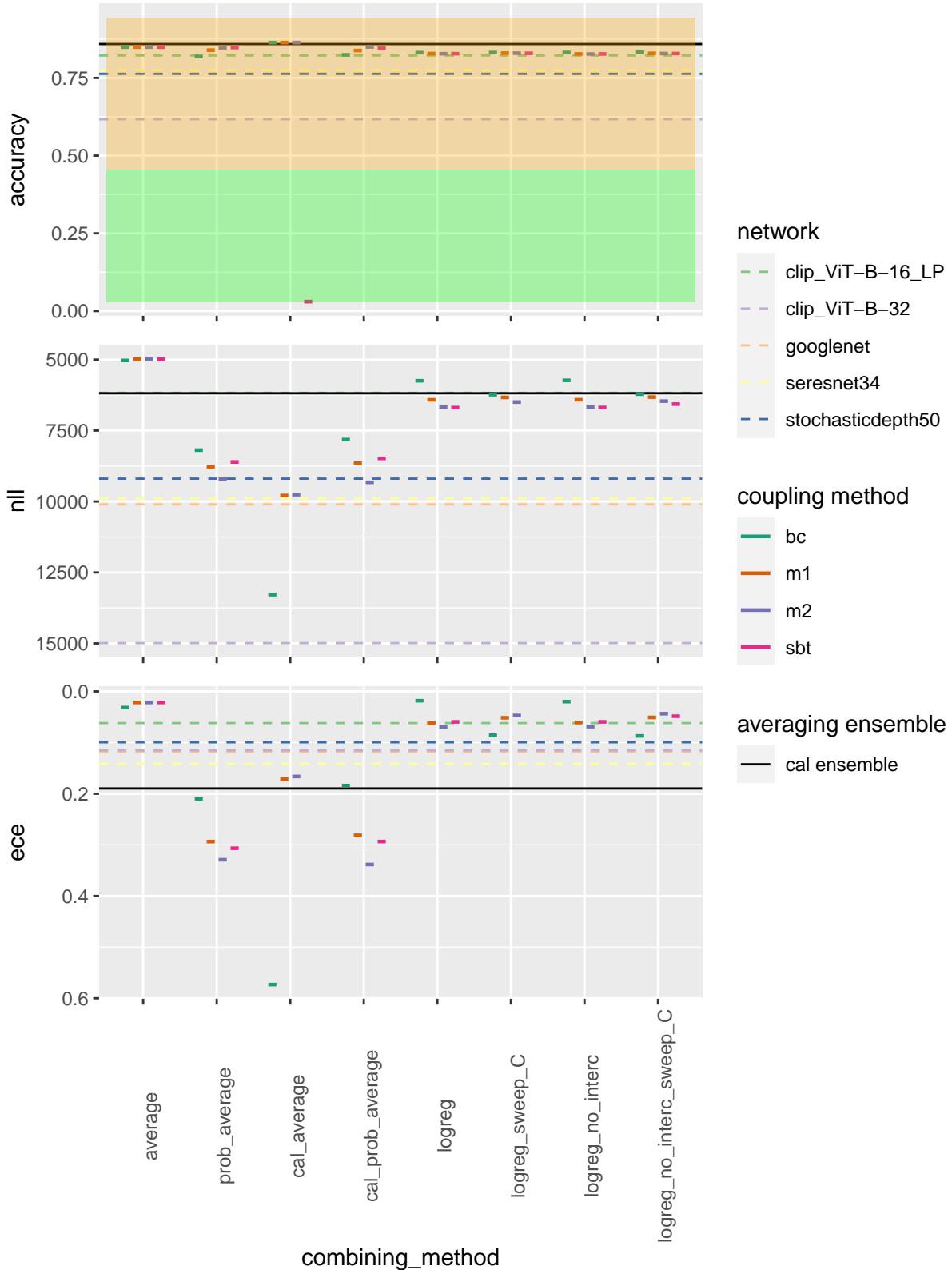
Ensemble metrics  
Error inconsistency 0.472699999809265



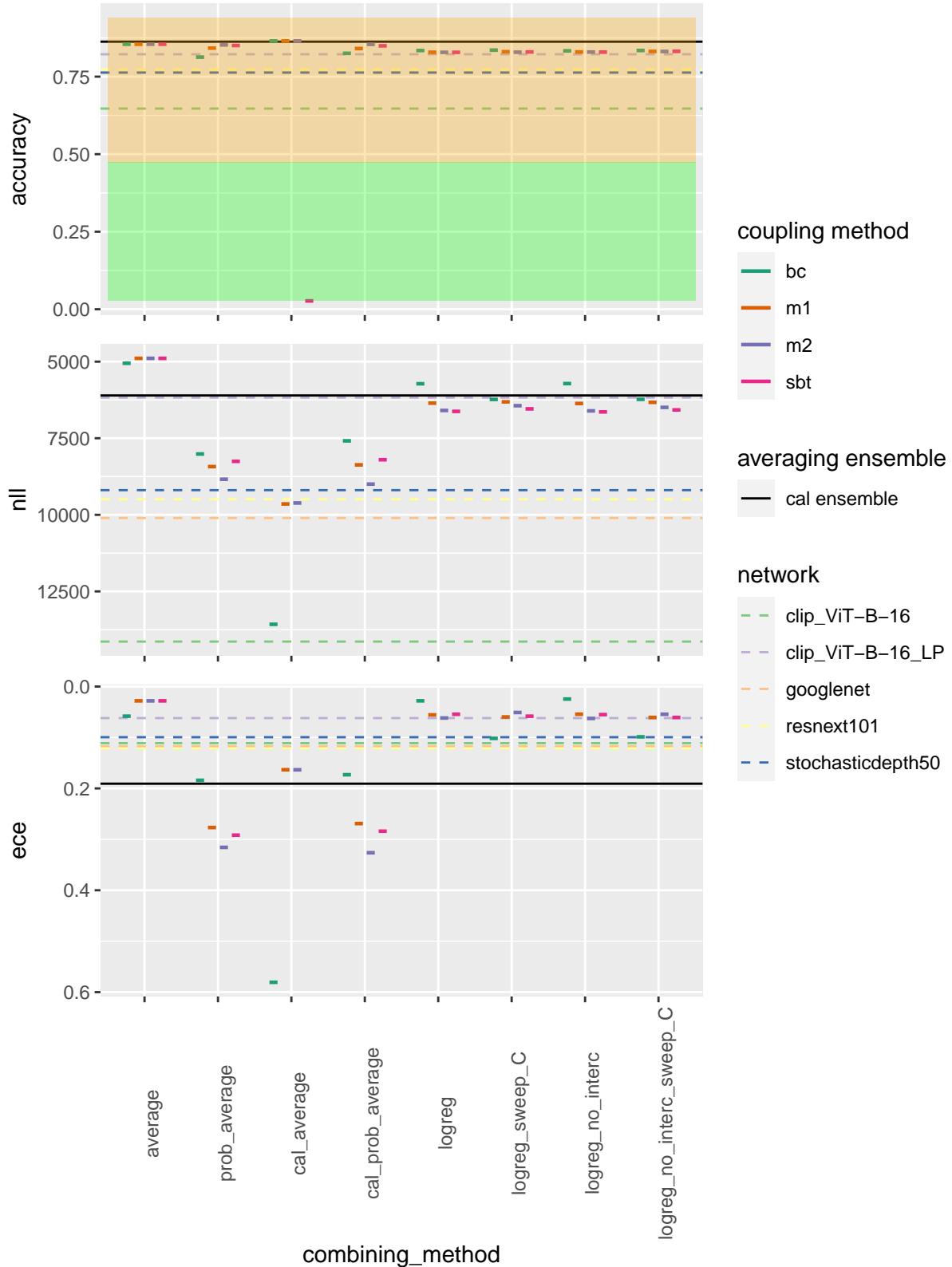
Ensemble metrics  
Error inconsistency 0.519199967384338



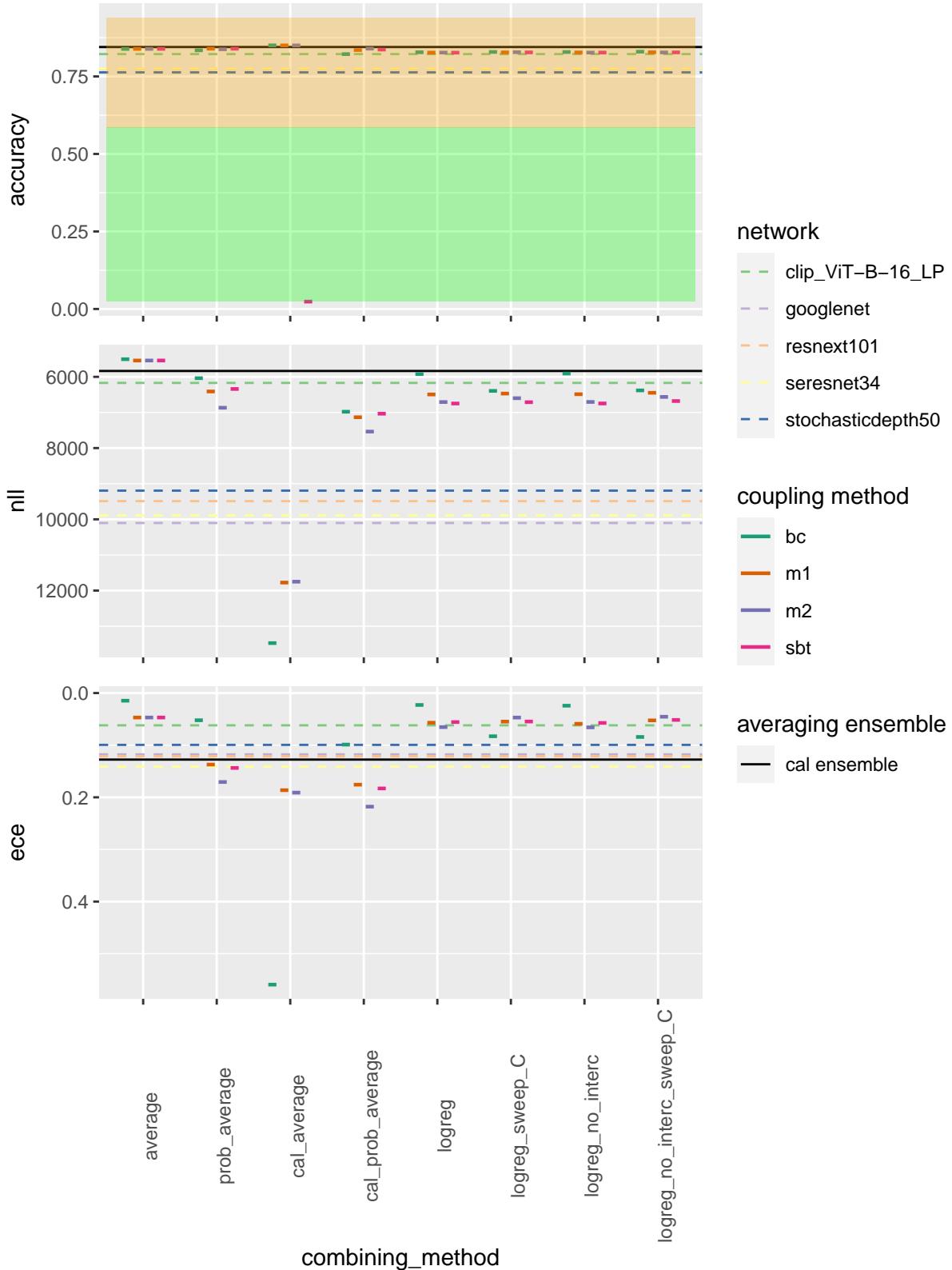
Ensemble metrics  
Error inconsistency 0.48870000243187



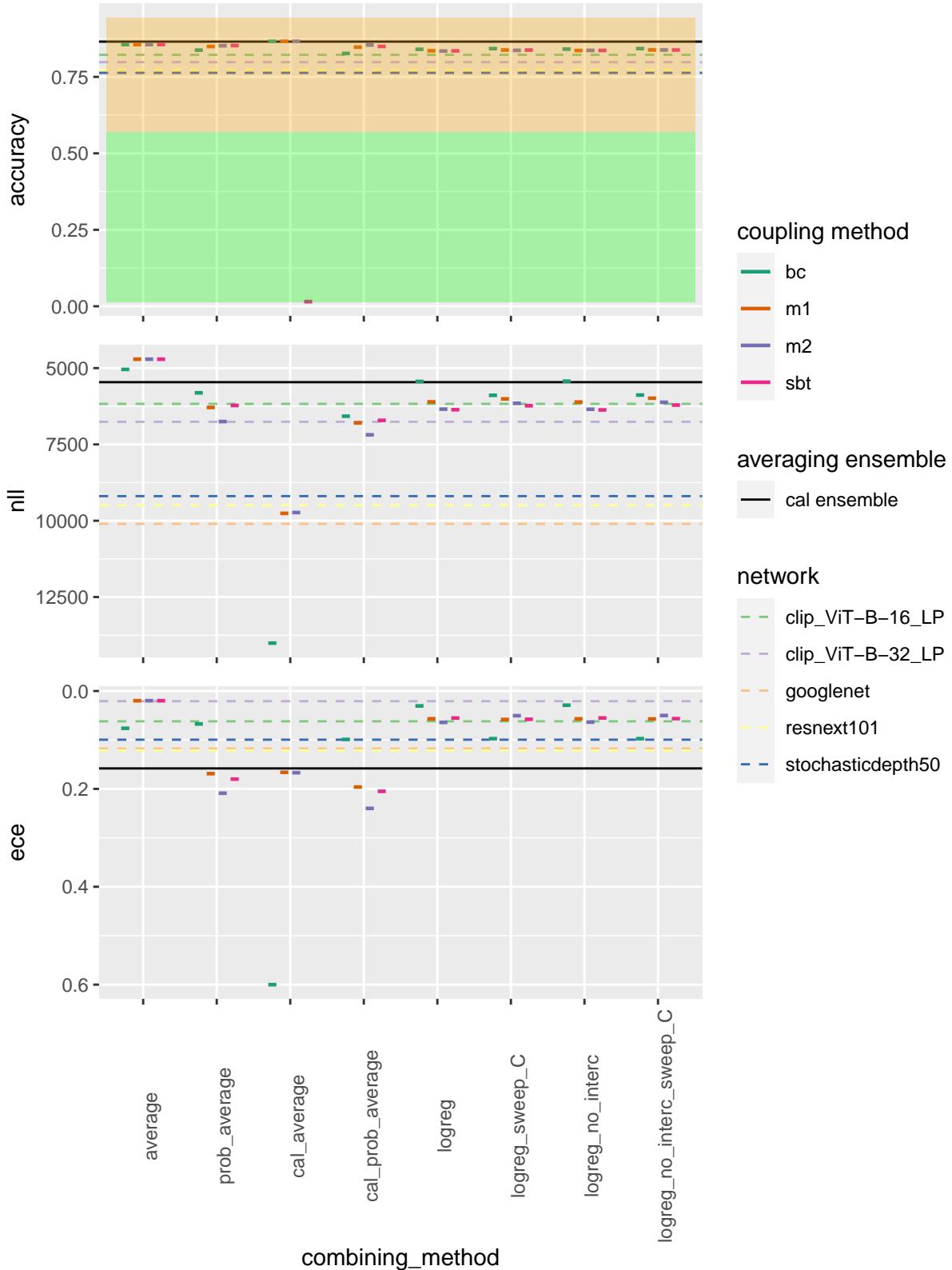
Ensemble metrics  
Error inconsistency 0.465700000524521



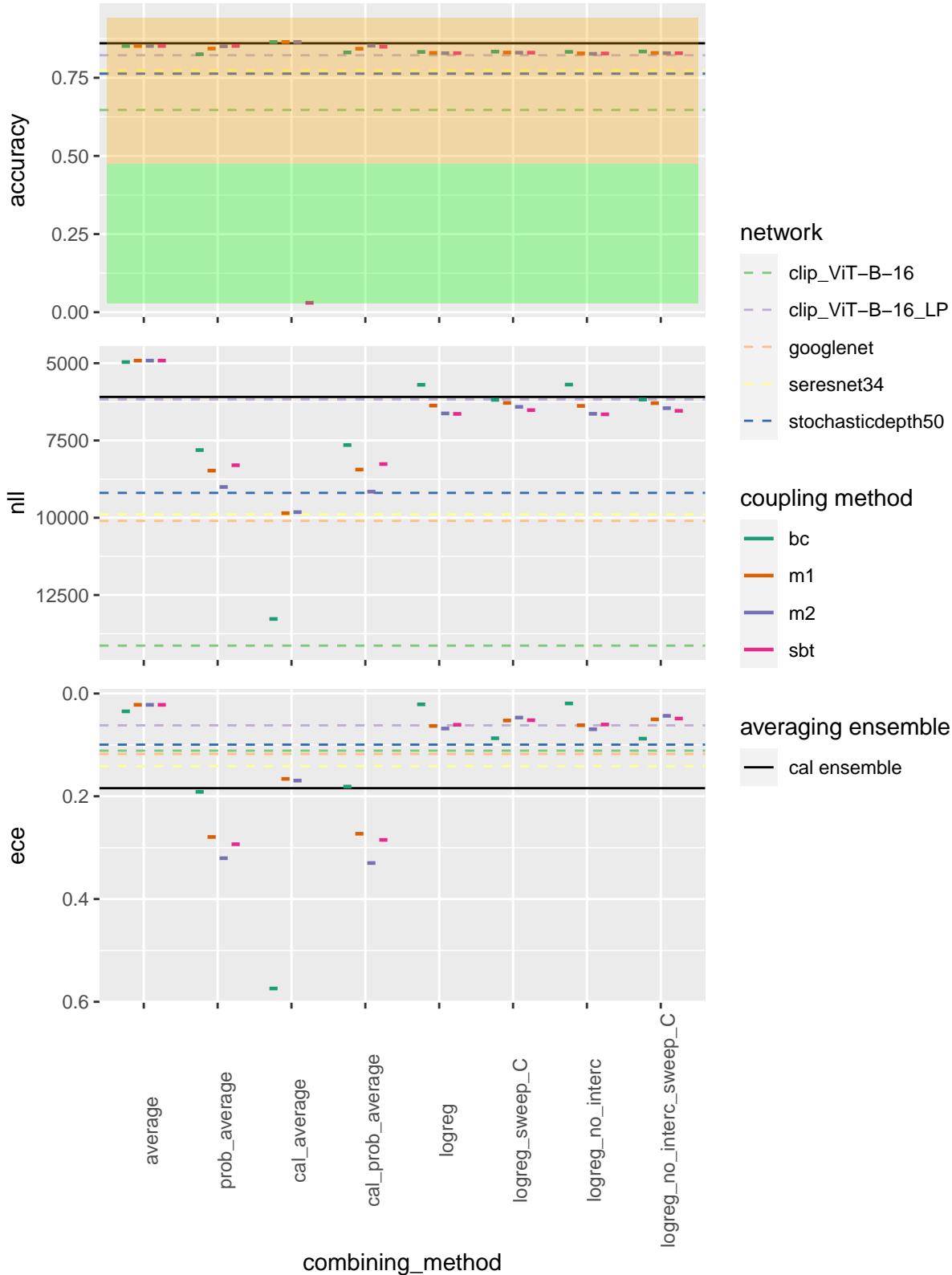
Ensemble metrics  
Error inconsistency 0.353499978780746



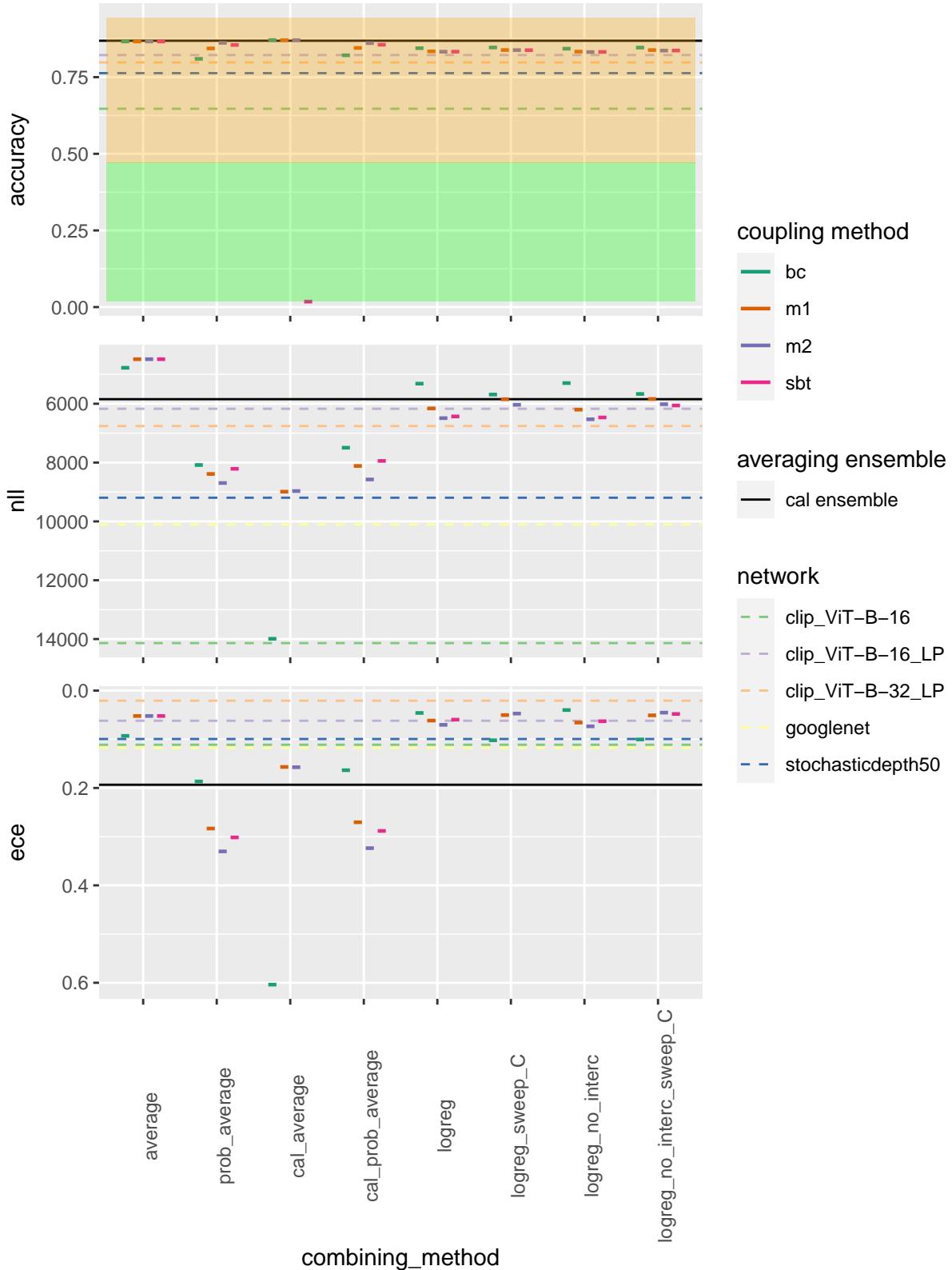
Ensemble metrics  
Error inconsistency 0.373499989509583



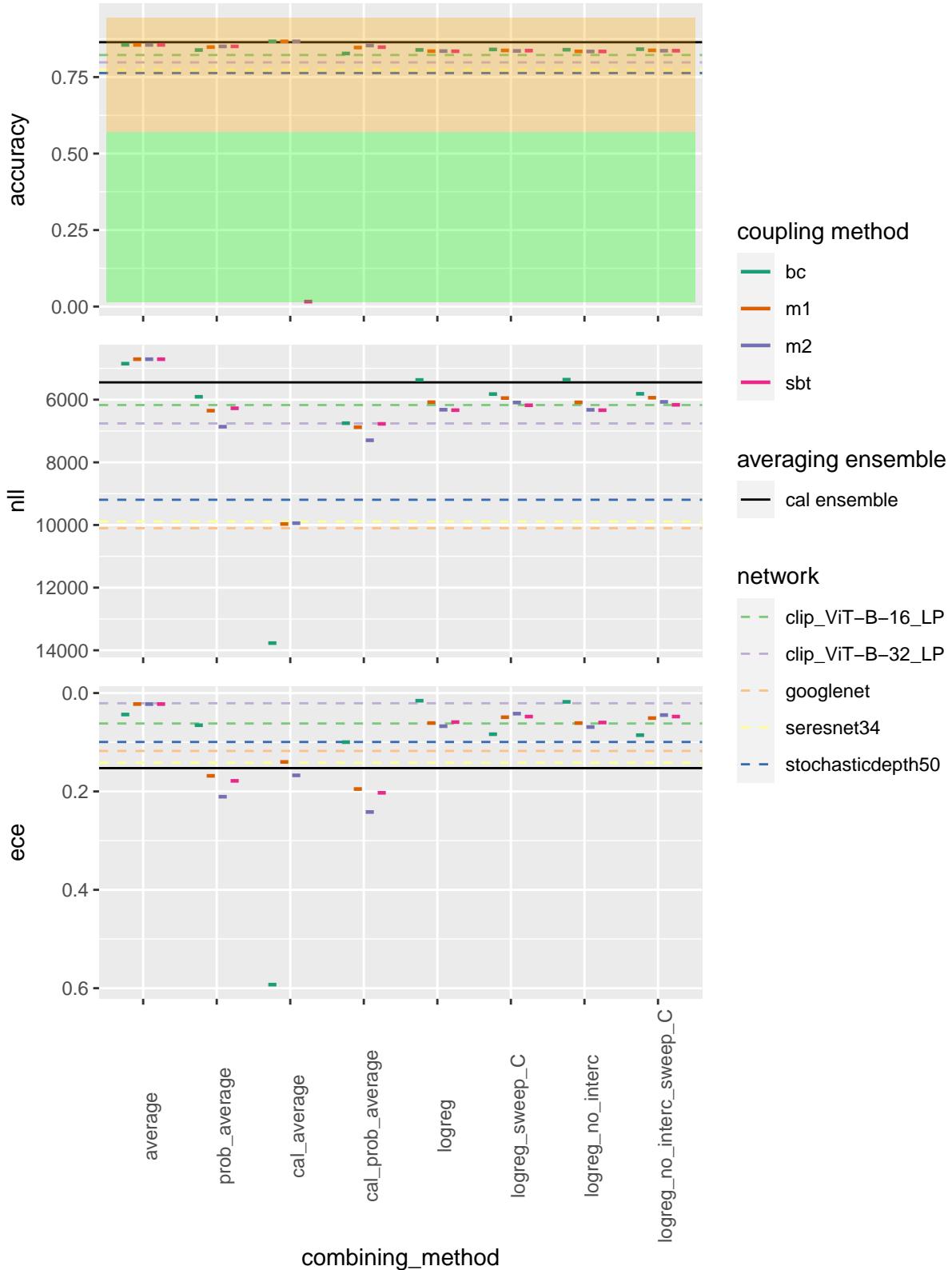
Ensemble metrics  
Error inconsistency 0.466899991035461



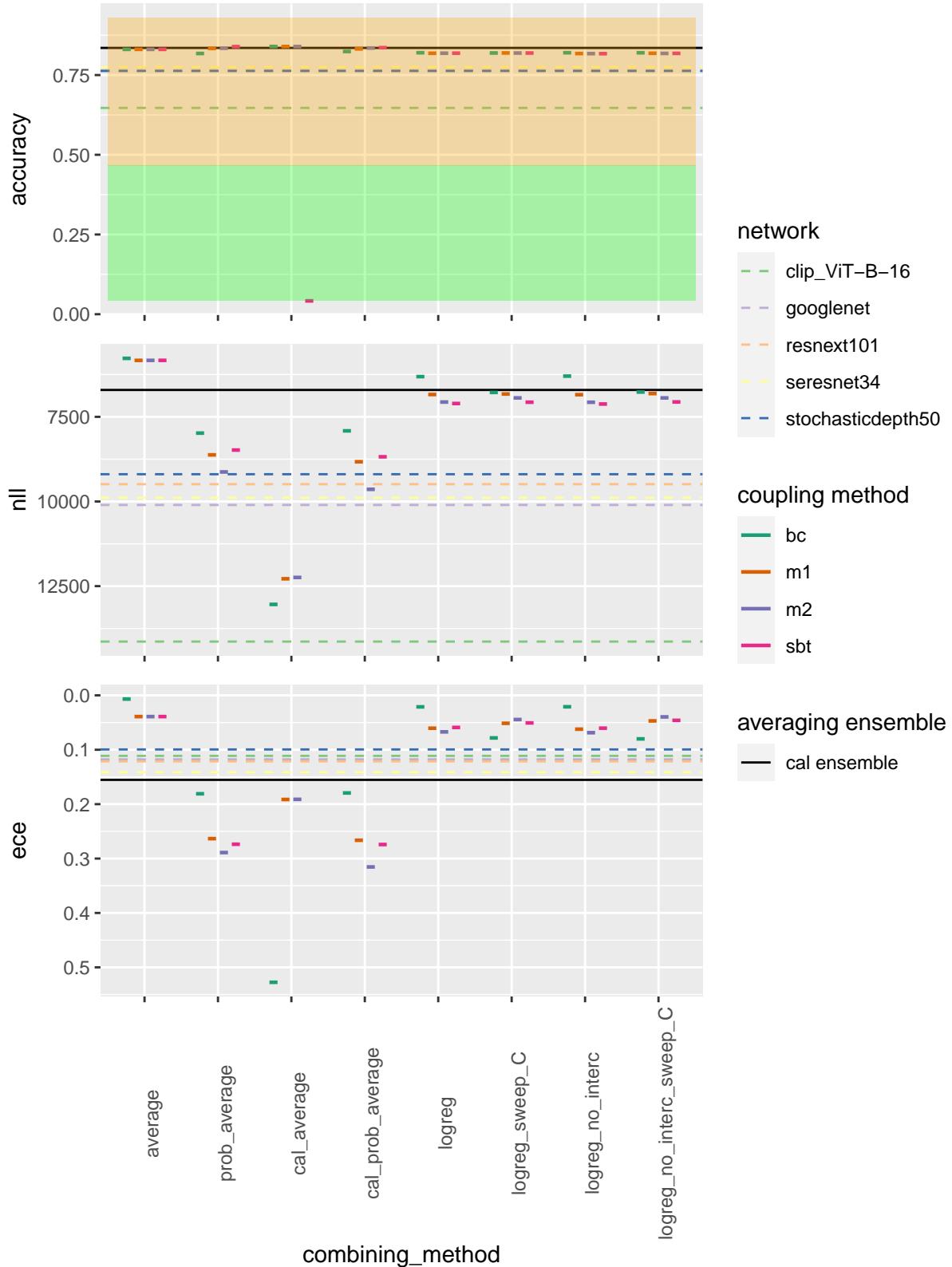
Ensemble metrics  
Error inconsistency 0.472199976444244



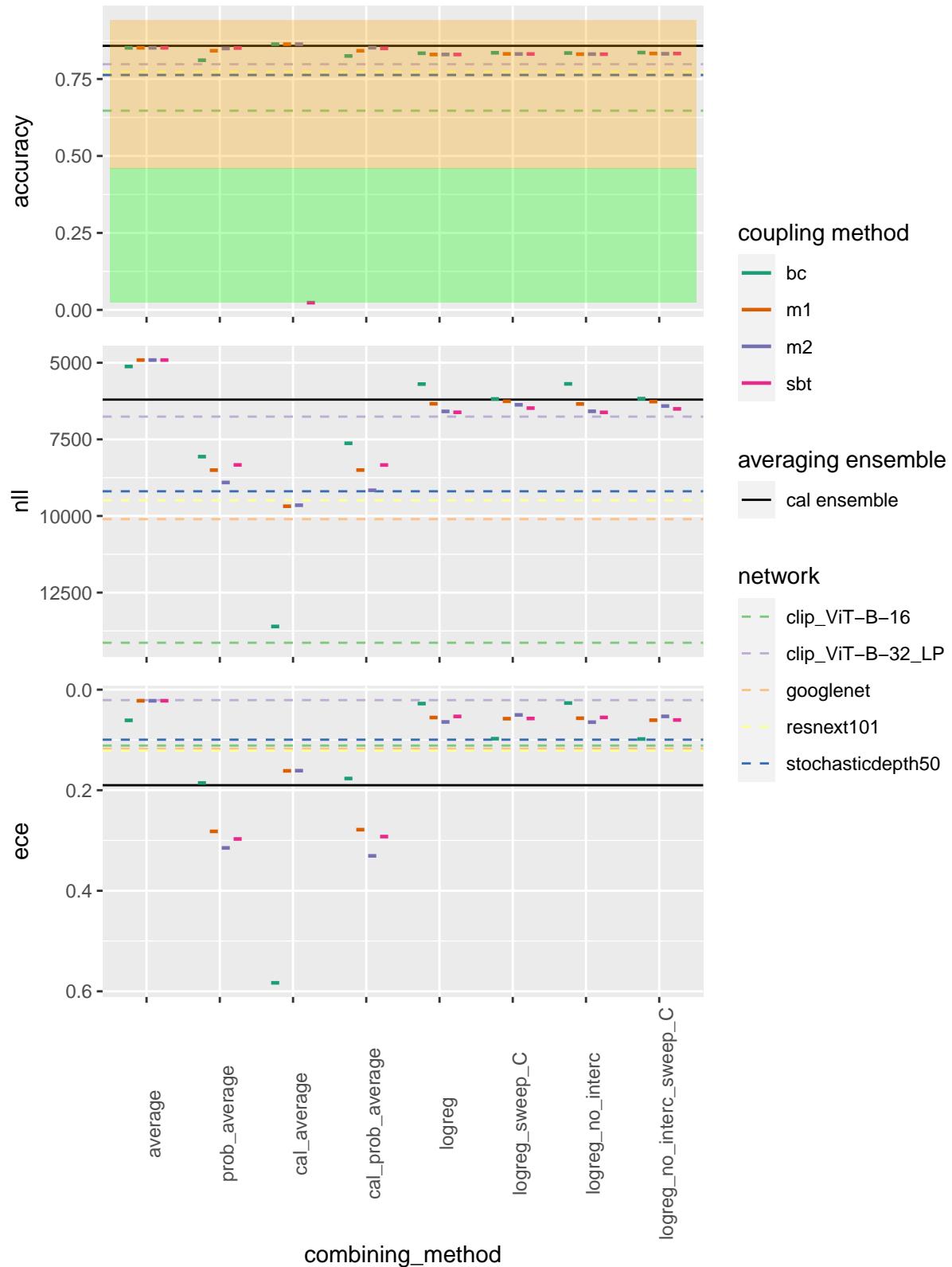
Ensemble metrics  
Error inconsistency 0.373400002717972



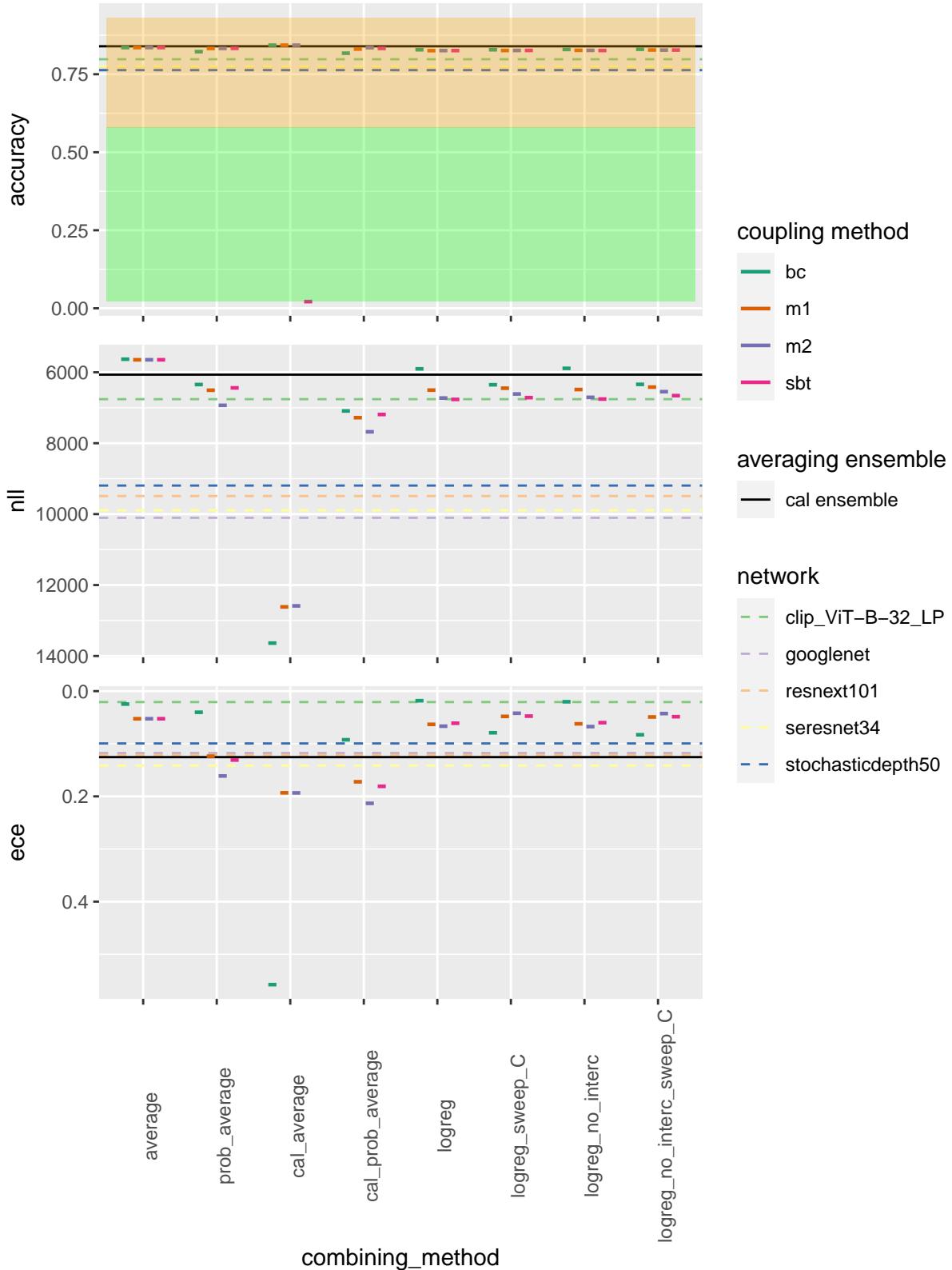
Ensemble metrics  
Error inconsistency 0.463299989700317



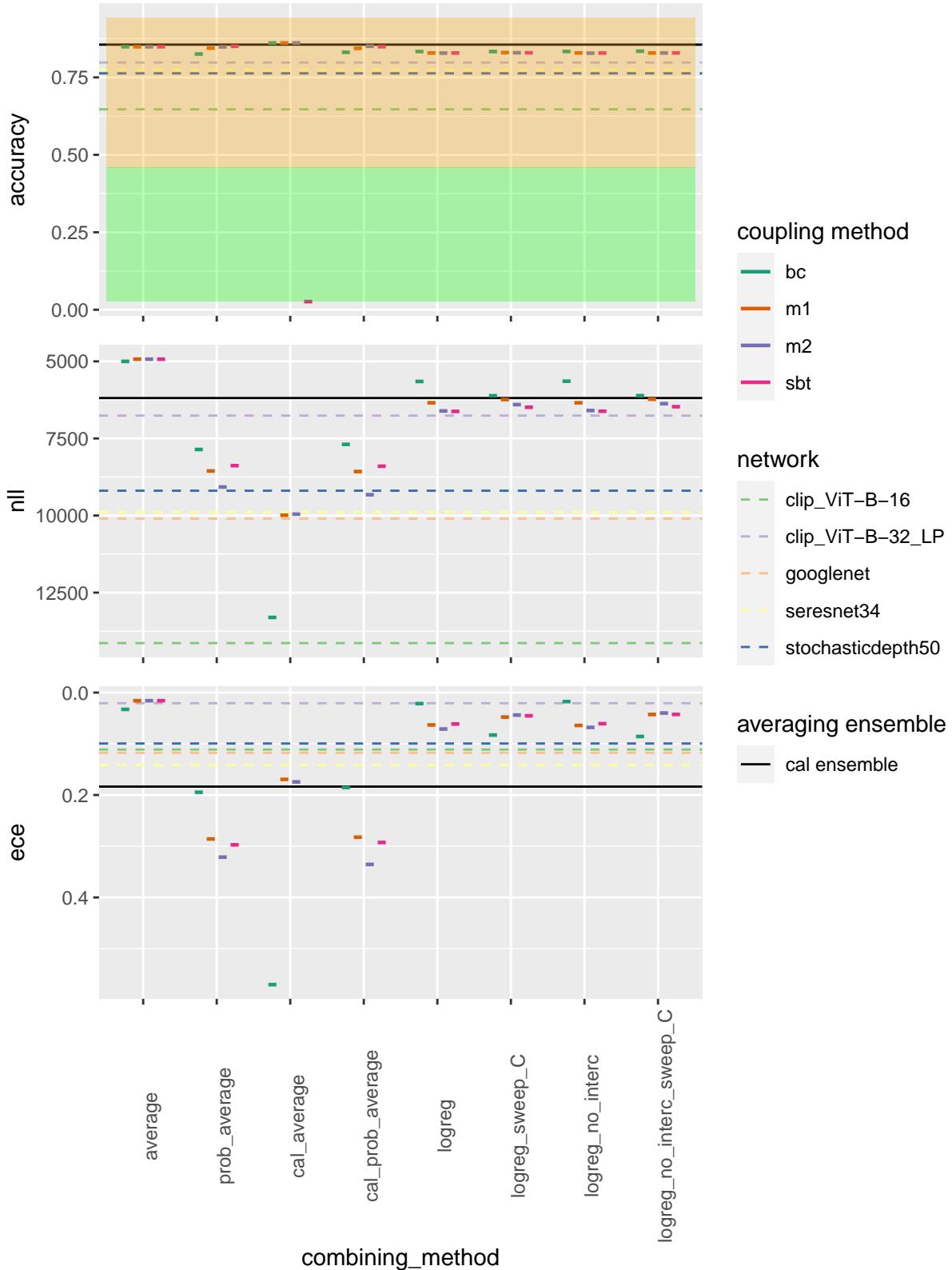
Ensemble metrics  
Error inconsistency 0.482899993658066



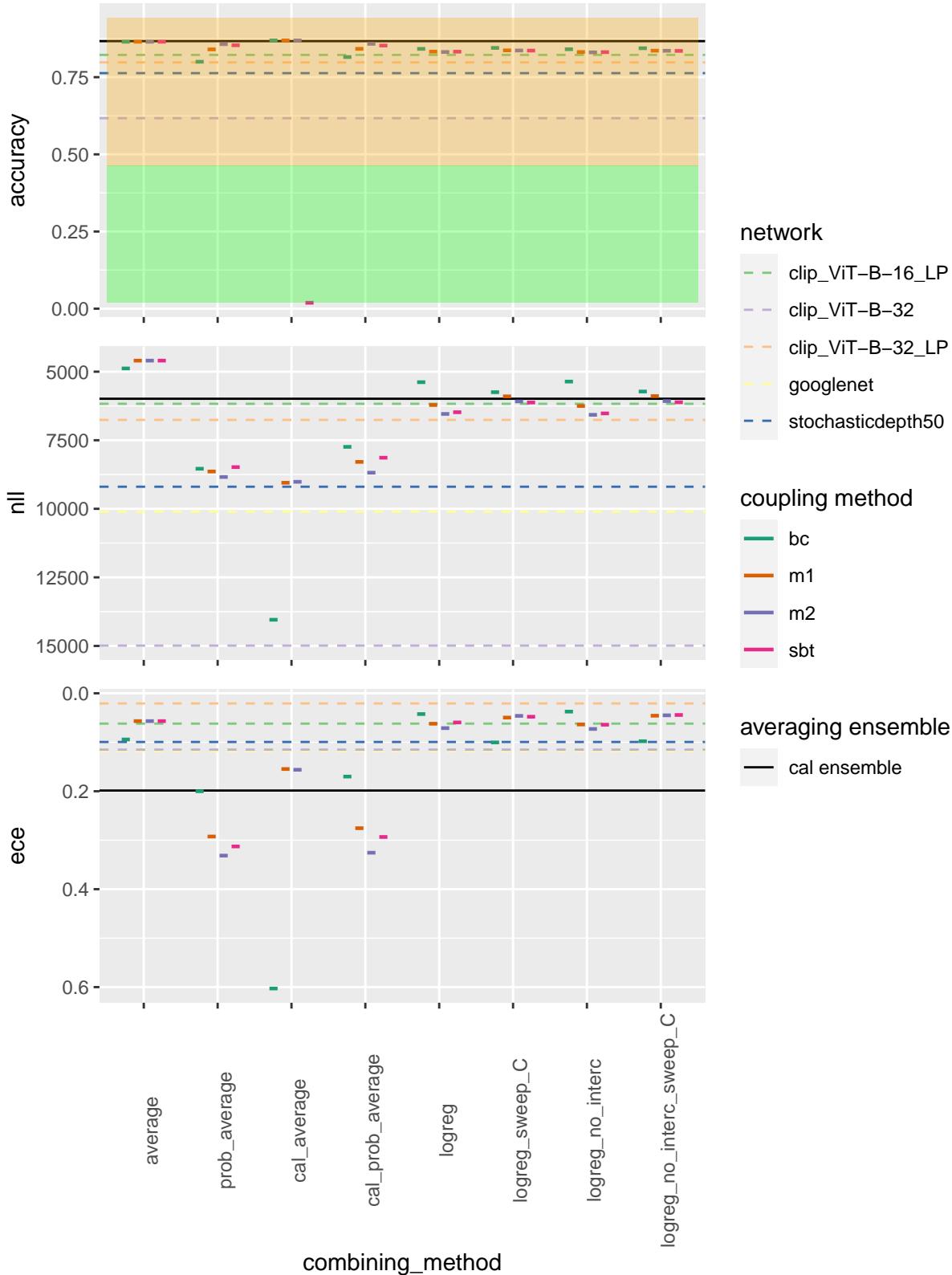
Ensemble metrics  
Error inconsistency 0.351300001144409



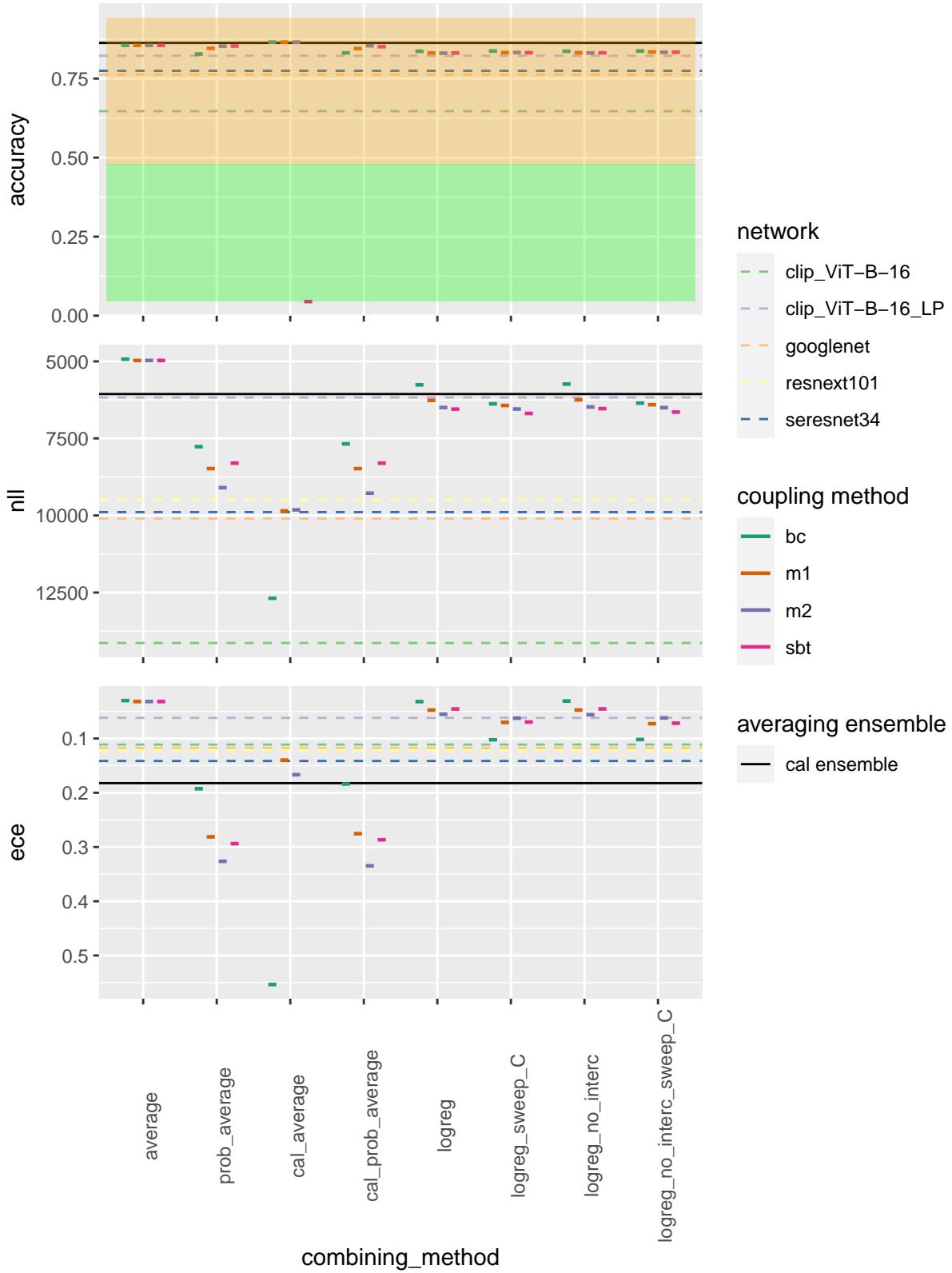
Ensemble metrics  
Error inconsistency 0.48319998383522



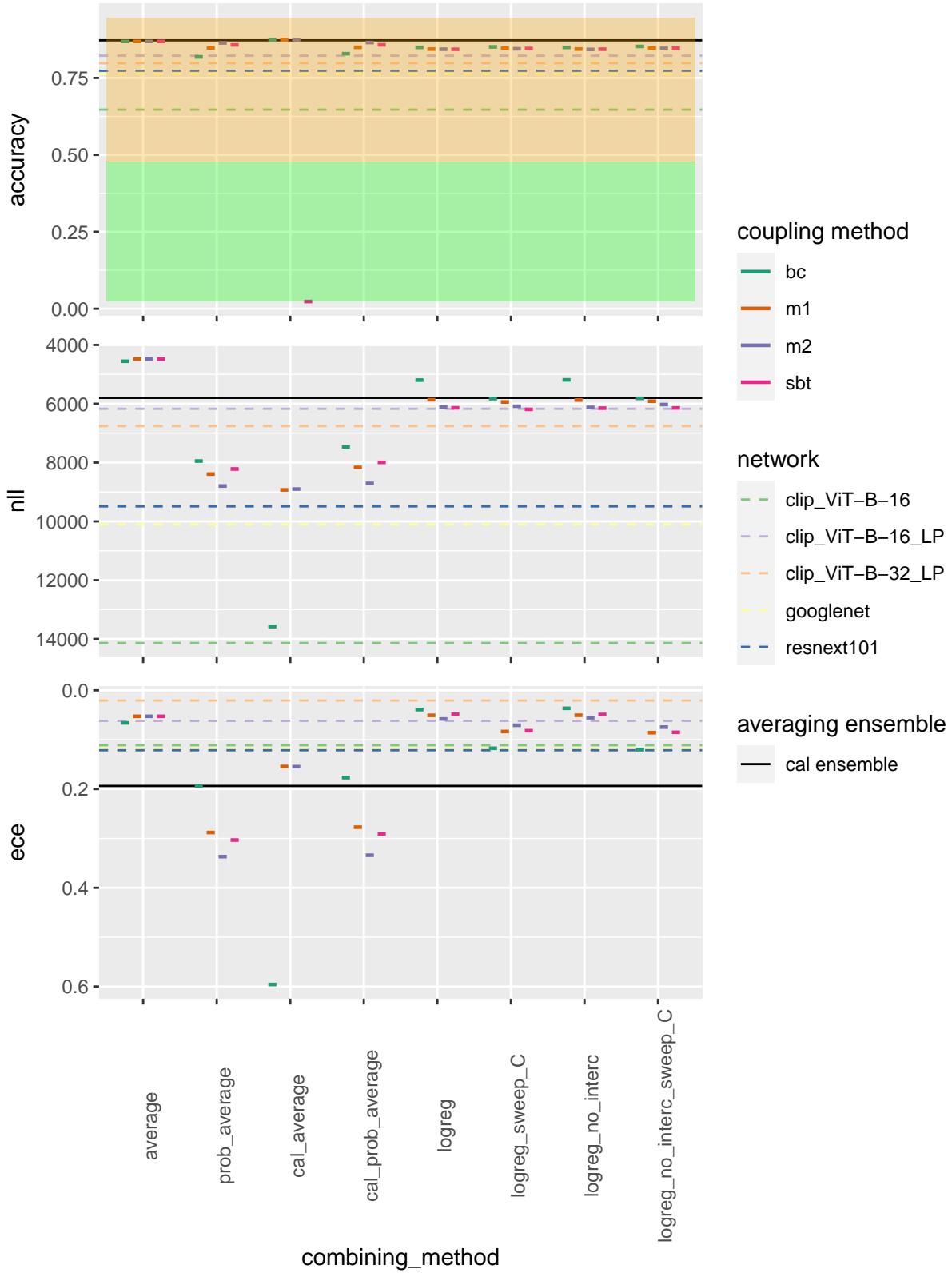
Ensemble metrics  
Error inconsistency 0.480499982833862



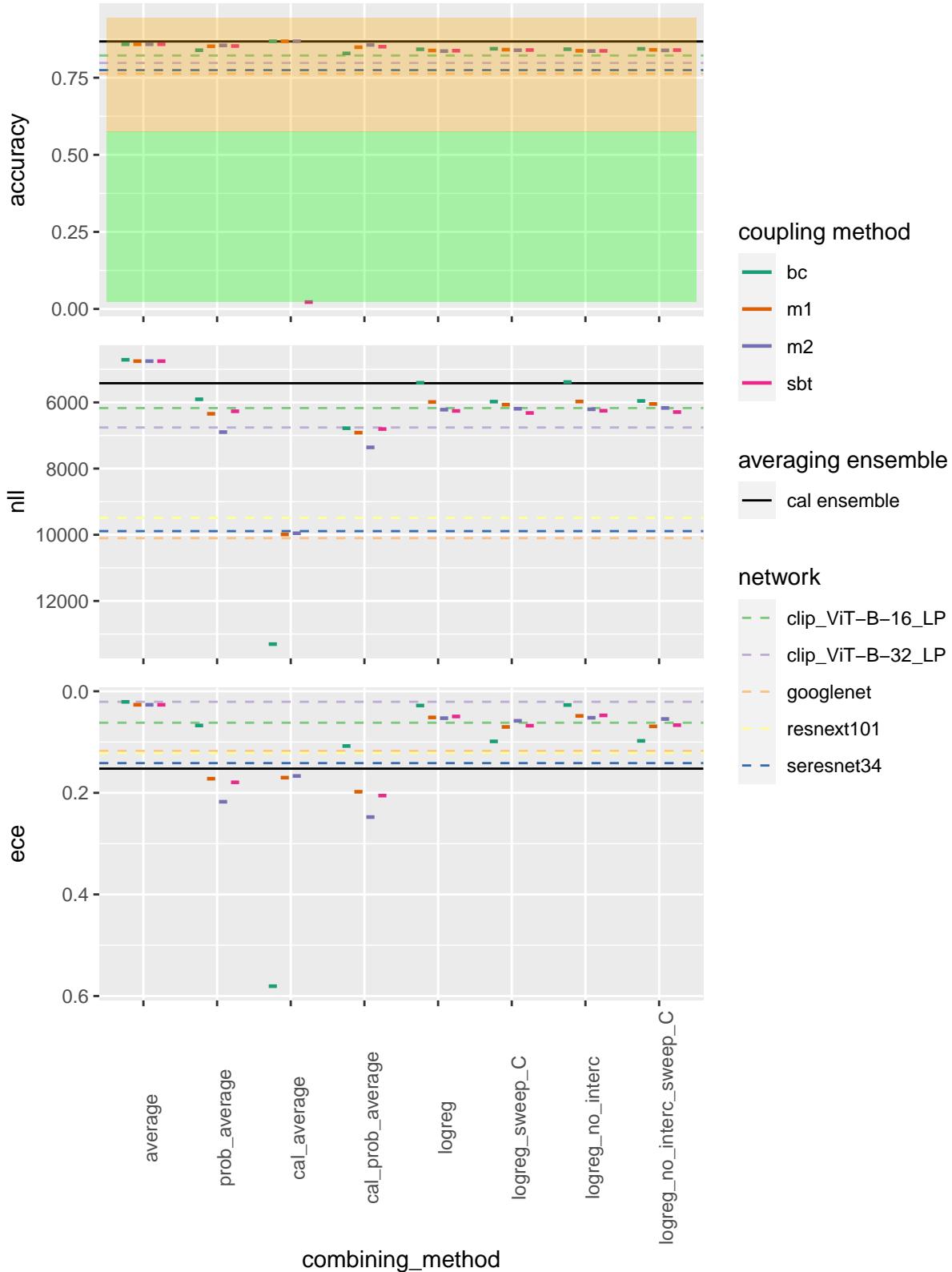
Ensemble metrics  
Error inconsistency 0.464699983596802



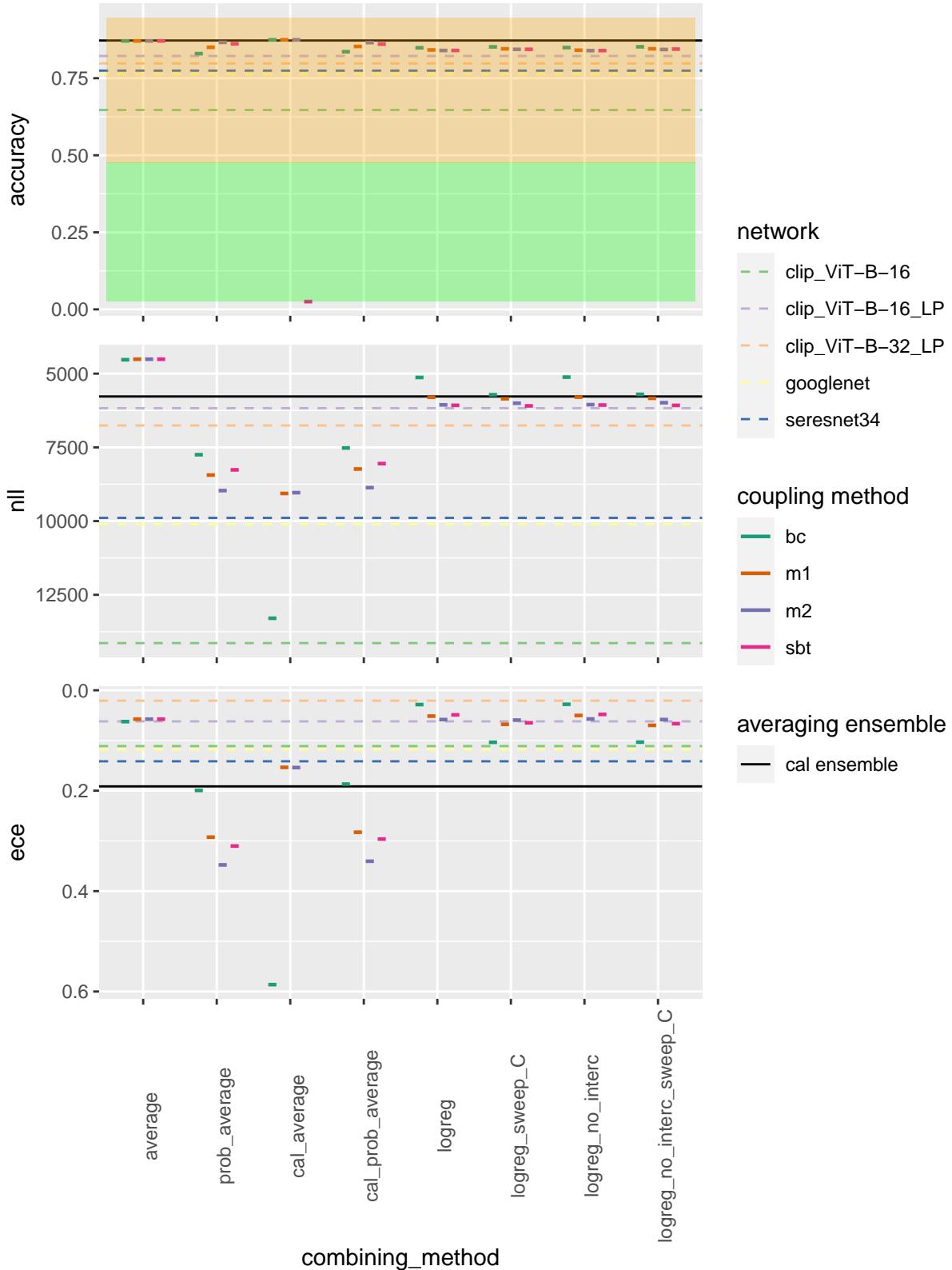
Ensemble metrics  
Error inconsistency 0.469300001859665



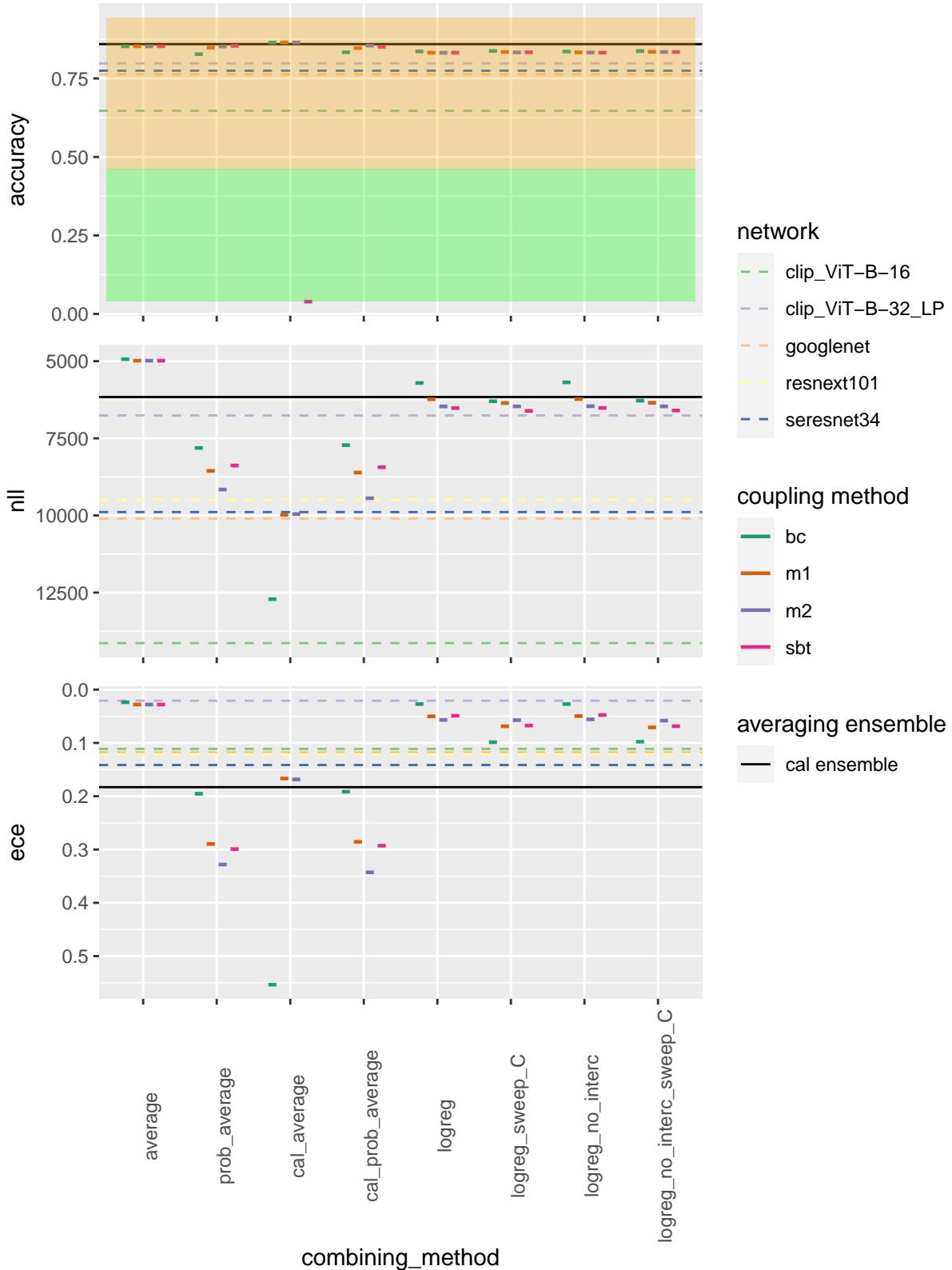
Ensemble metrics  
Error inconsistency 0.369699984788895



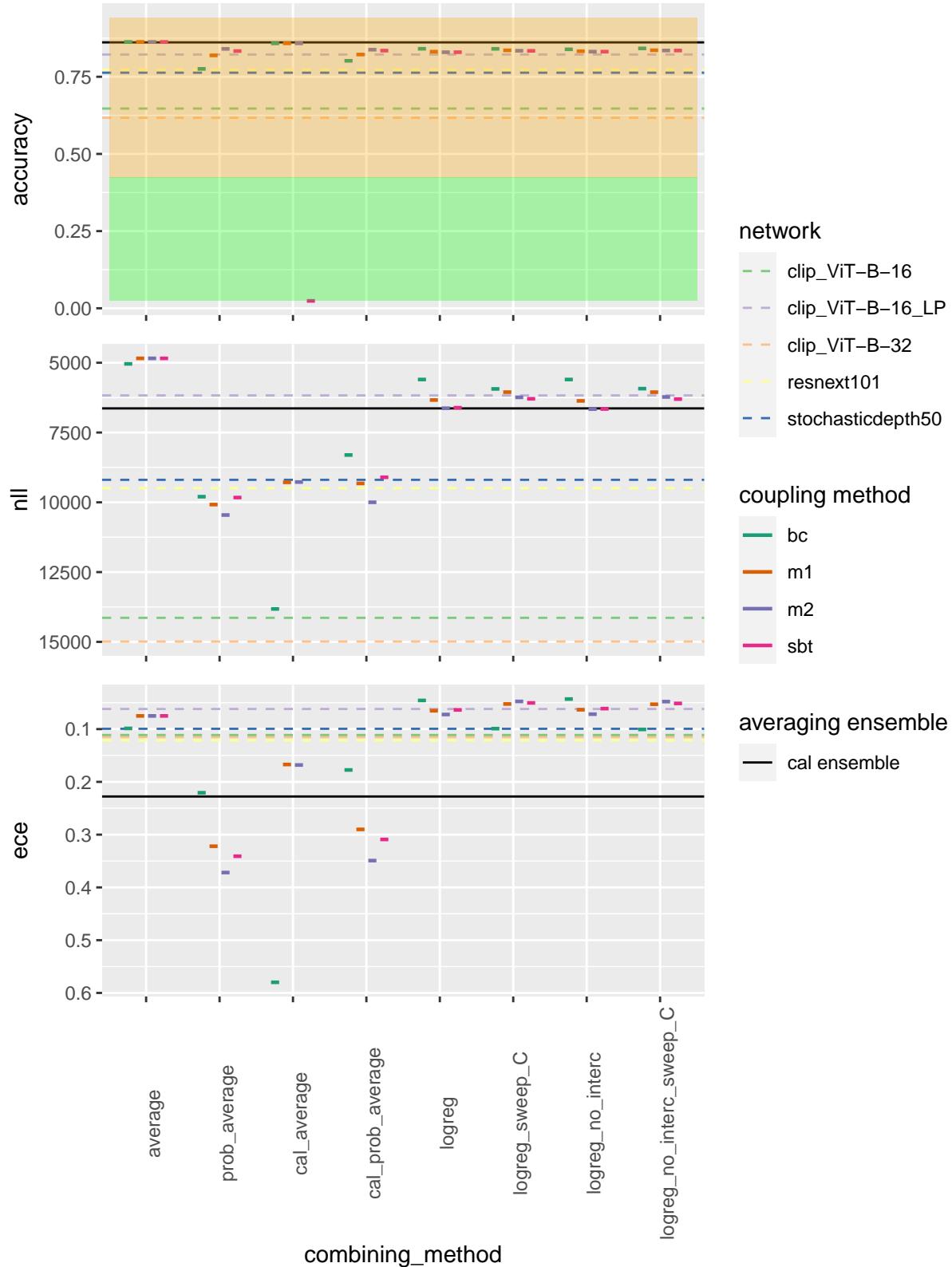
Ensemble metrics  
Error inconsistency 0.470899999141693



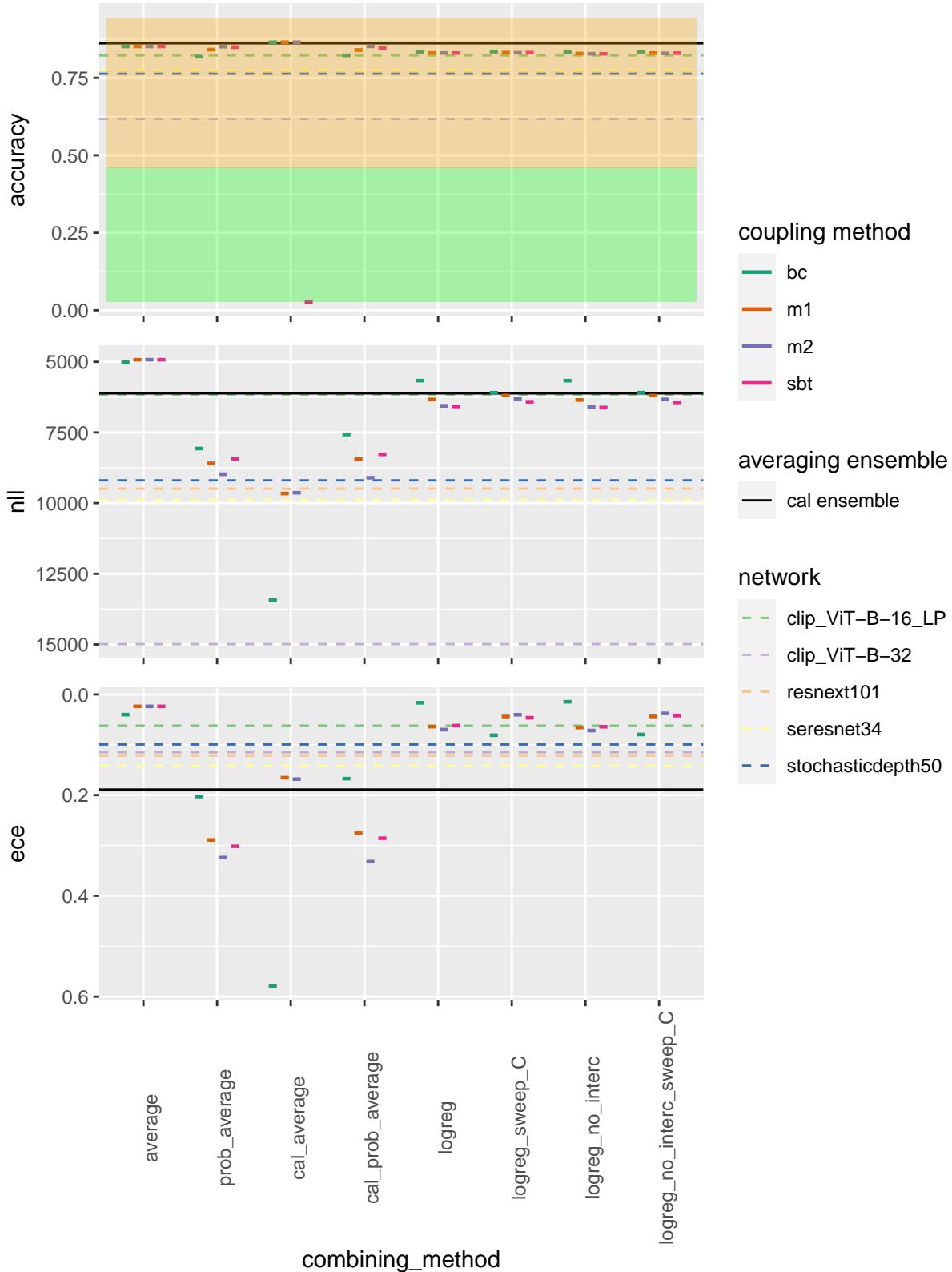
Ensemble metrics  
Error inconsistency 0.481499999761581



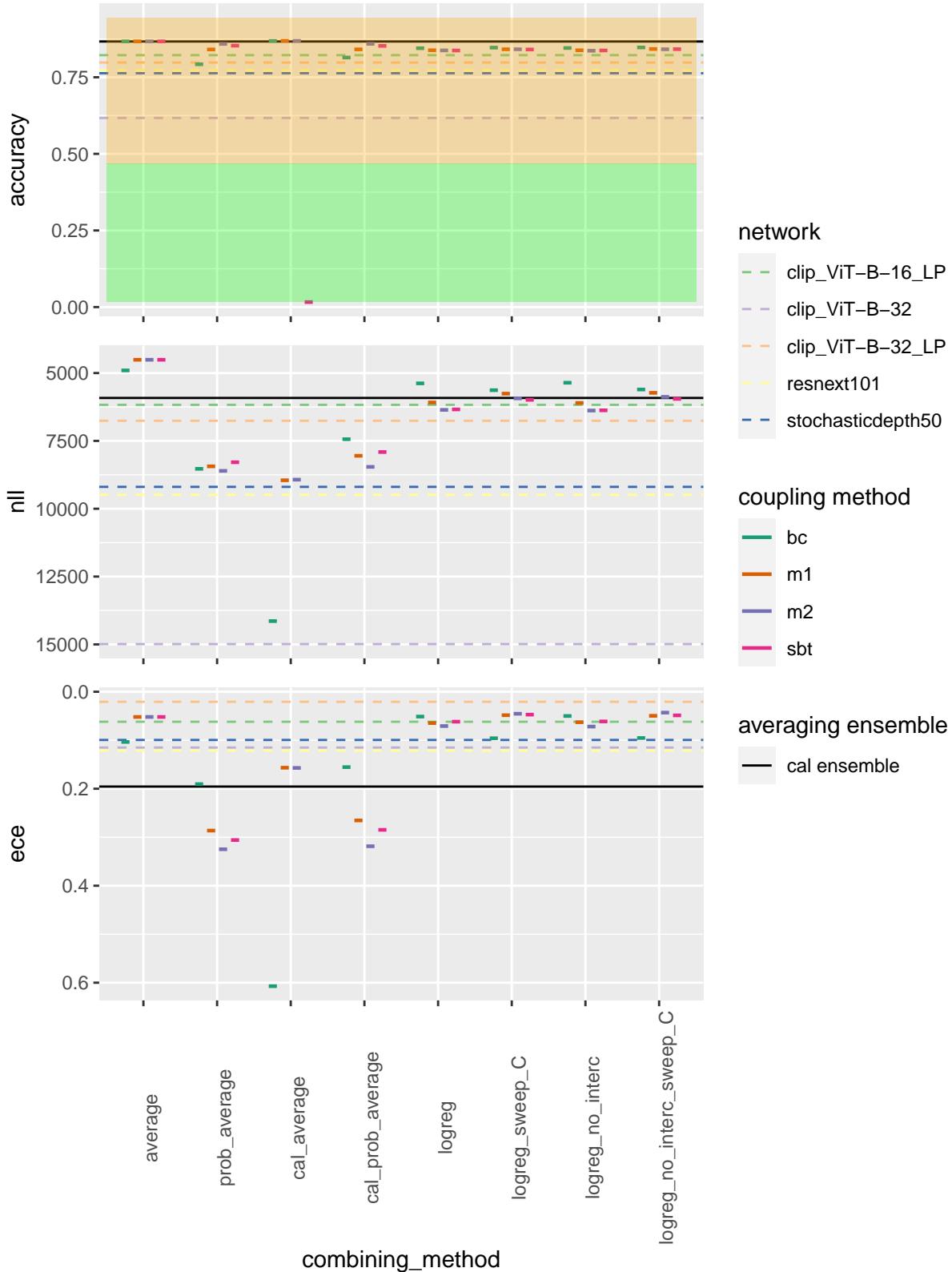
Ensemble metrics  
Error inconsistency 0.517099976539612



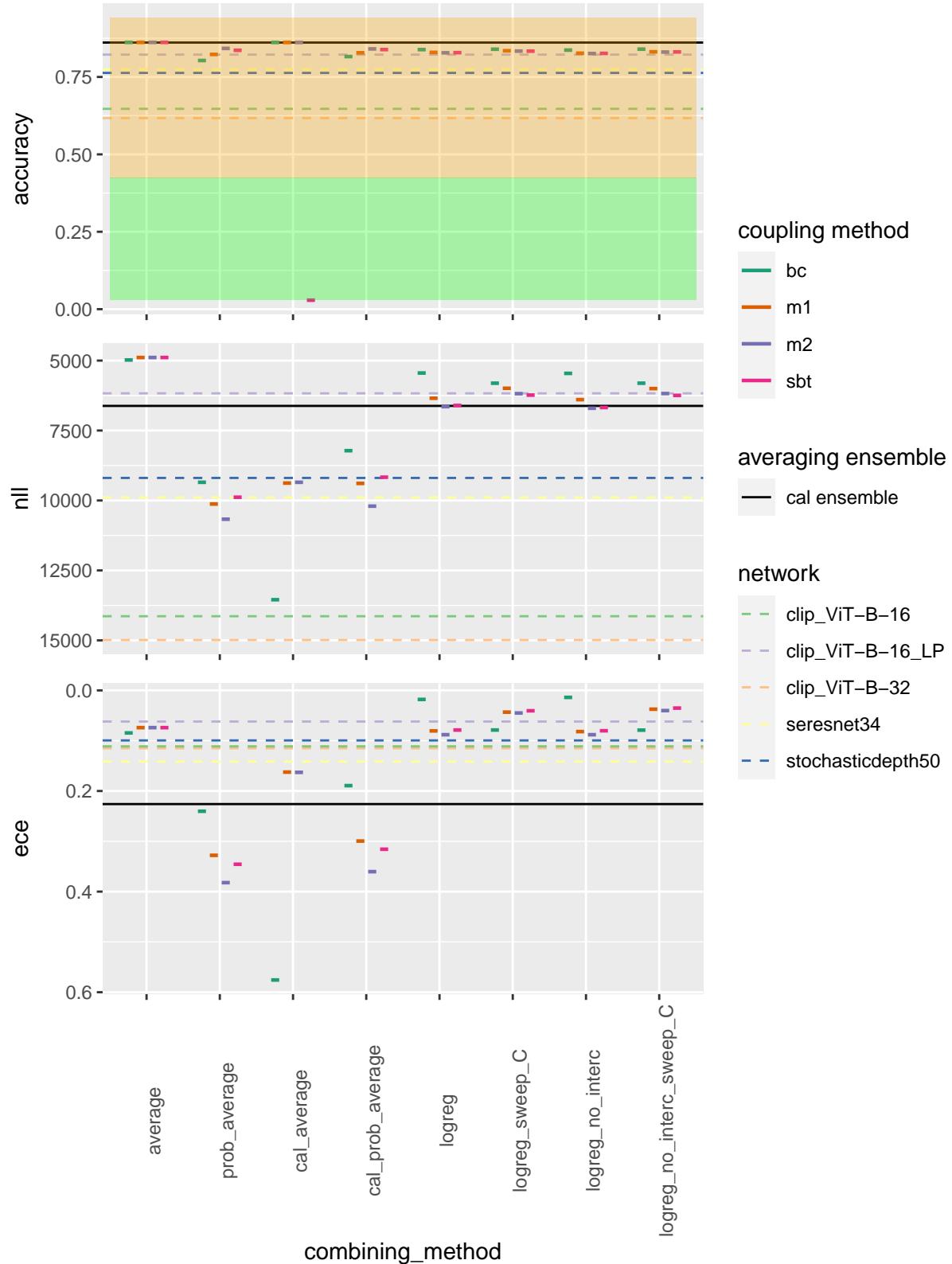
Ensemble metrics  
Error inconsistency 0.484399974346161



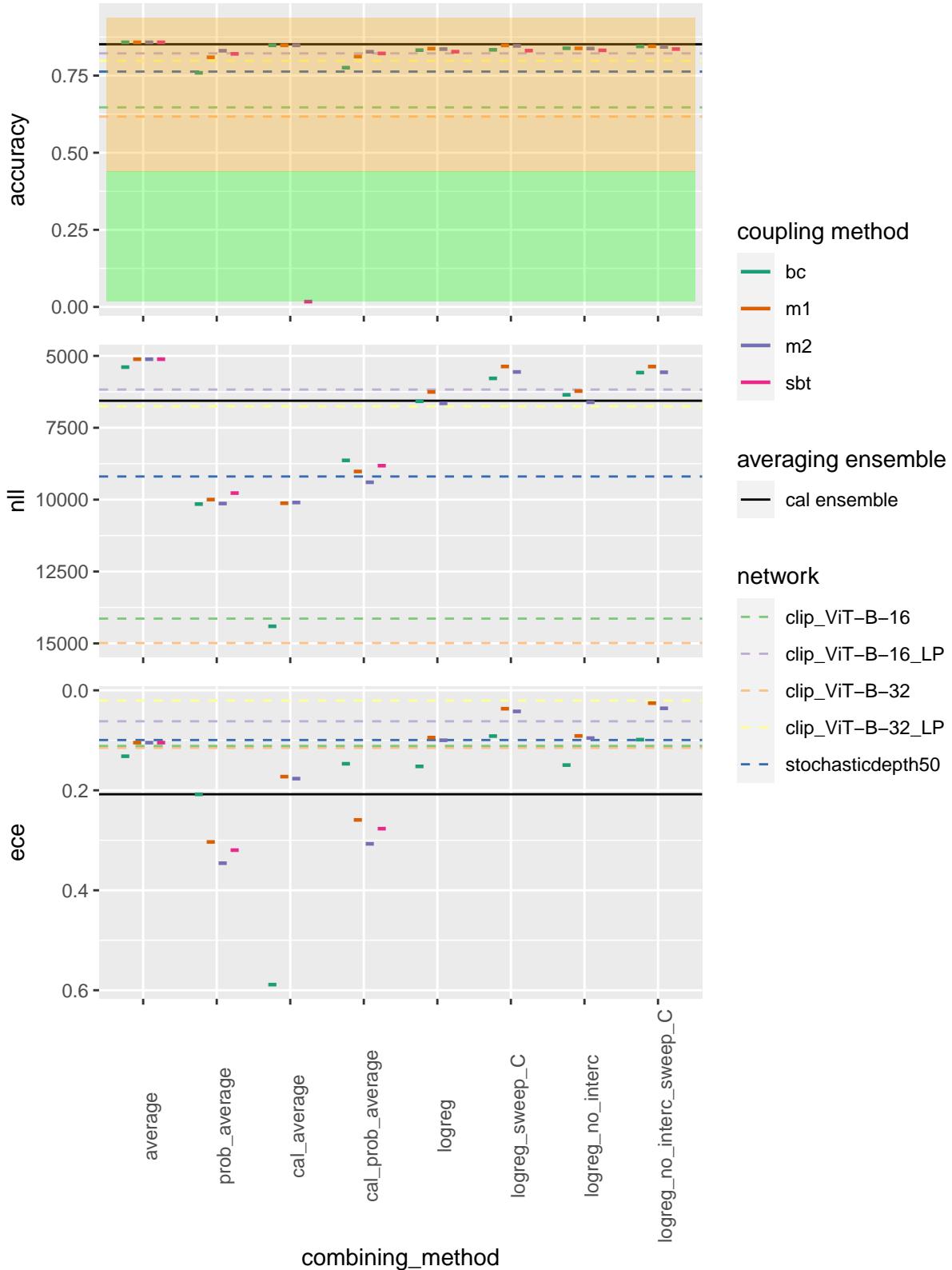
Ensemble metrics  
Error inconsistency 0.476999998092651



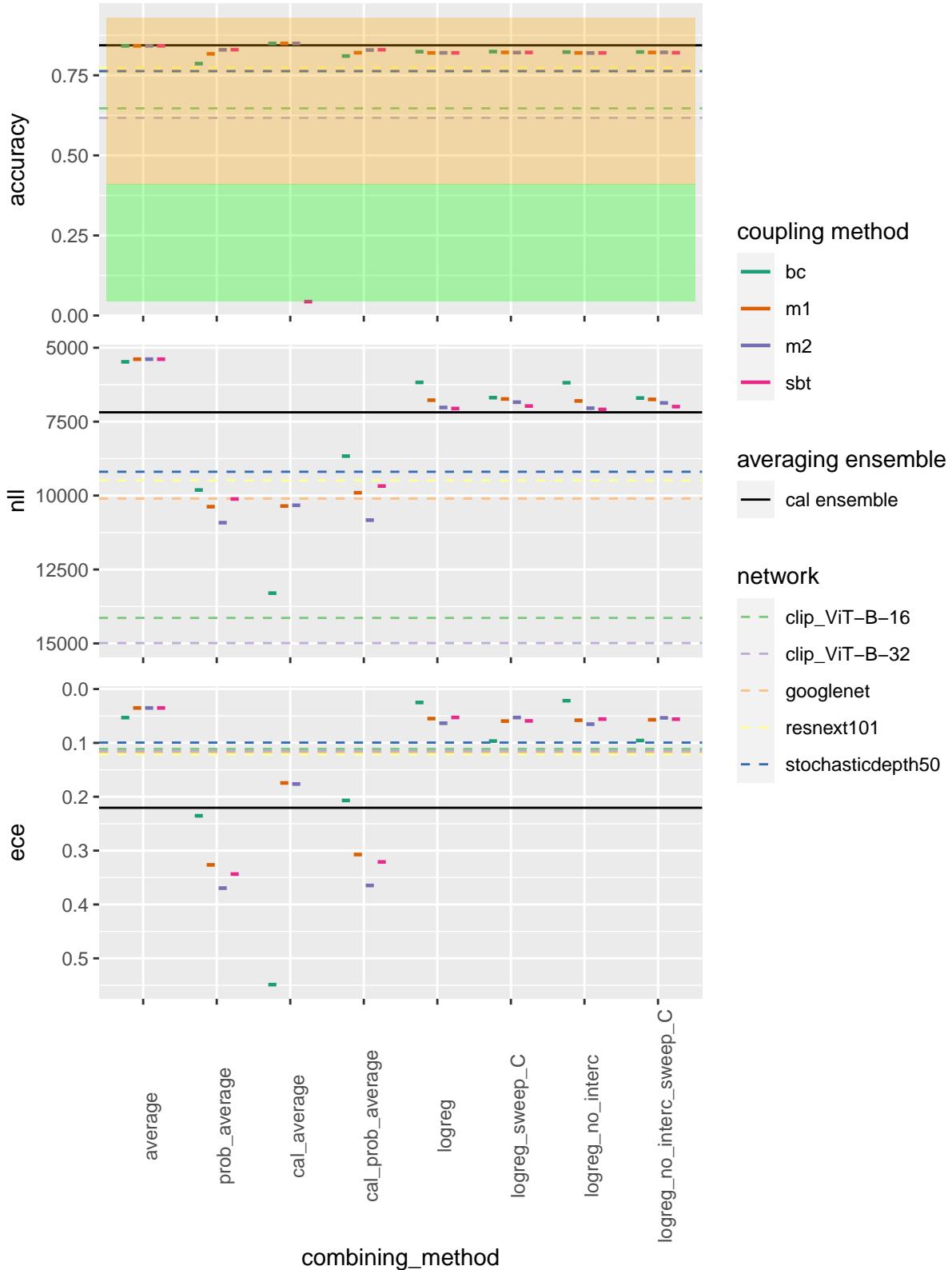
Ensemble metrics  
Error inconsistency 0.51690000295639



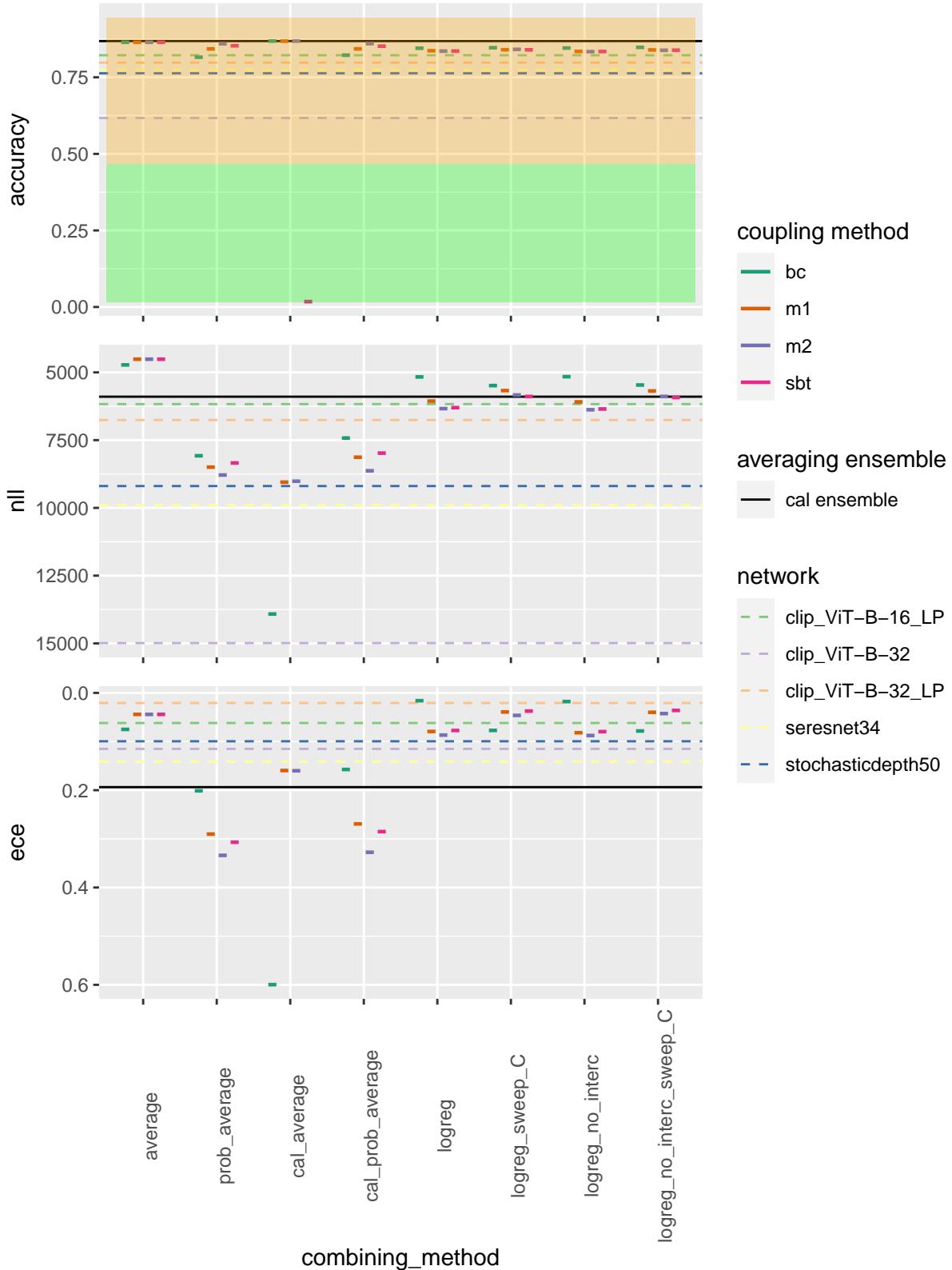
Ensemble metrics  
Error inconsistency 0.499299973249435



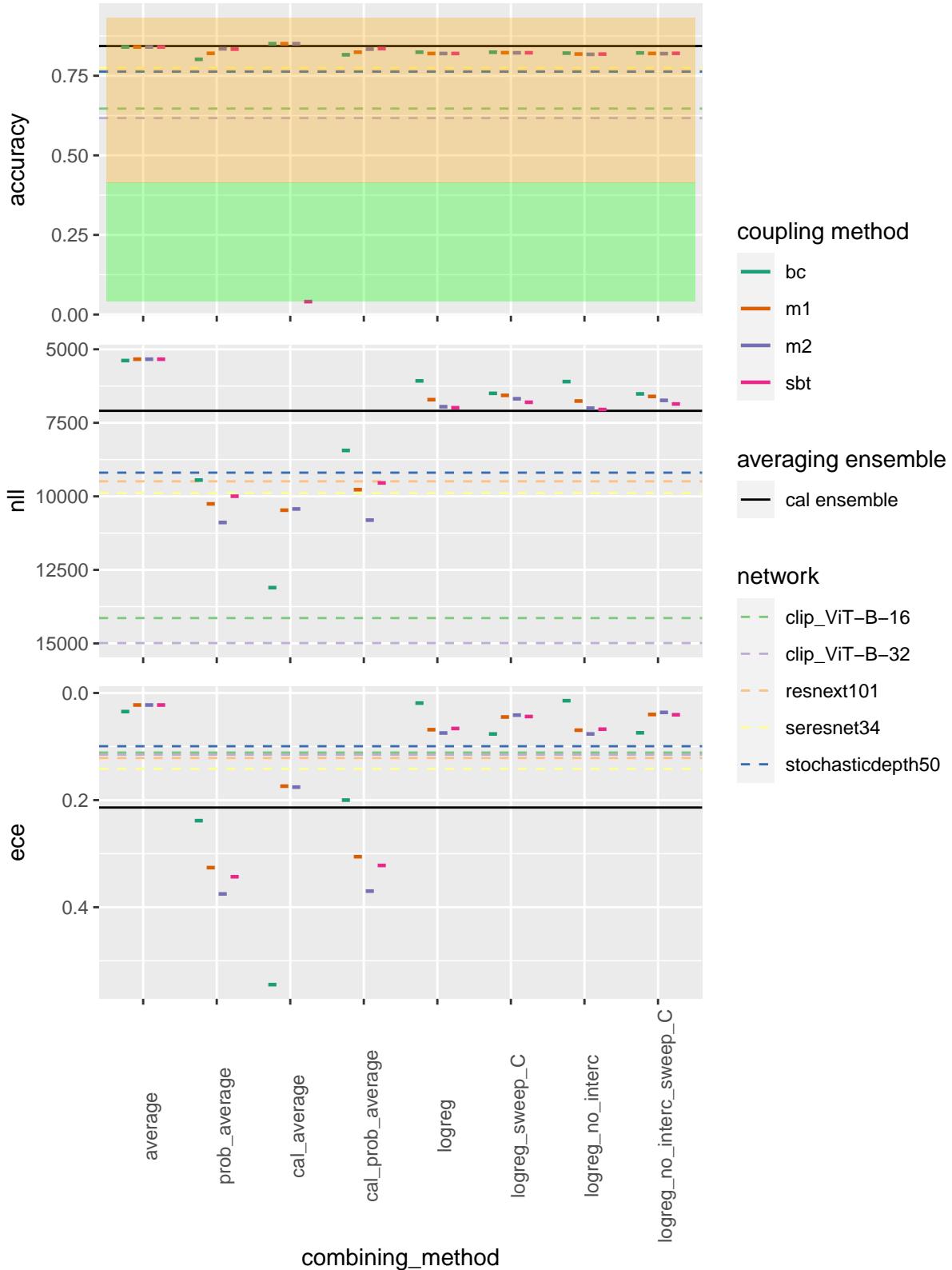
Ensemble metrics  
Error inconsistency 0.520900011062622



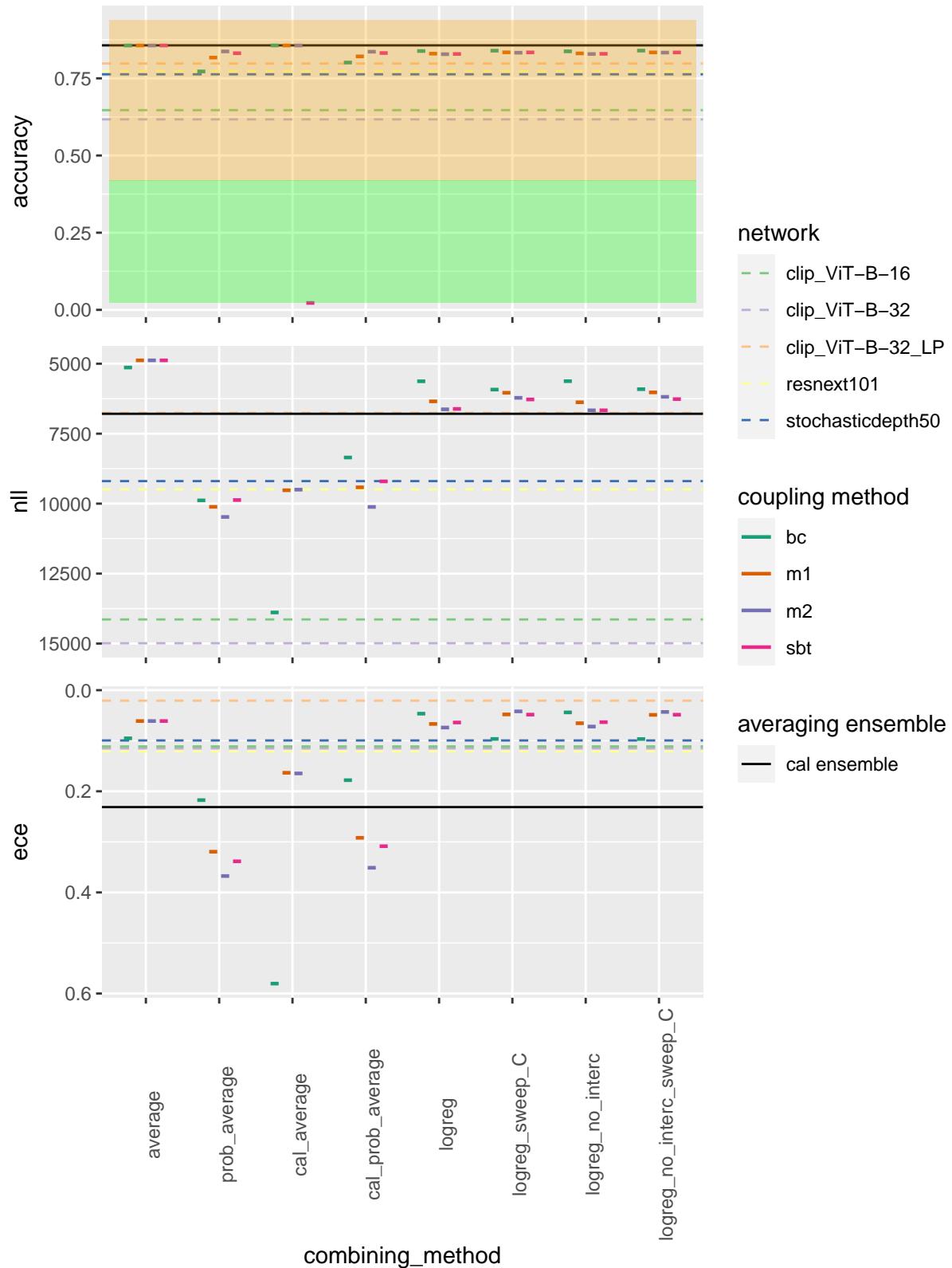
Ensemble metrics  
Error inconsistency 0.477699995040893



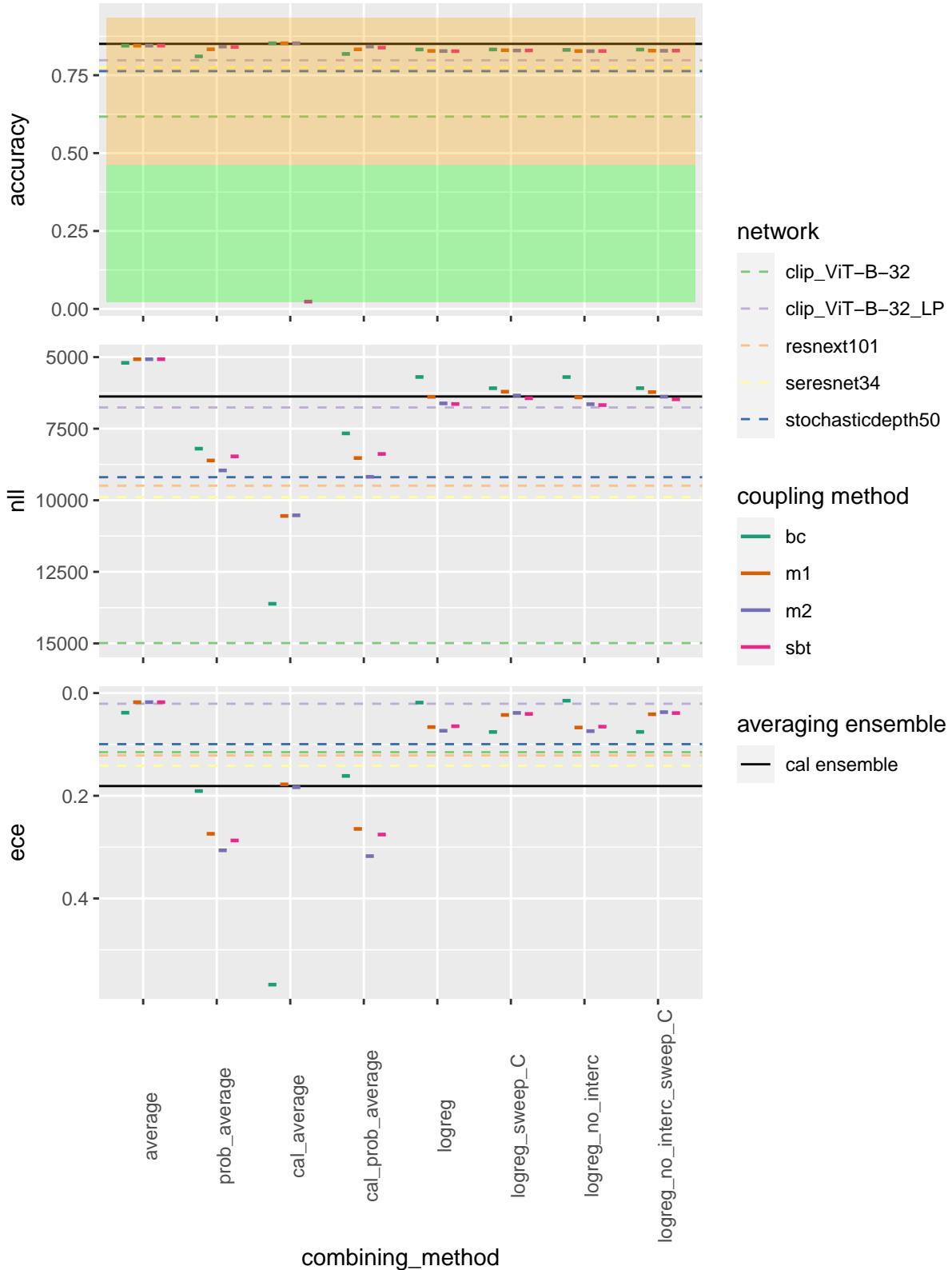
Ensemble metrics  
Error inconsistency 0.517700016498566



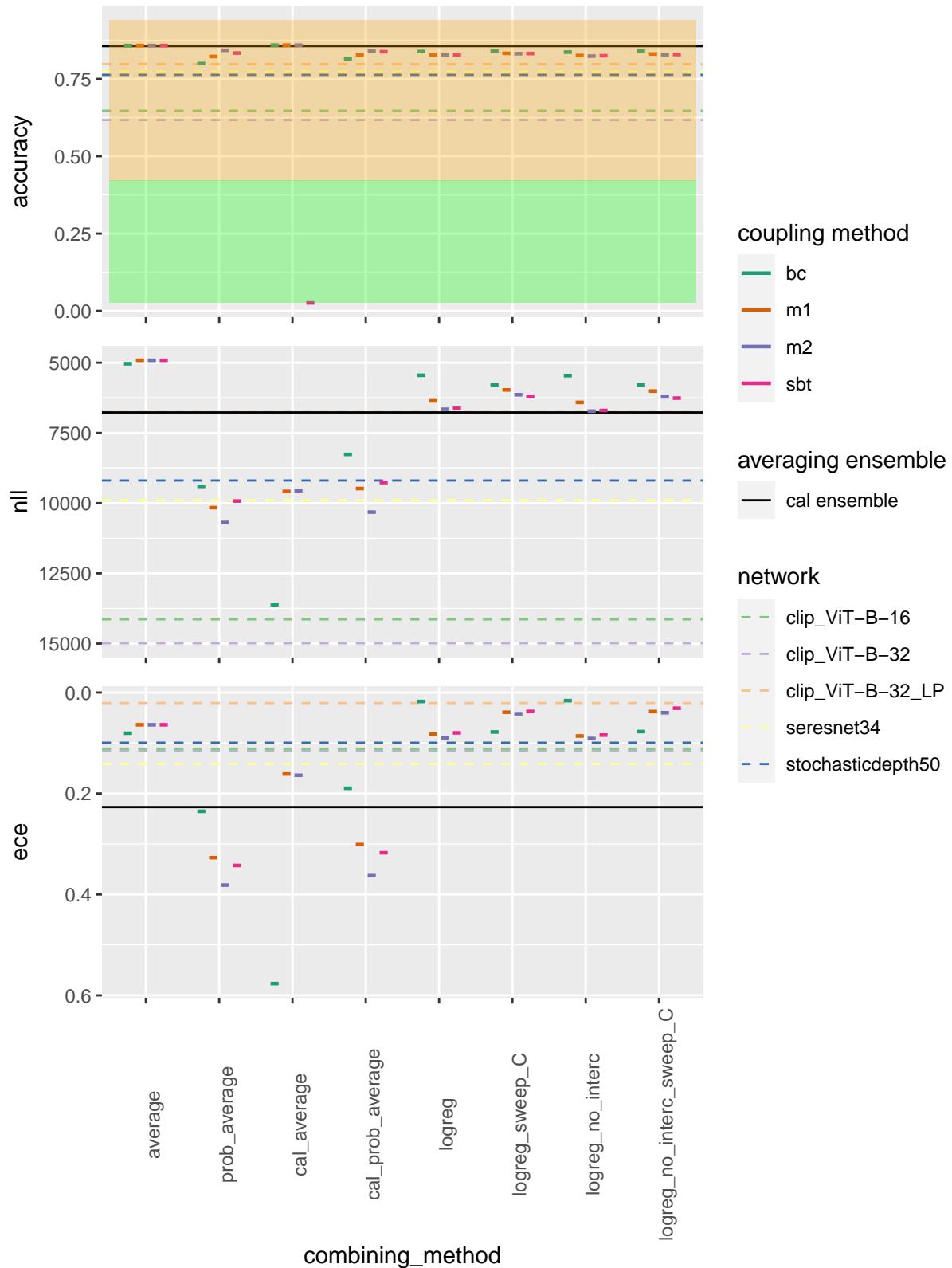
Ensemble metrics  
Error inconsistency 0.518999993801117



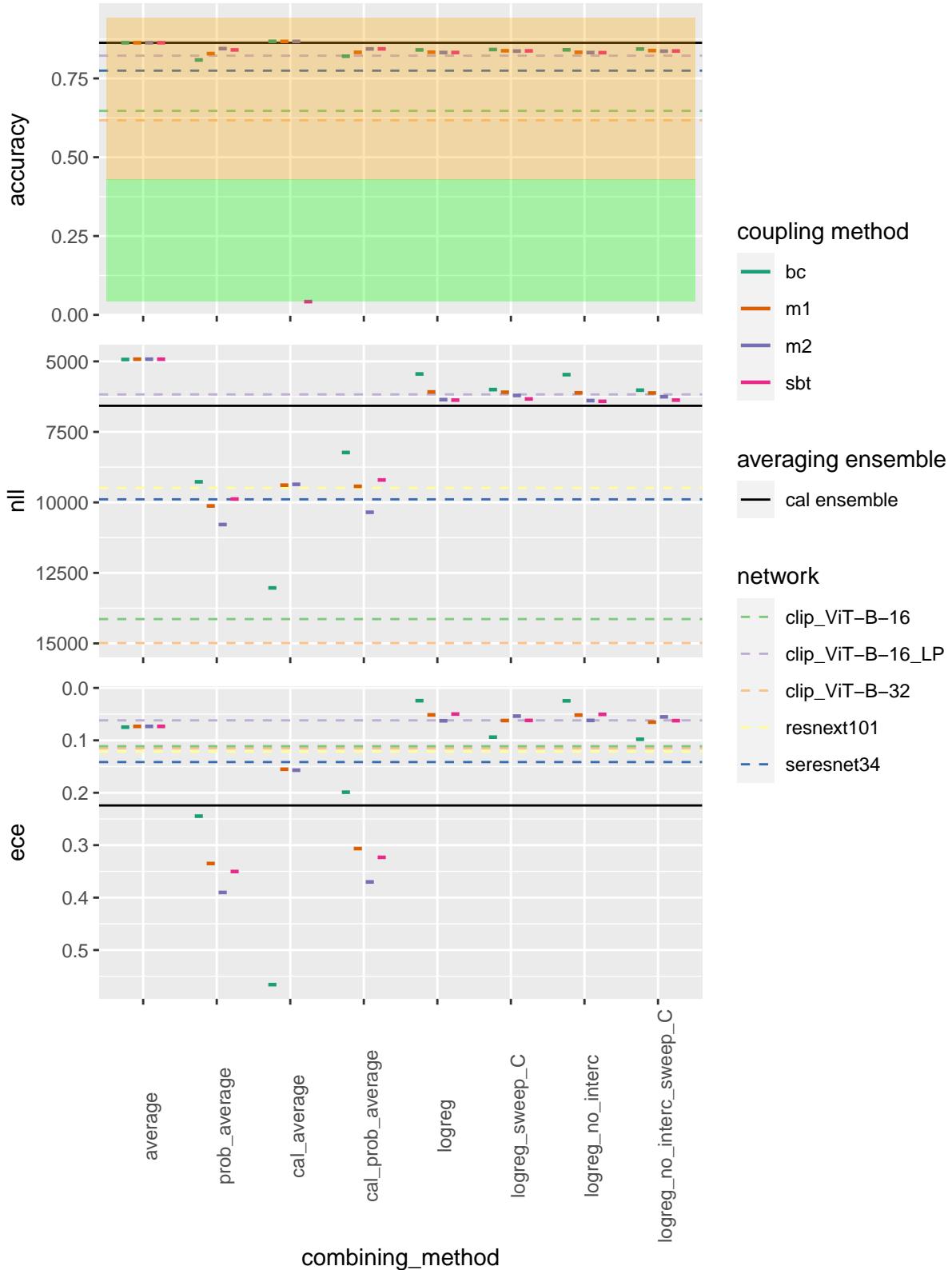
Ensemble metrics  
Error inconsistency 0.472199976444244



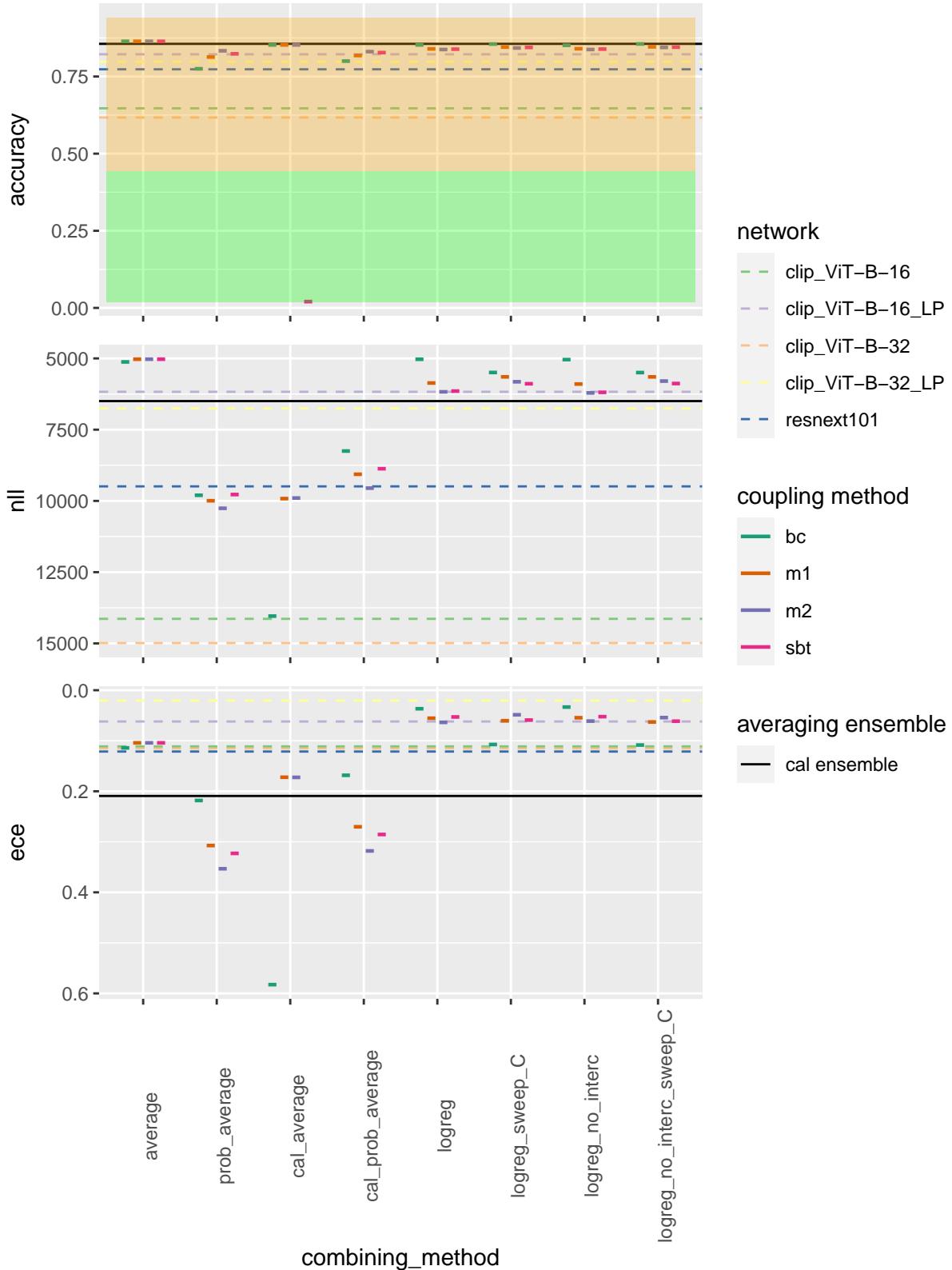
Ensemble metrics  
Error inconsistency 0.519400000572205



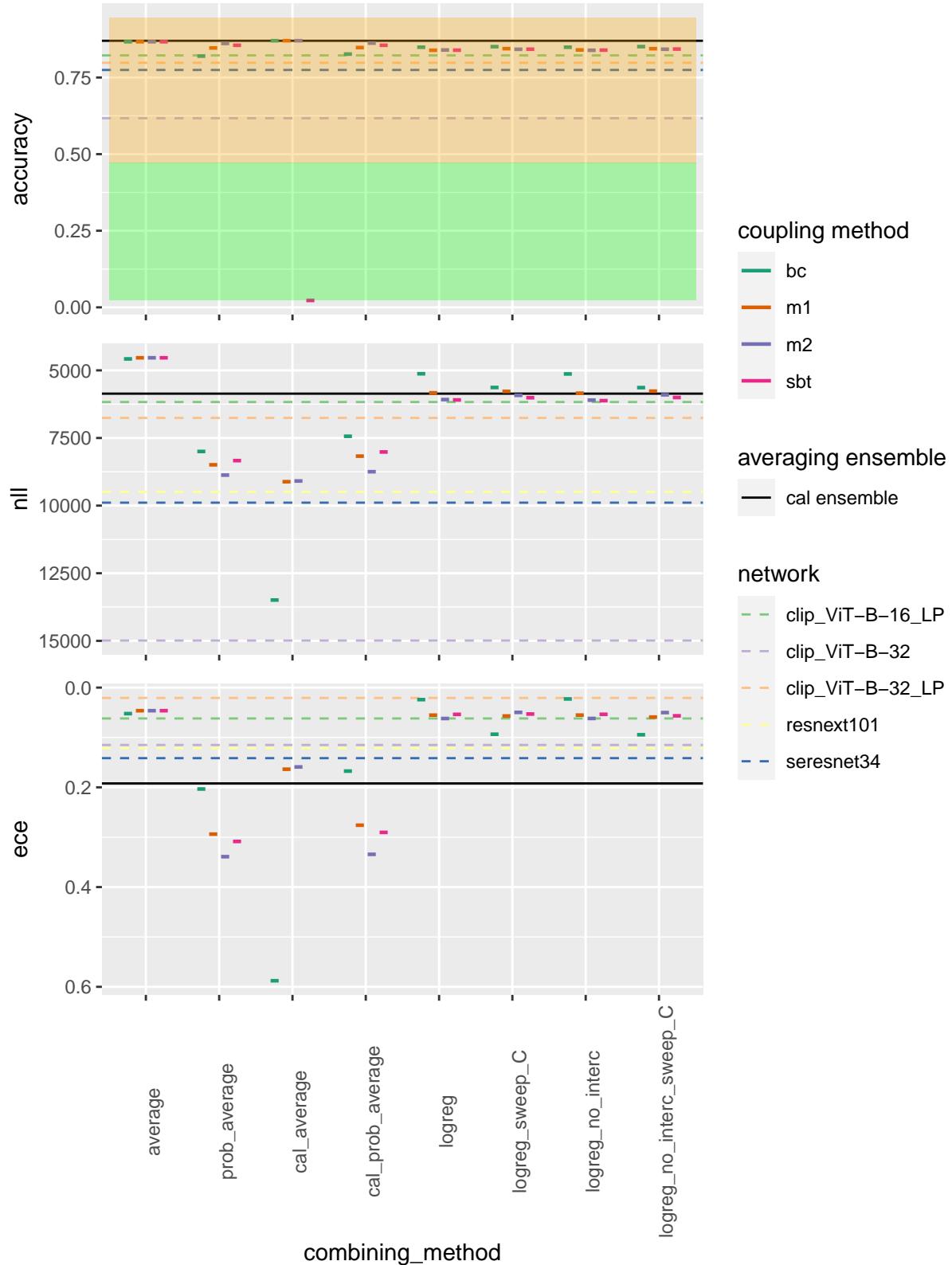
Ensemble metrics  
Error inconsistency 0.513999998569489



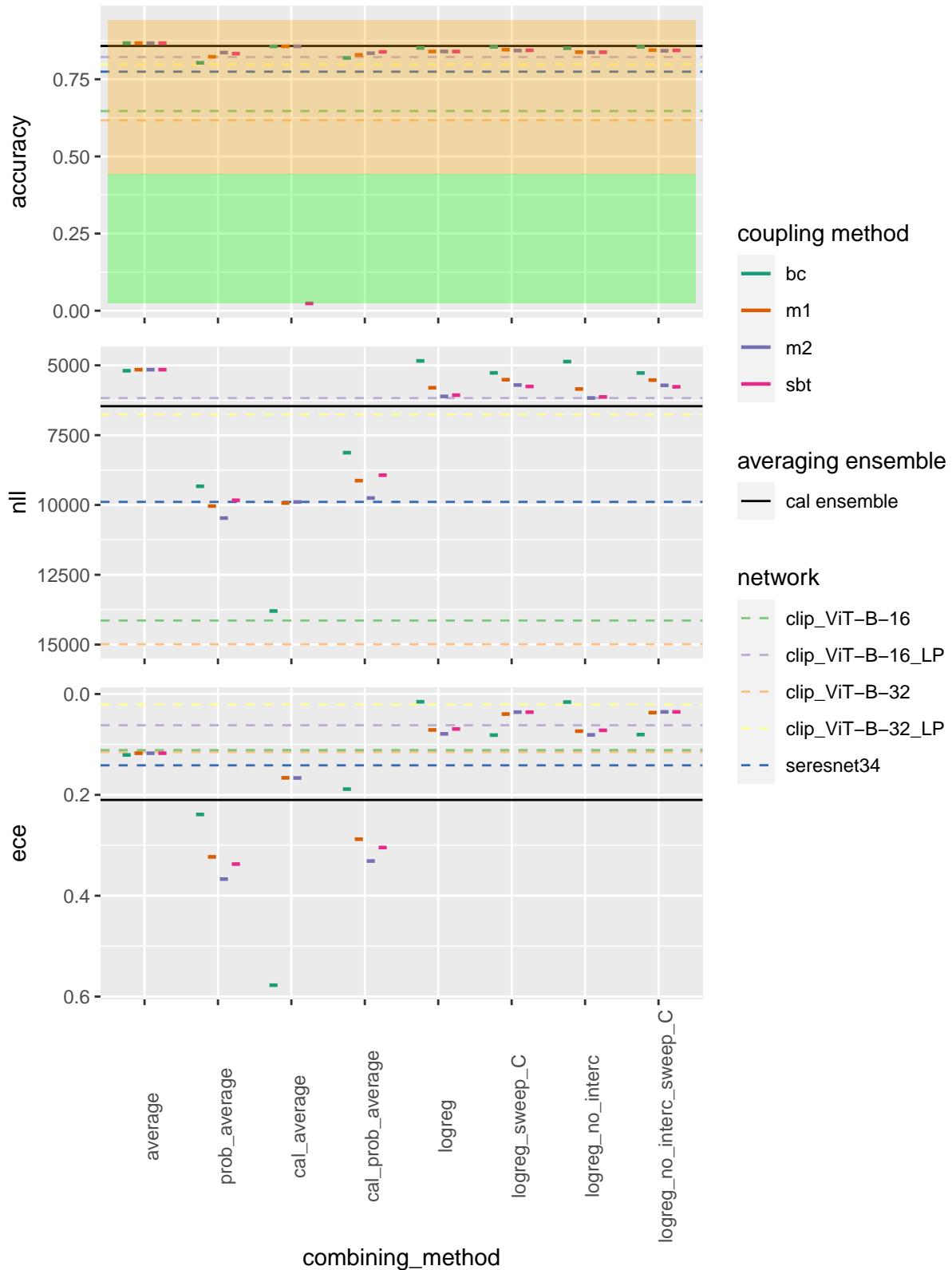
Ensemble metrics  
Error inconsistency 0.497599989175797



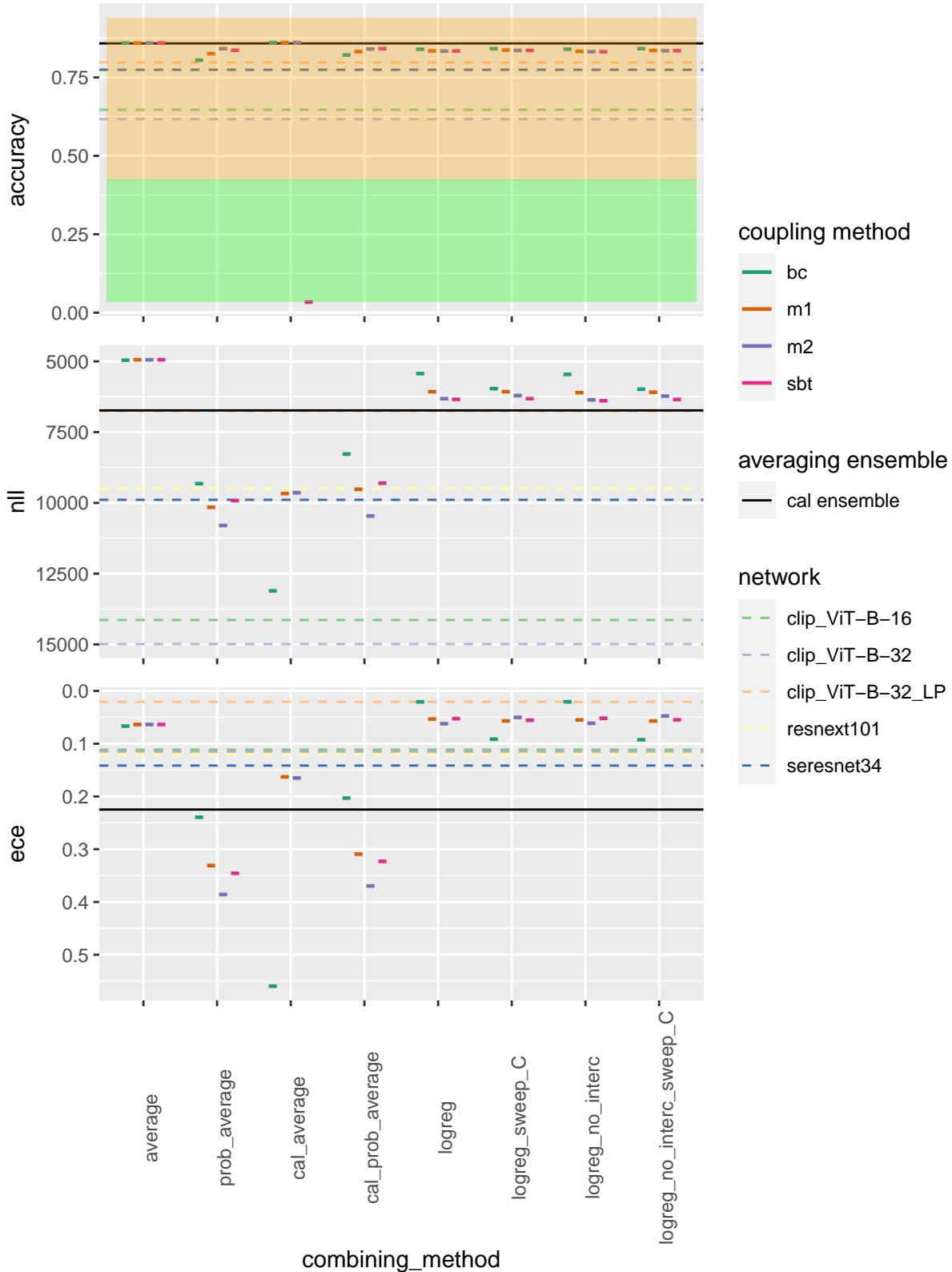
Ensemble metrics  
Error inconsistency 0.473799973726273



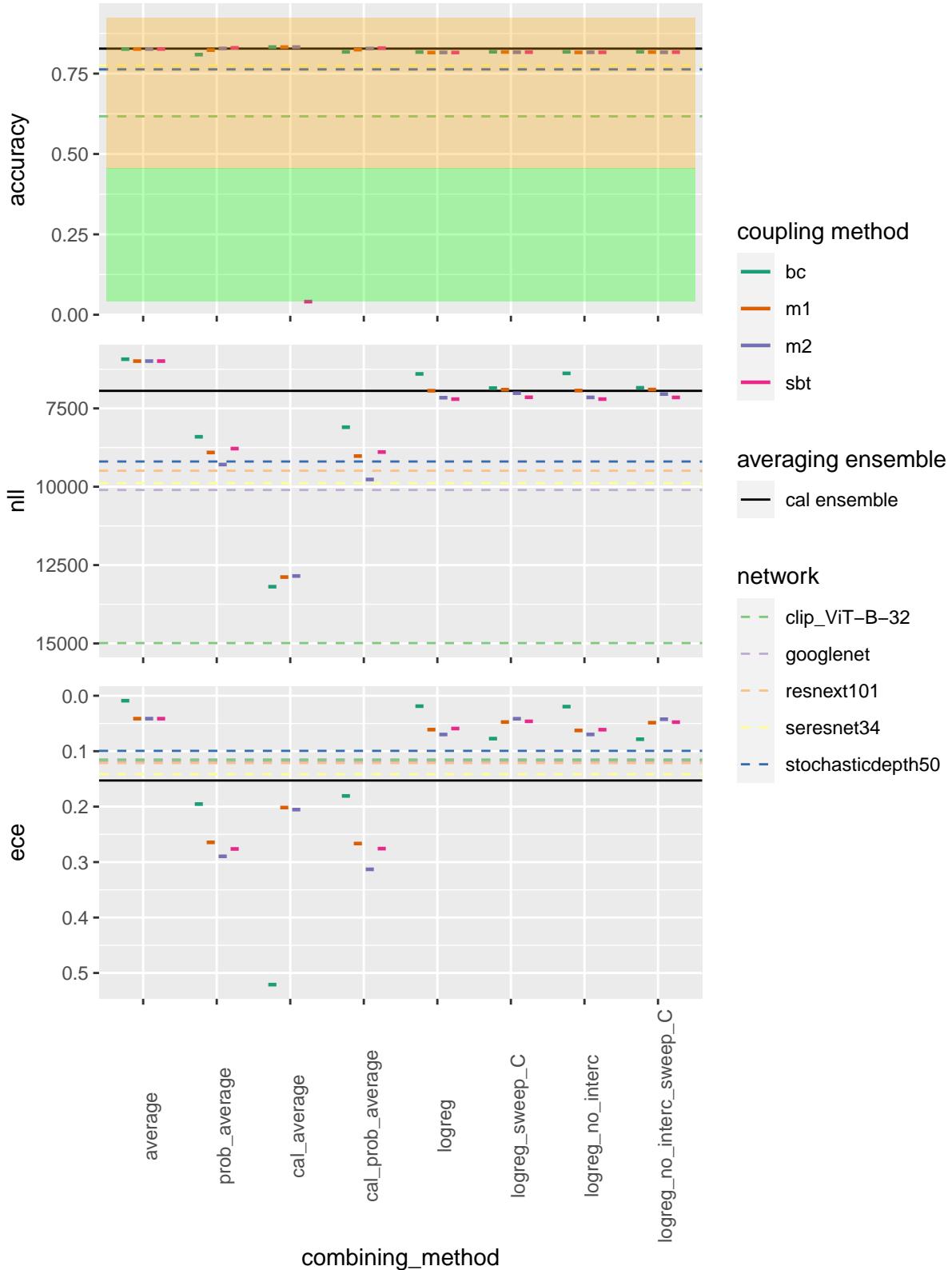
Ensemble metrics  
Error inconsistency 0.501199960708618



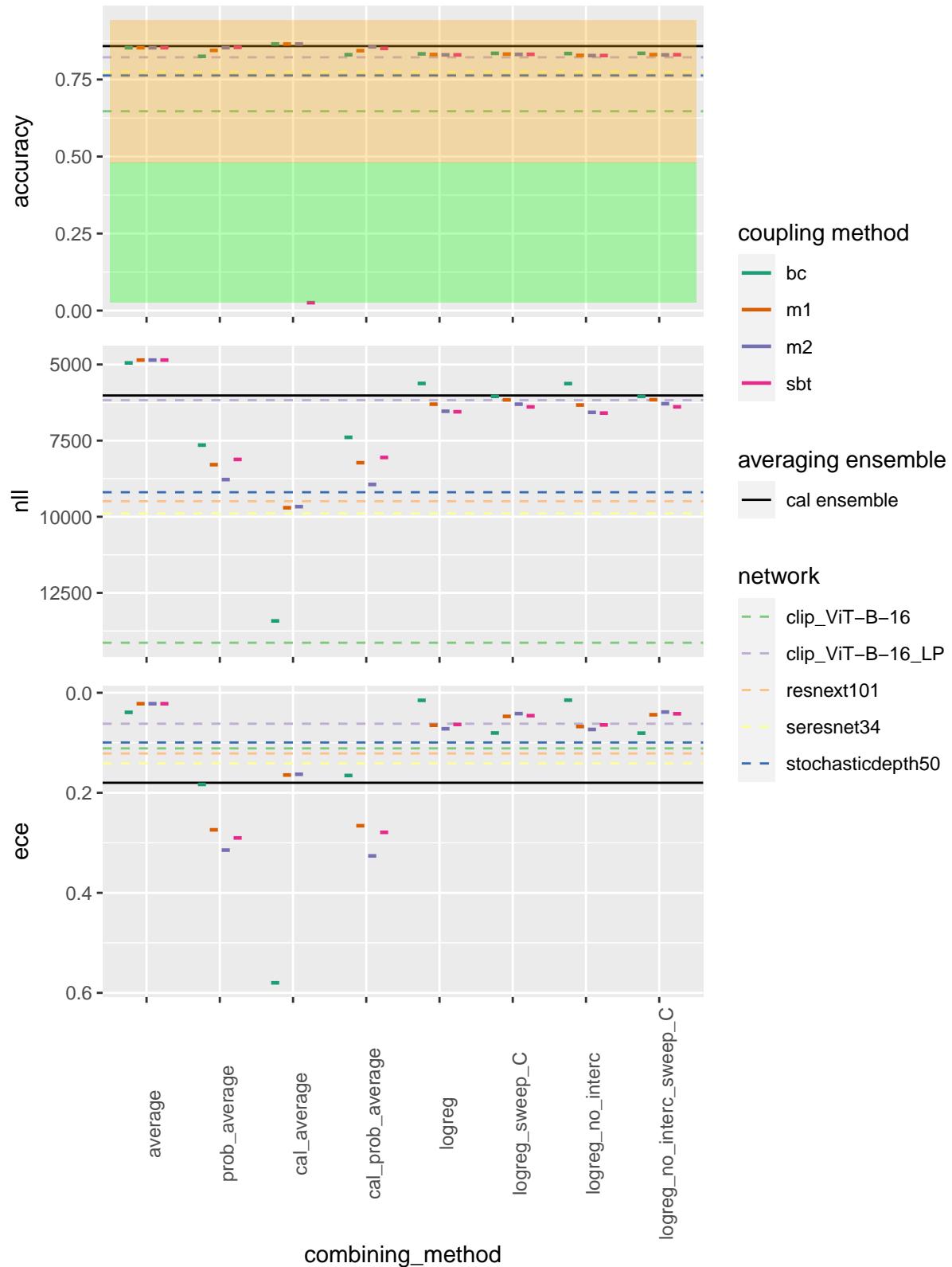
Ensemble metrics  
Error inconsistency 0.516799986362457



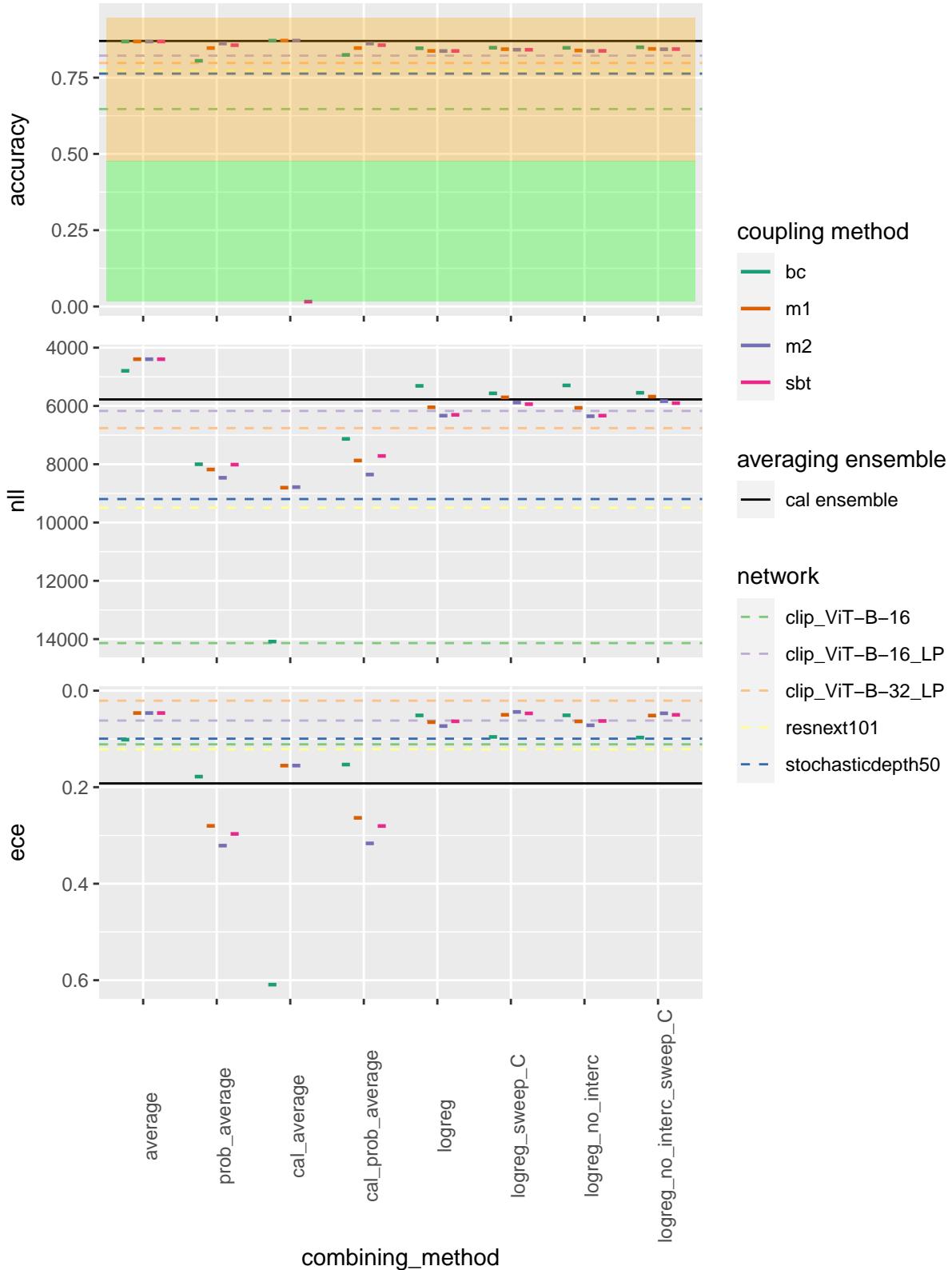
Ensemble metrics  
Error inconsistency 0.468099981546402



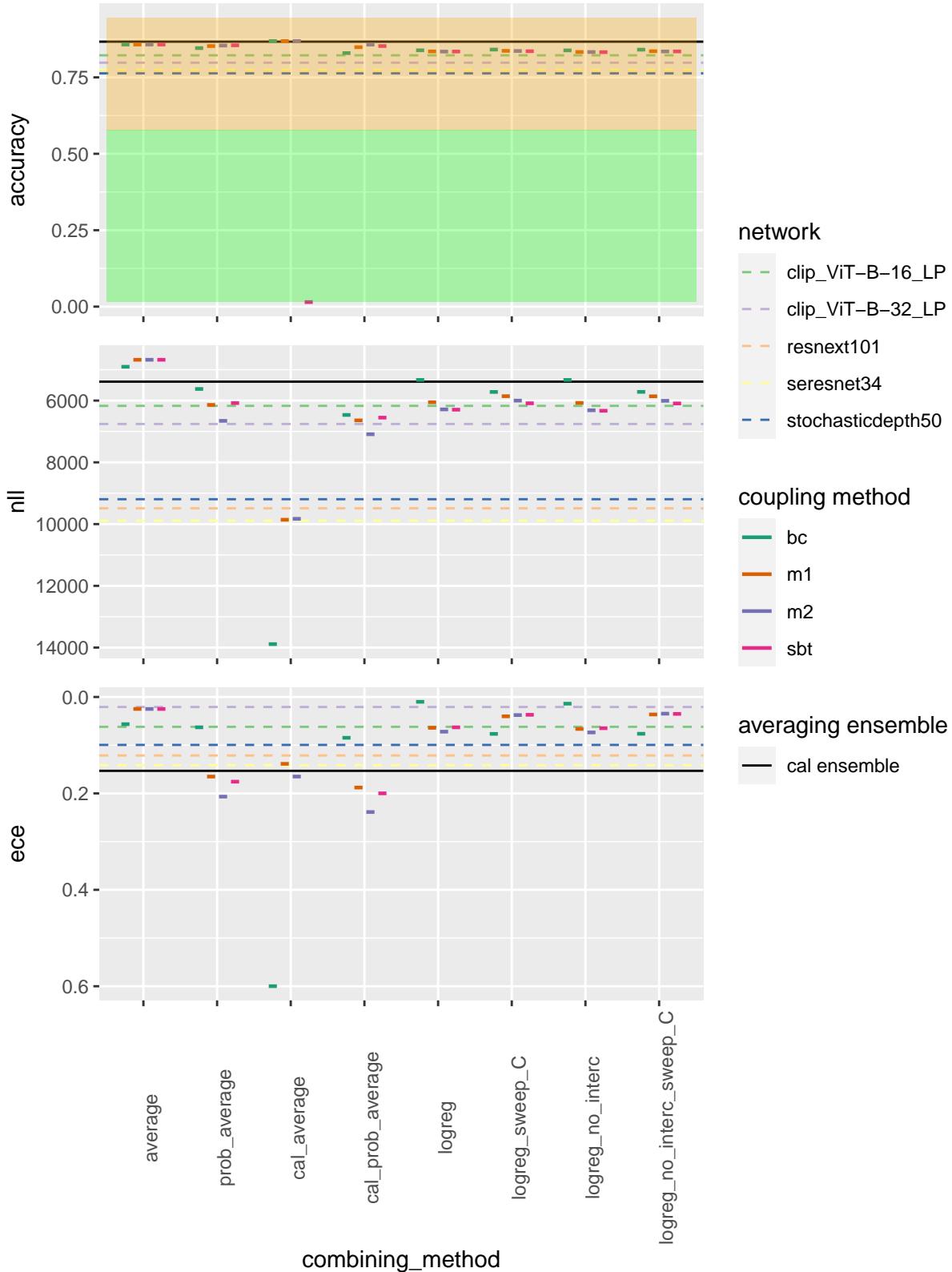
Ensemble metrics  
Error inconsistency 0.46289998292923



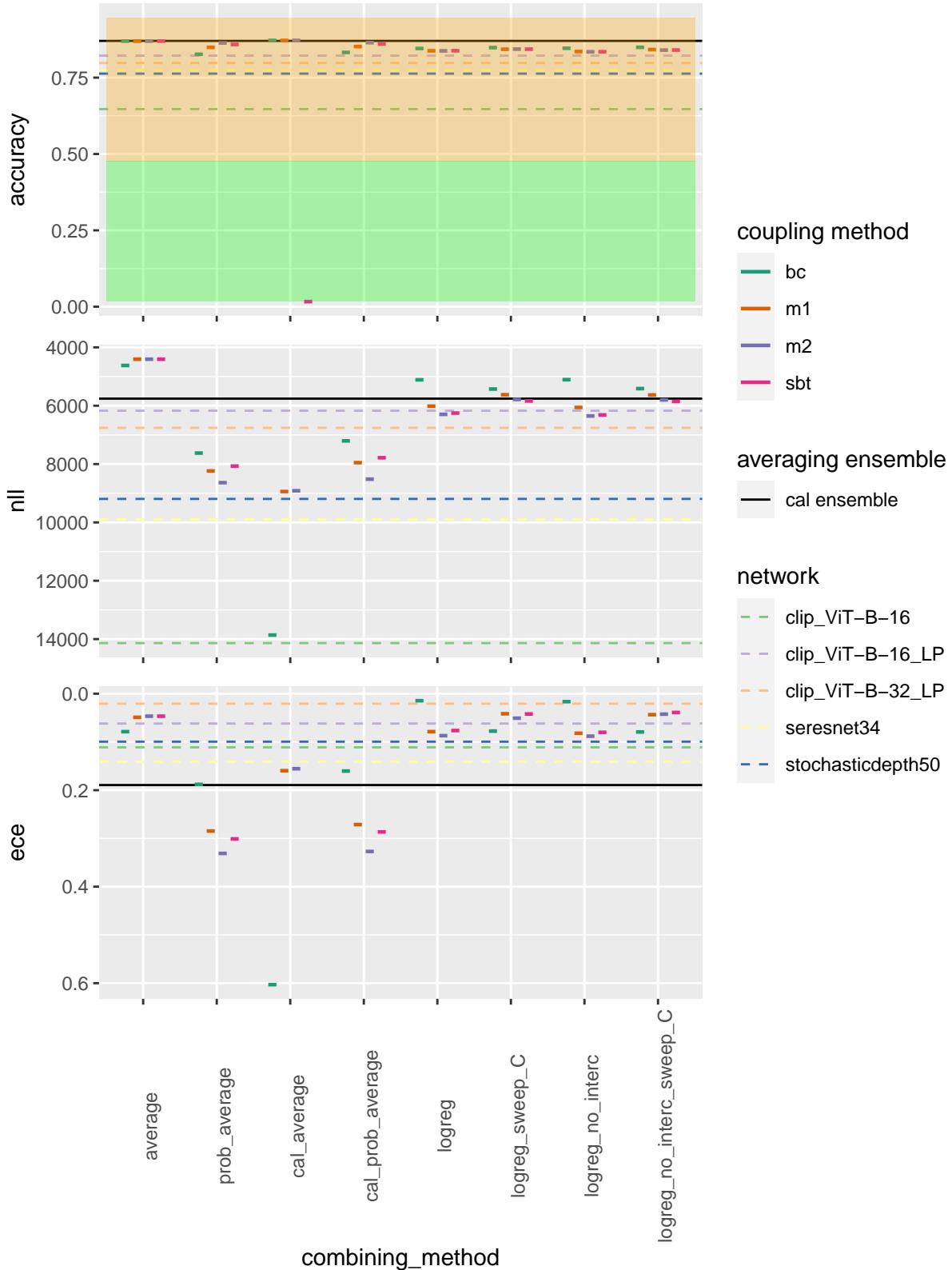
Ensemble metrics  
Error inconsistency 0.469300001859665



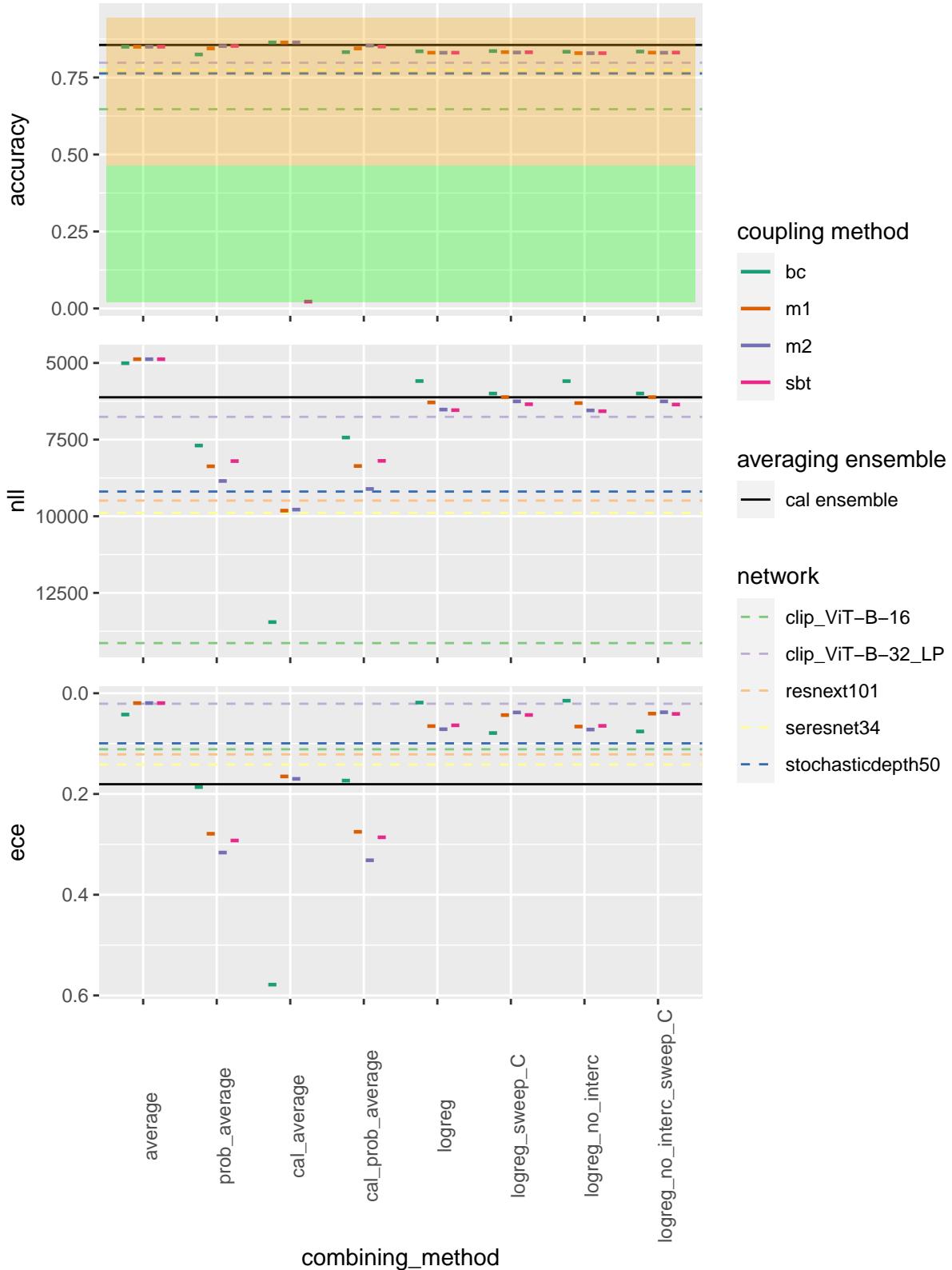
Ensemble metrics  
Error inconsistency 0.368099987506866



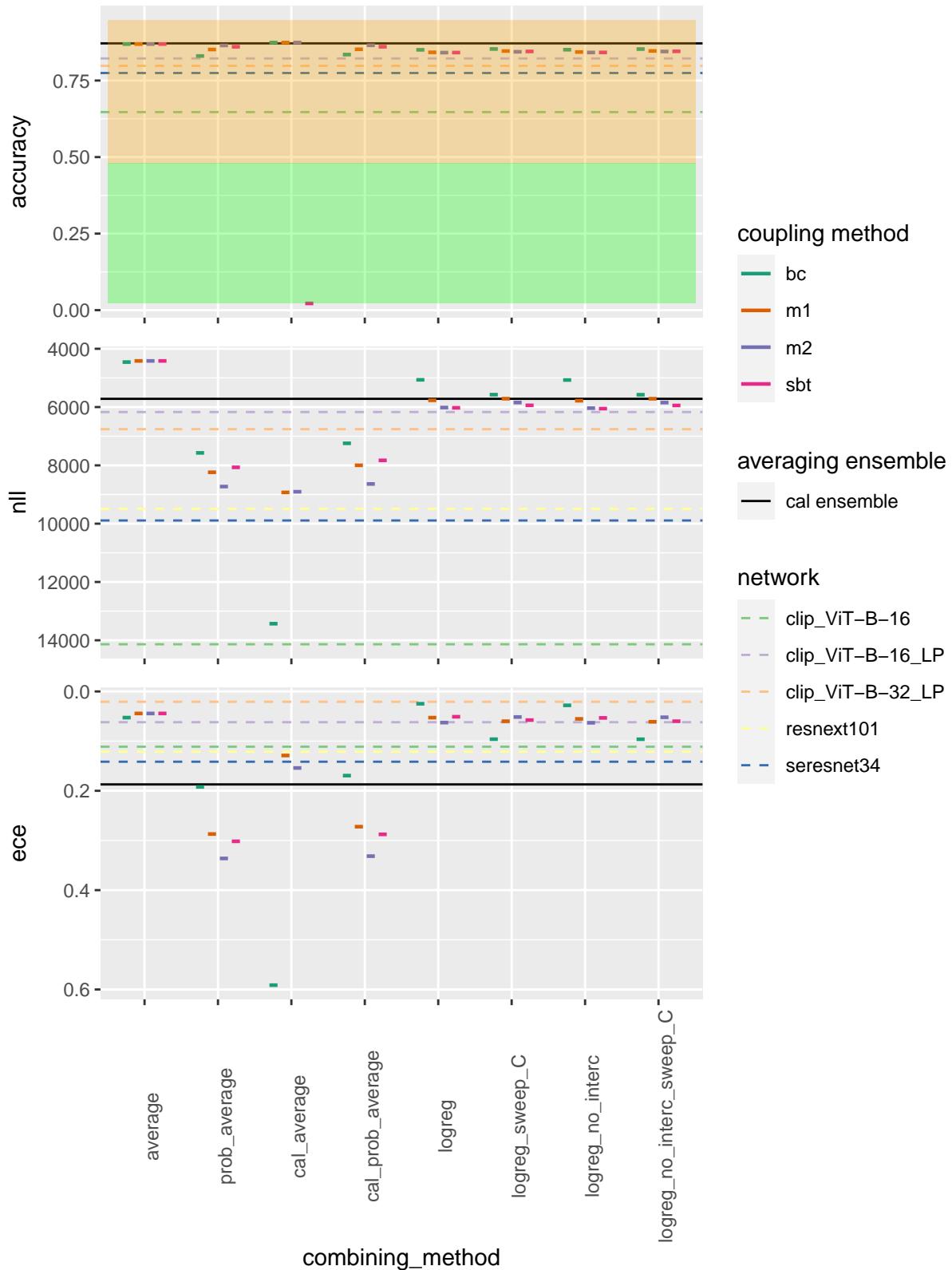
Ensemble metrics  
Error inconsistency 0.469999998807907



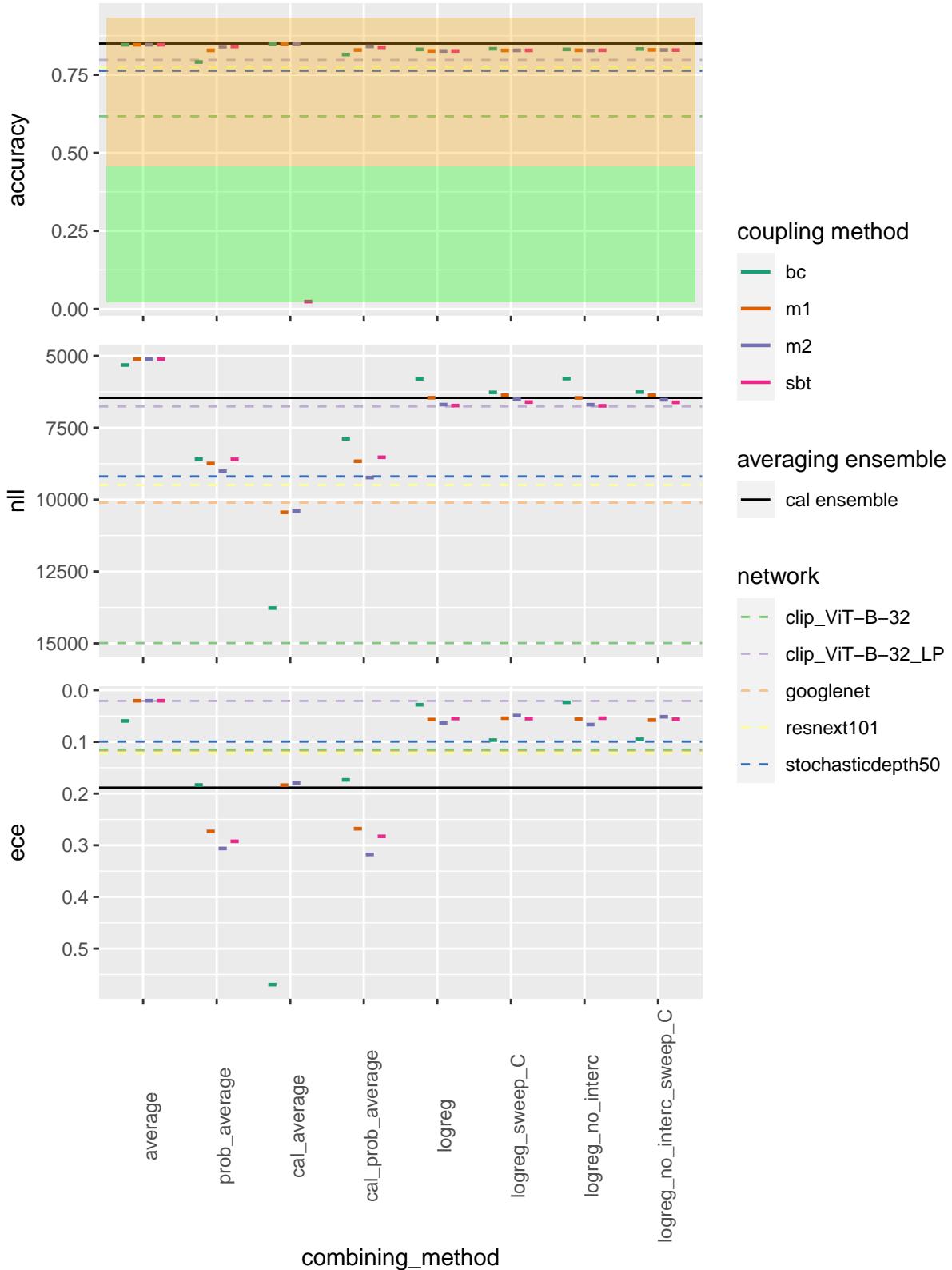
Ensemble metrics  
Error inconsistency 0.480299979448318



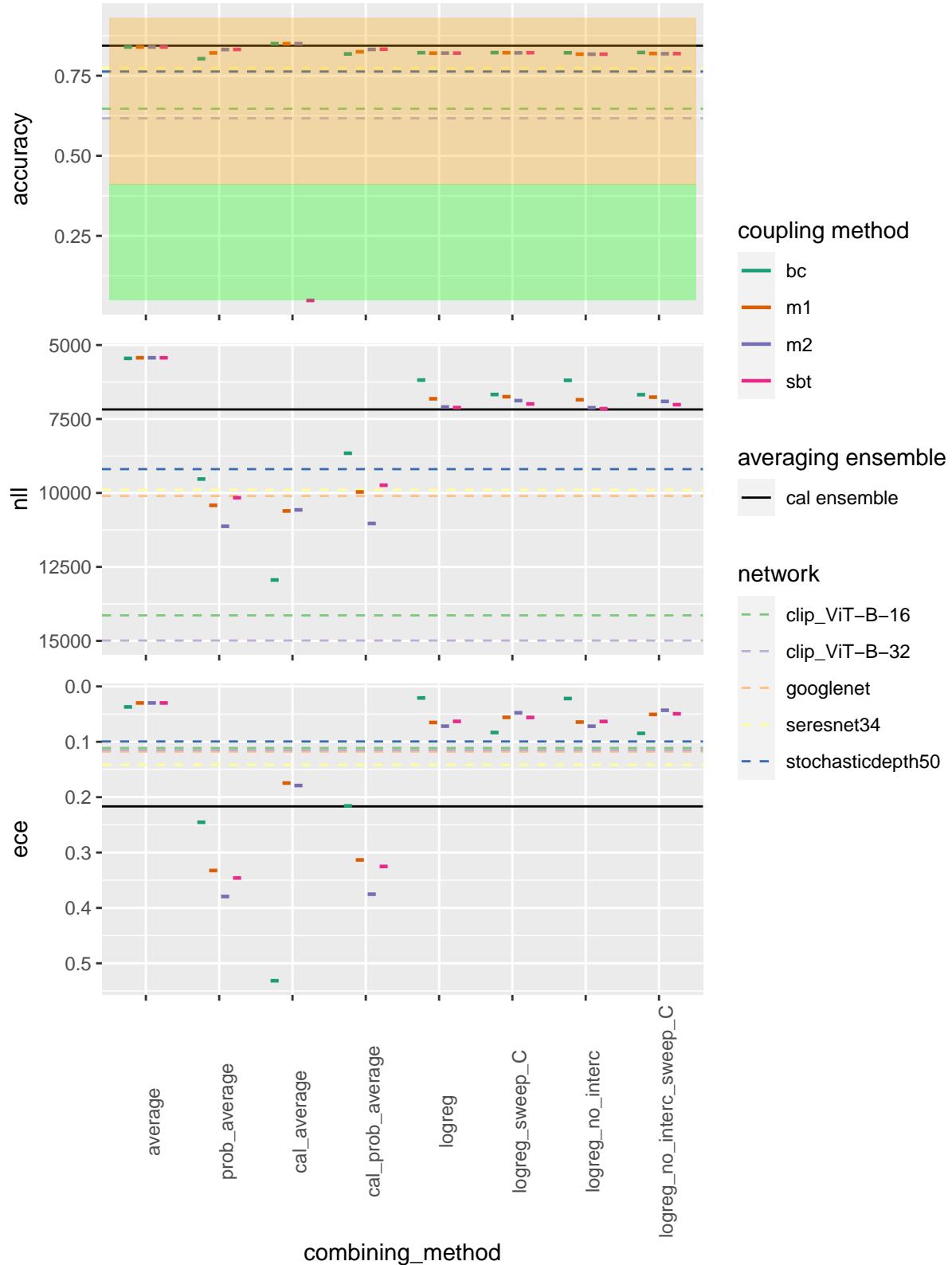
Ensemble metrics  
Error inconsistency 0.467299997806549



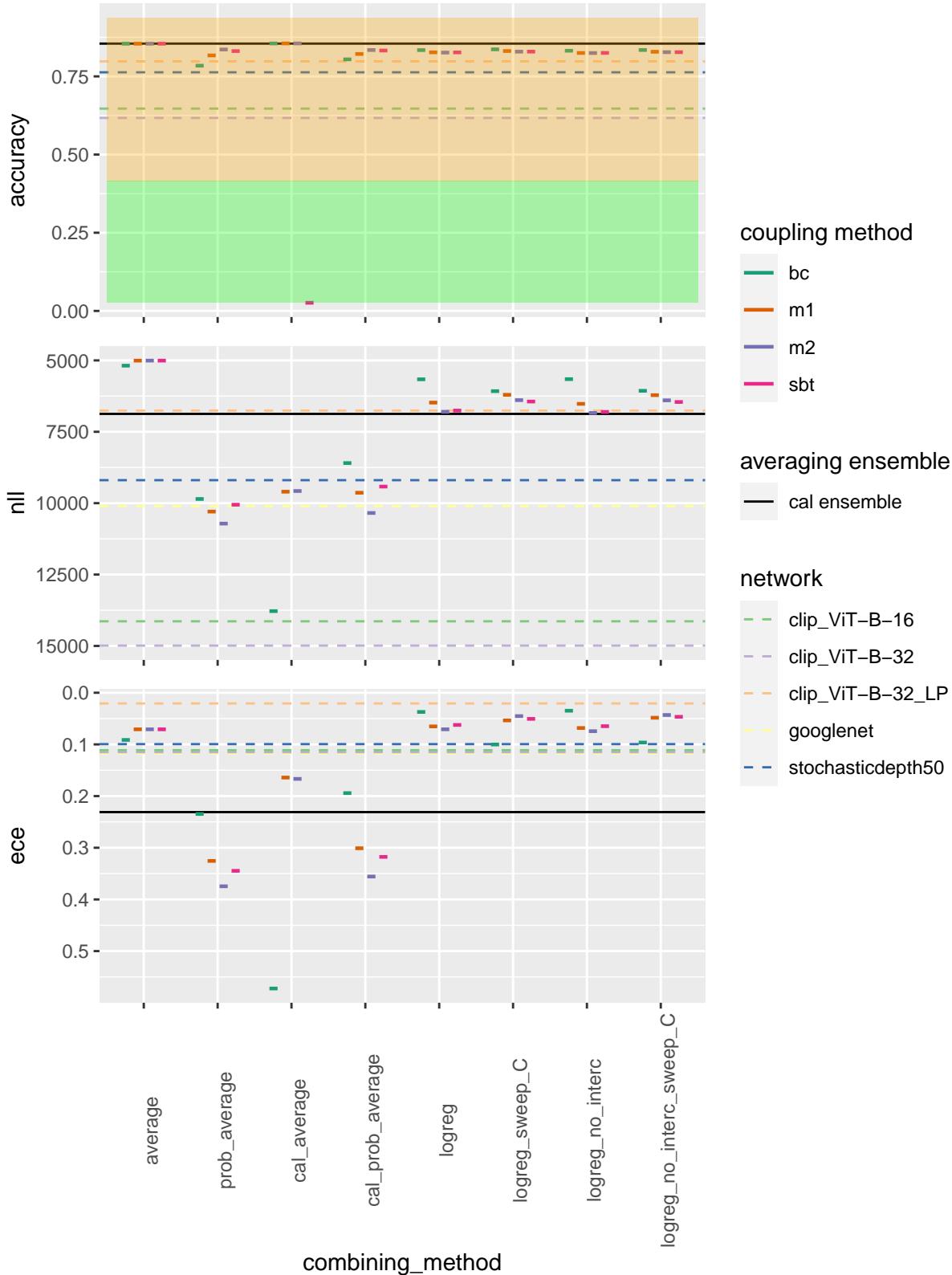
Ensemble metrics  
Error inconsistency 0.476300001144409



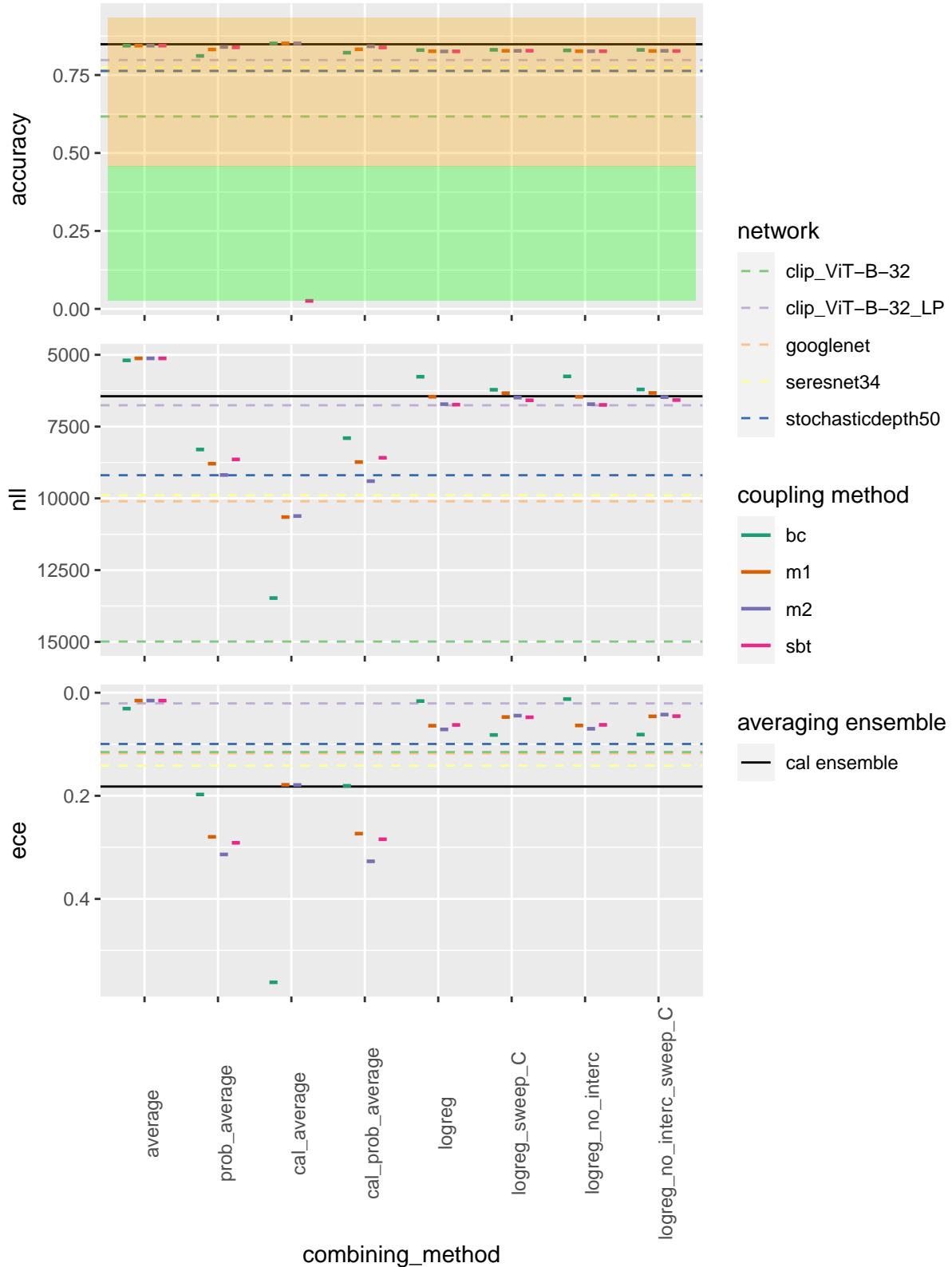
Ensemble metrics  
Error inconsistency 0.521200001239777



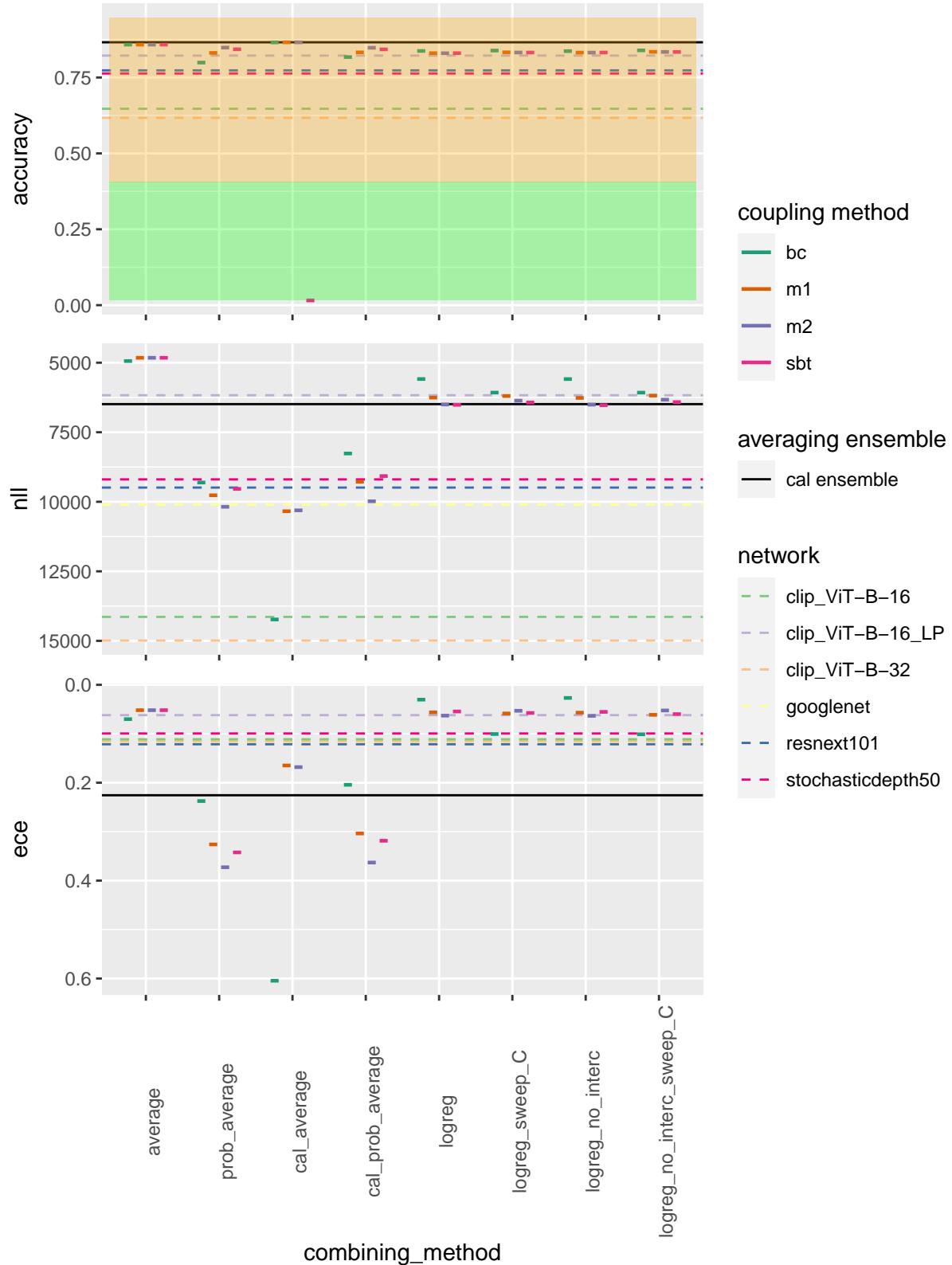
Ensemble metrics  
Error inconsistency 0.521600008010864



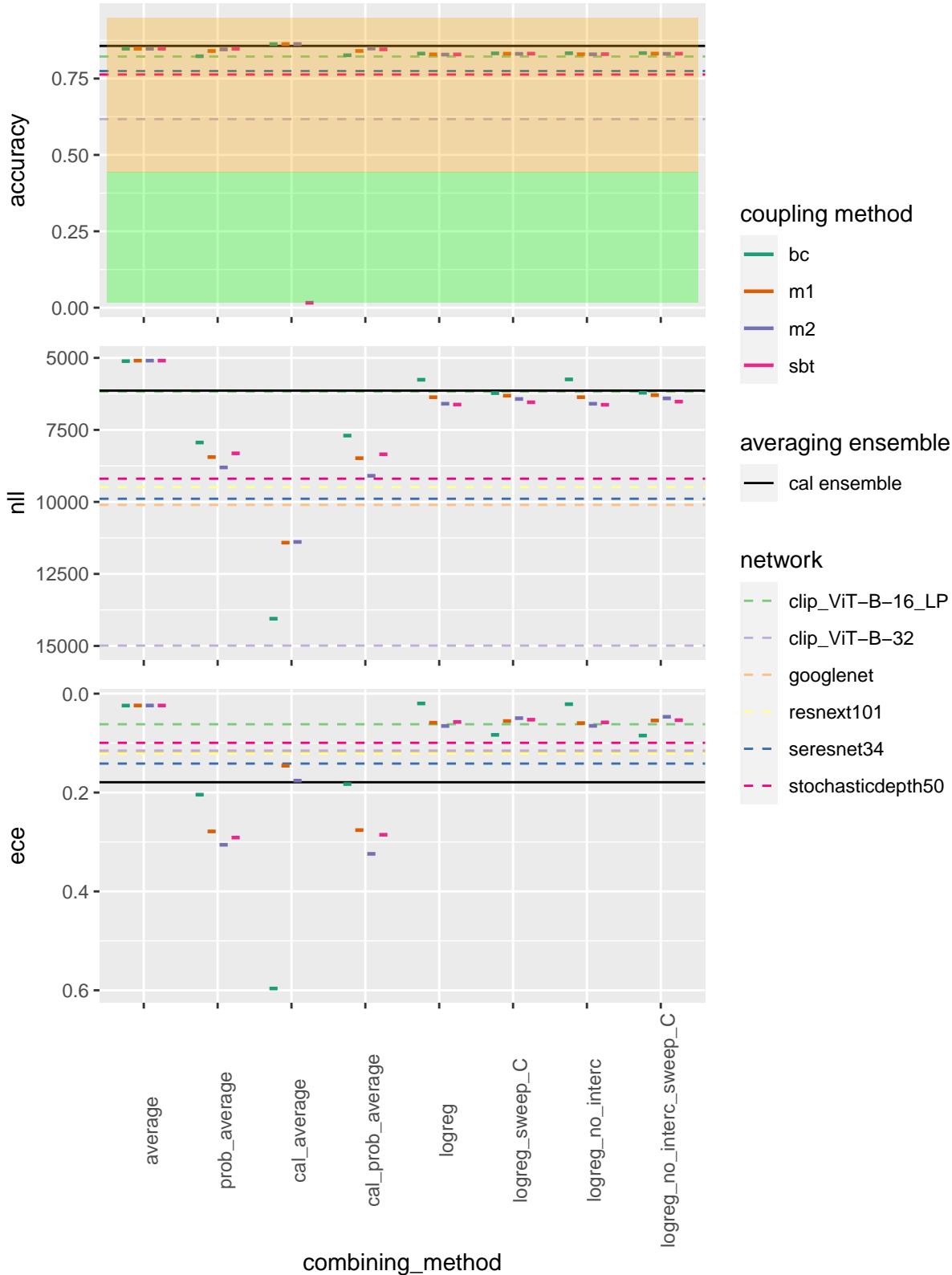
Ensemble metrics  
Error inconsistency 0.476199984550476



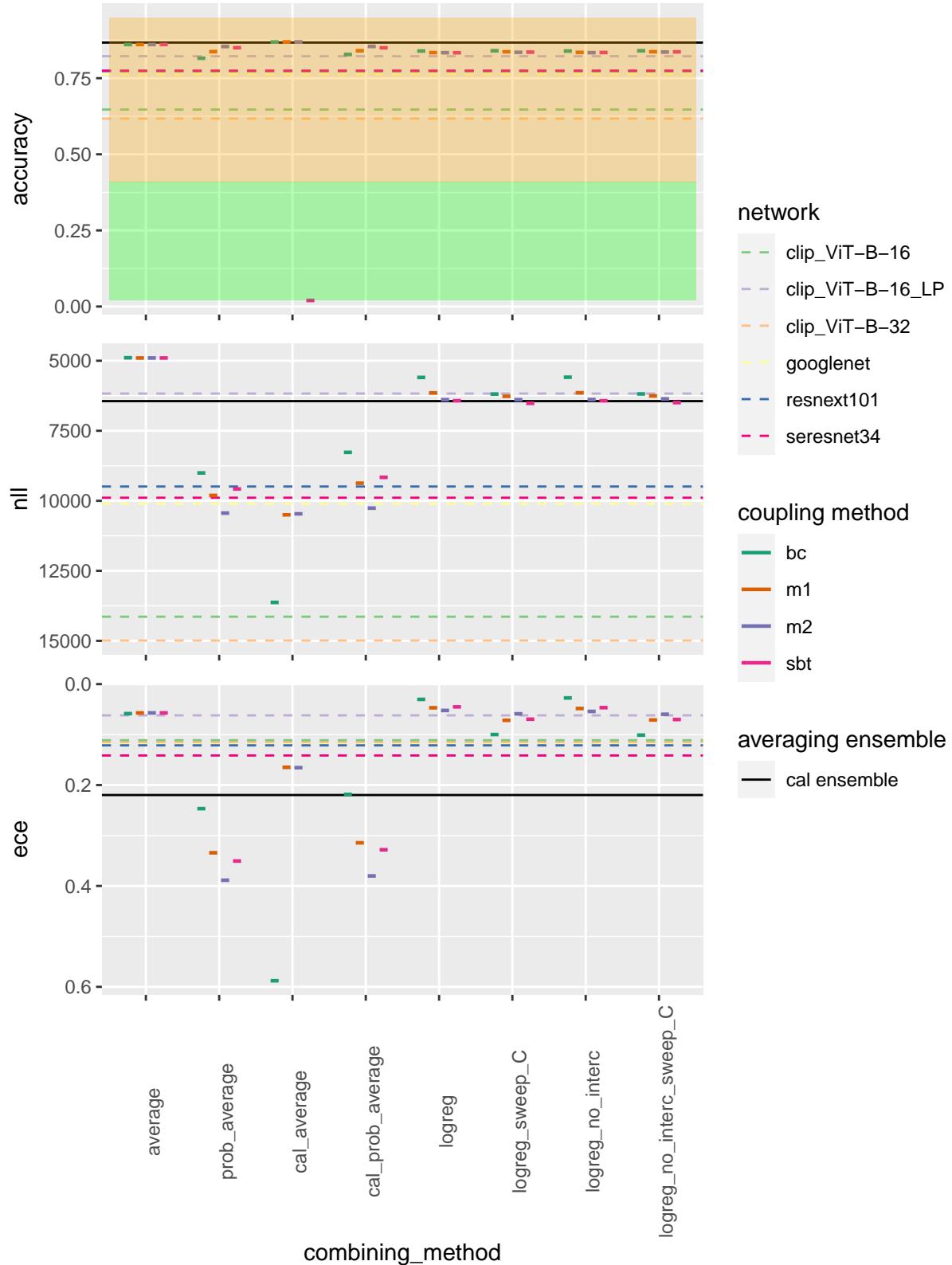
Ensemble metrics  
Error inconsistency 0.541799962520599



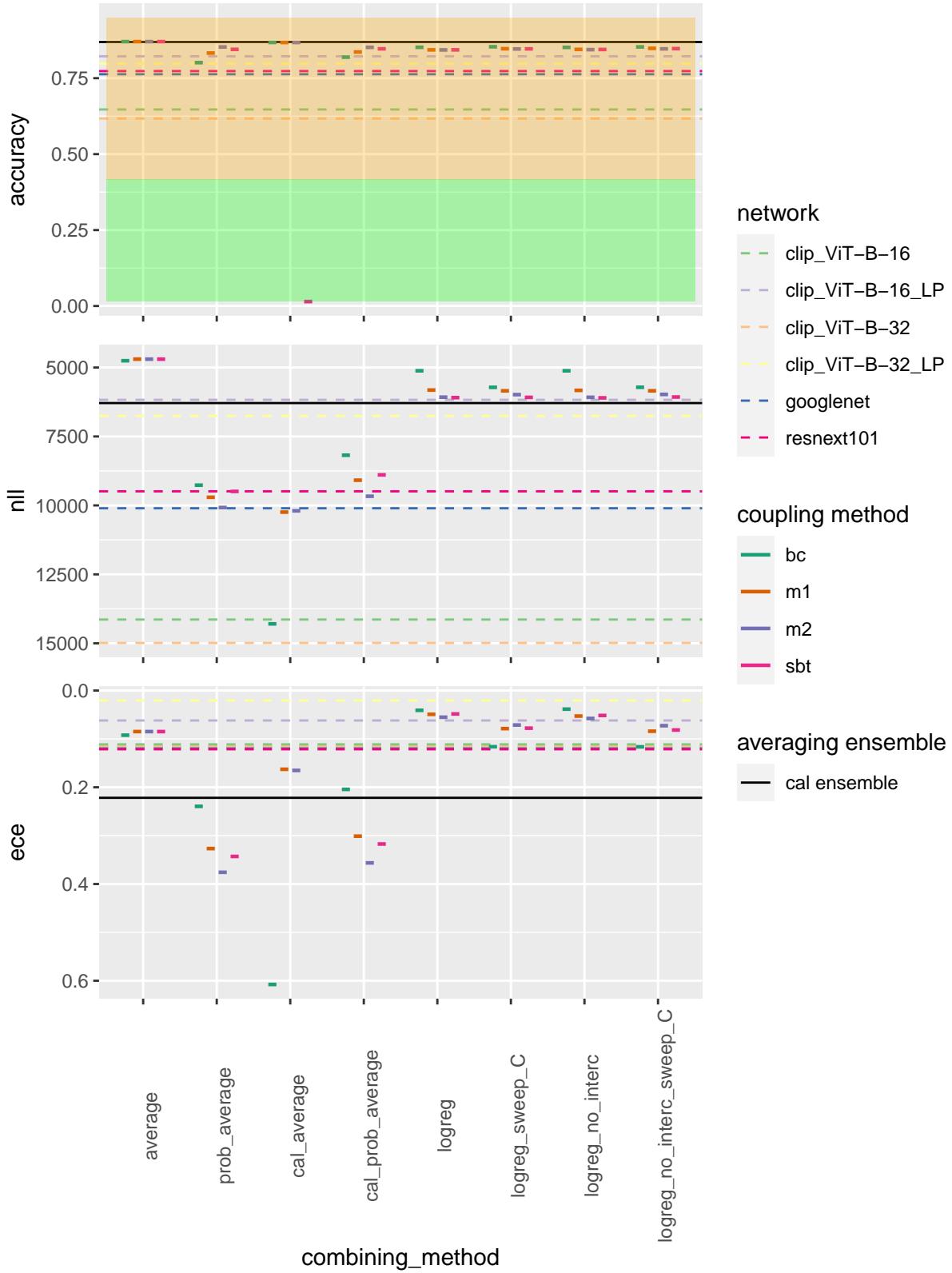
Ensemble metrics  
Error inconsistency 0.506099998950958



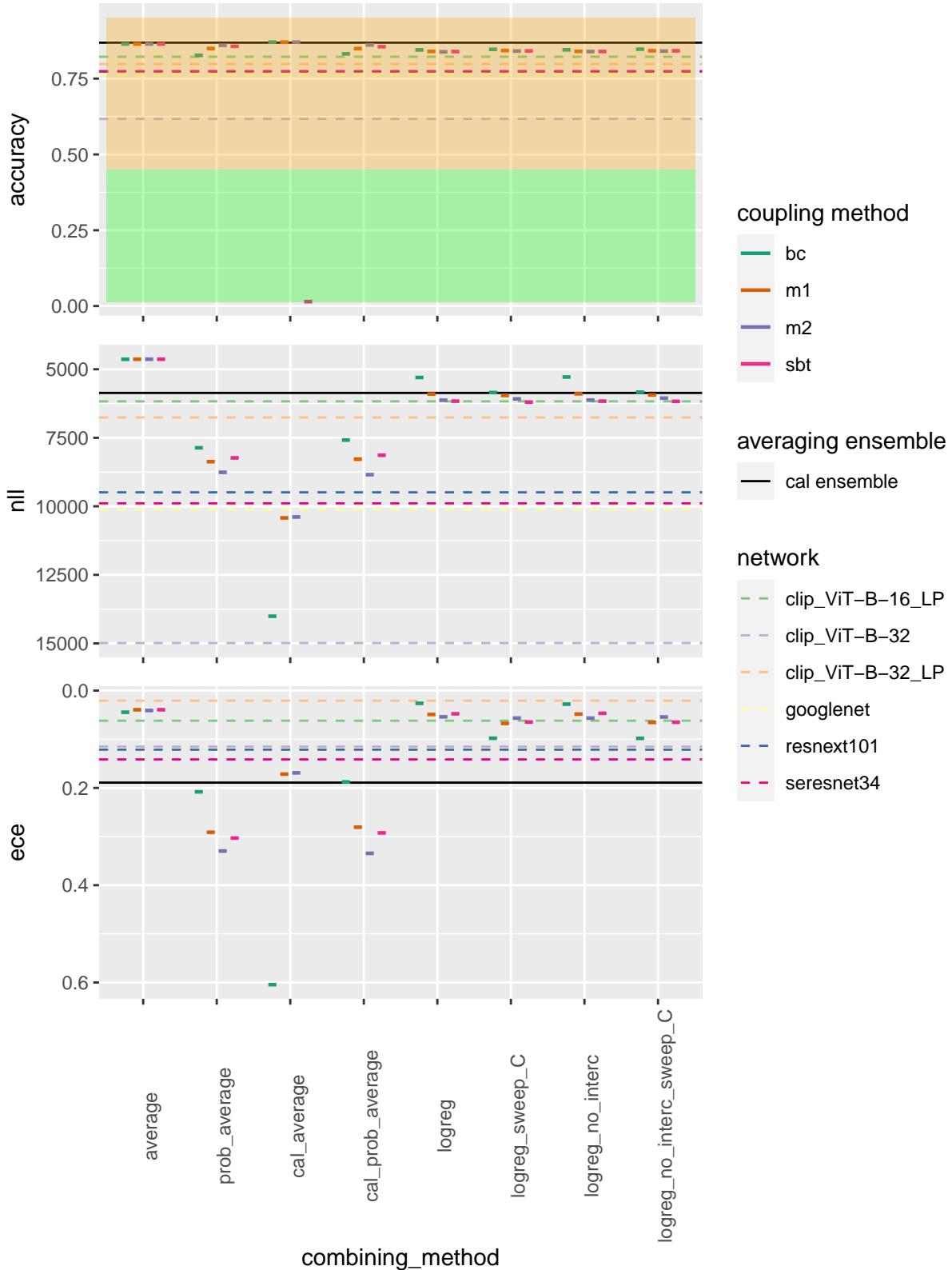
Ensemble metrics  
Error inconsistency 0.539799988269806



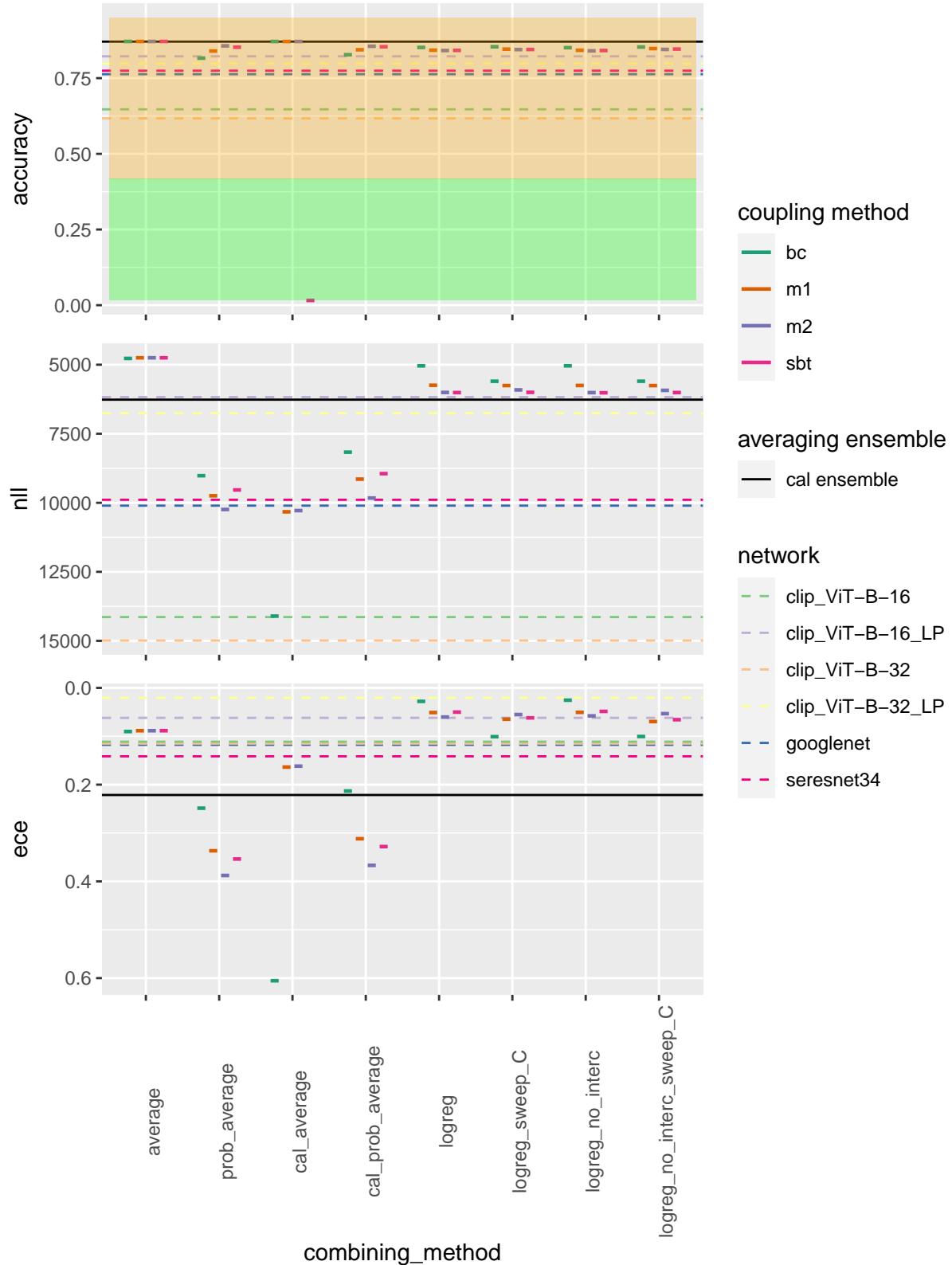
Ensemble metrics  
Error inconsistency 0.532000005245209



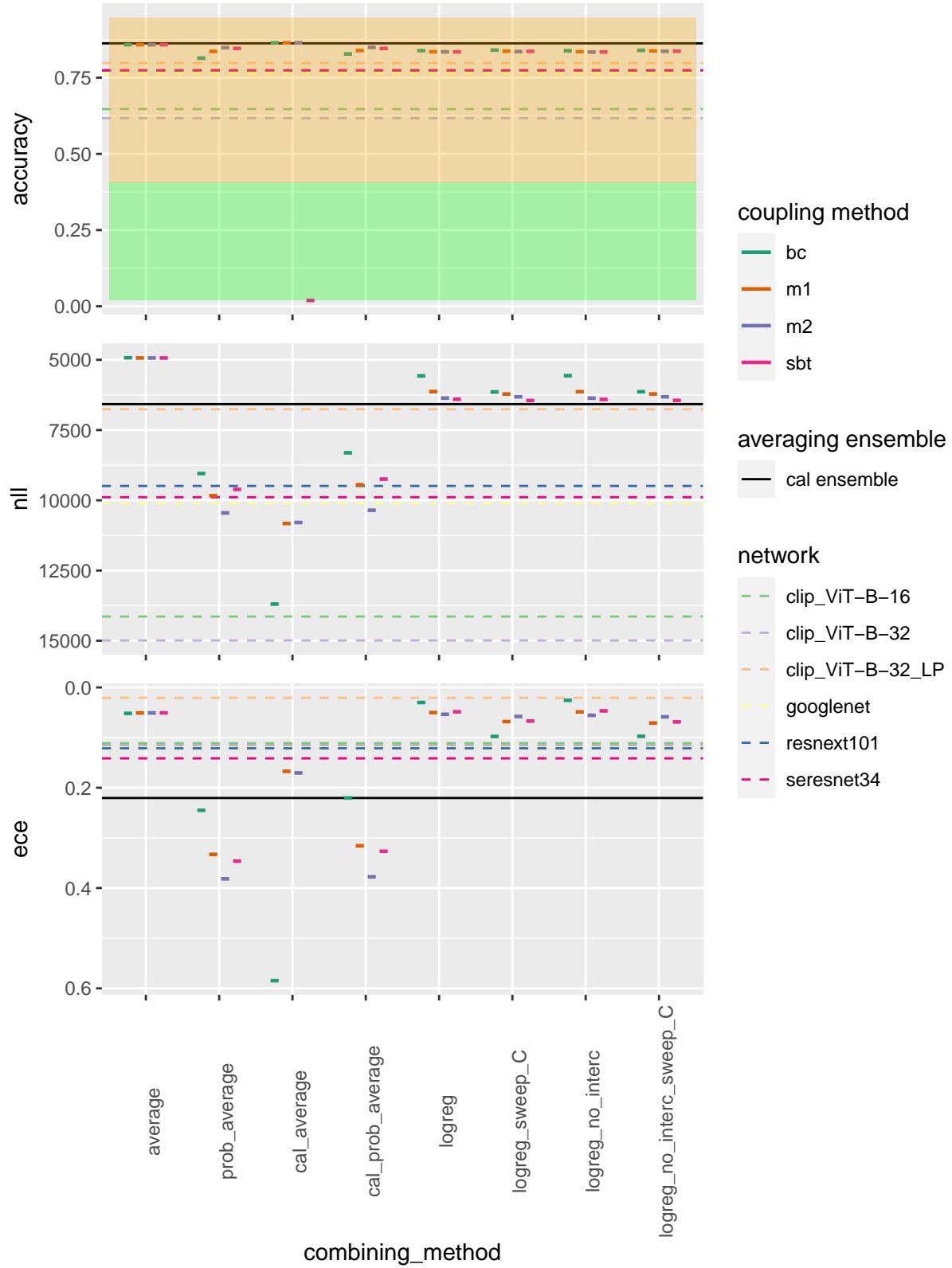
Ensemble metrics  
Error inconsistency 0.500400006771088



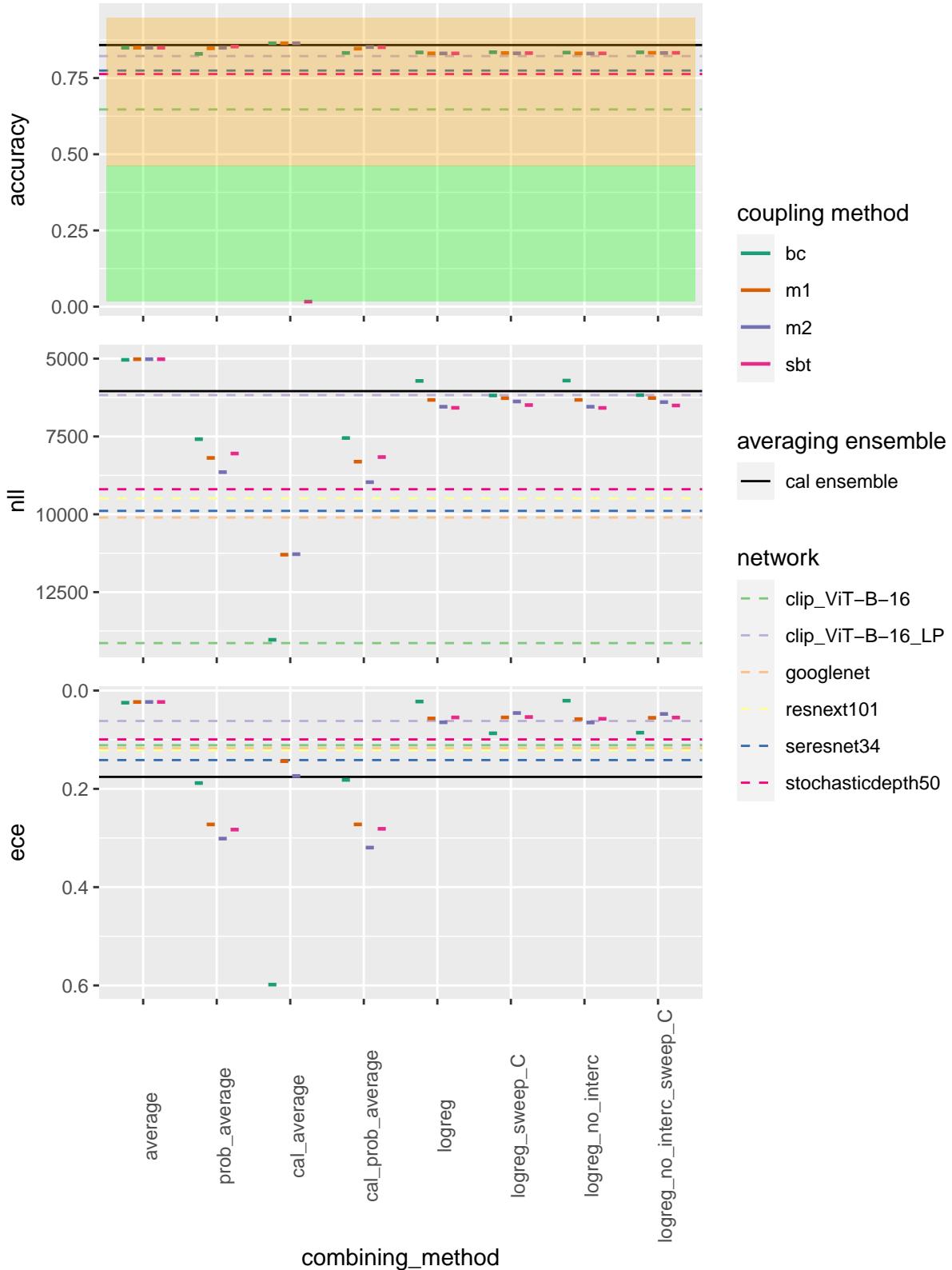
Ensemble metrics  
Error inconsistency 0.533199965953827



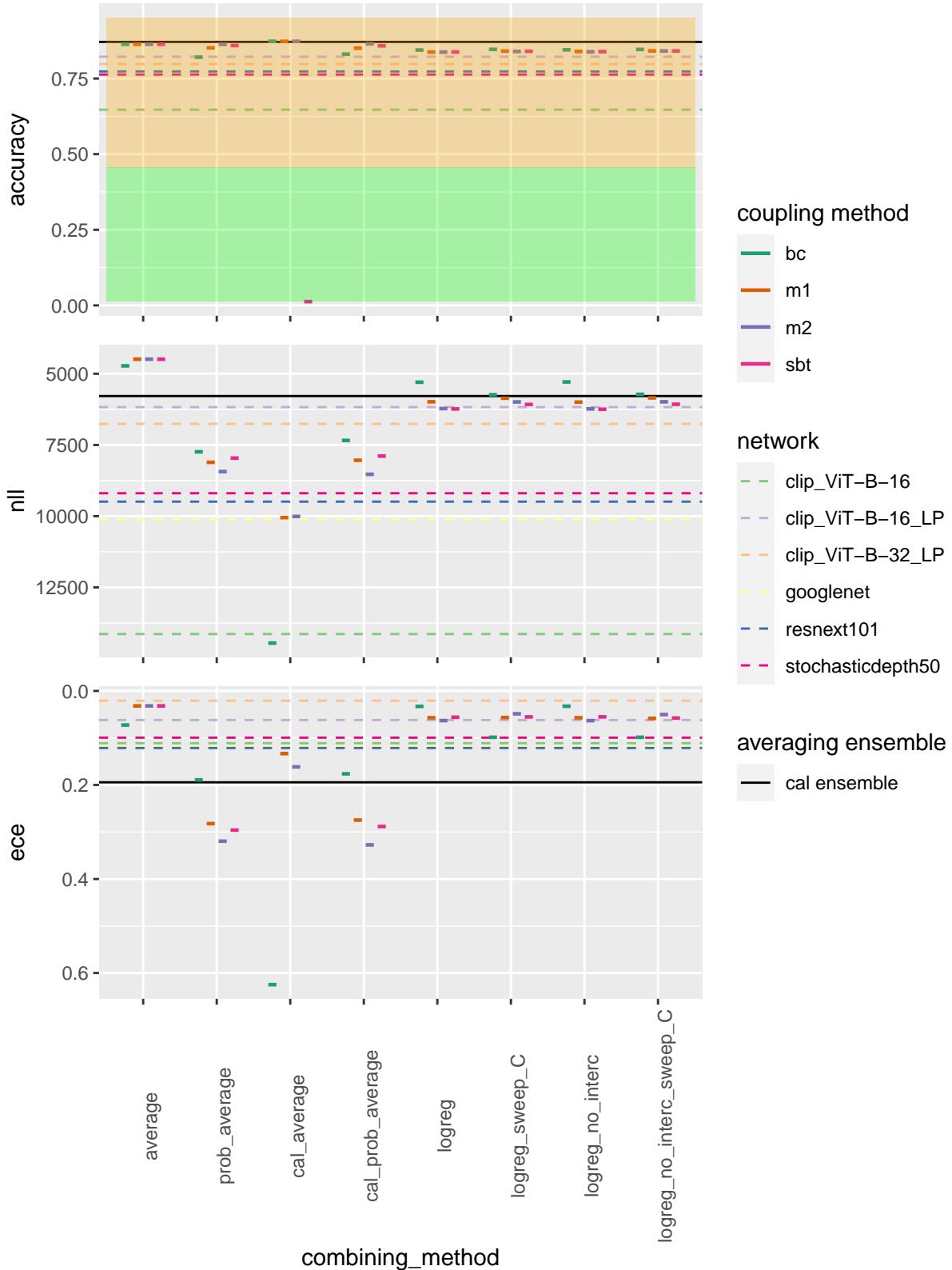
Ensemble metrics  
Error inconsistency 0.541400015354156



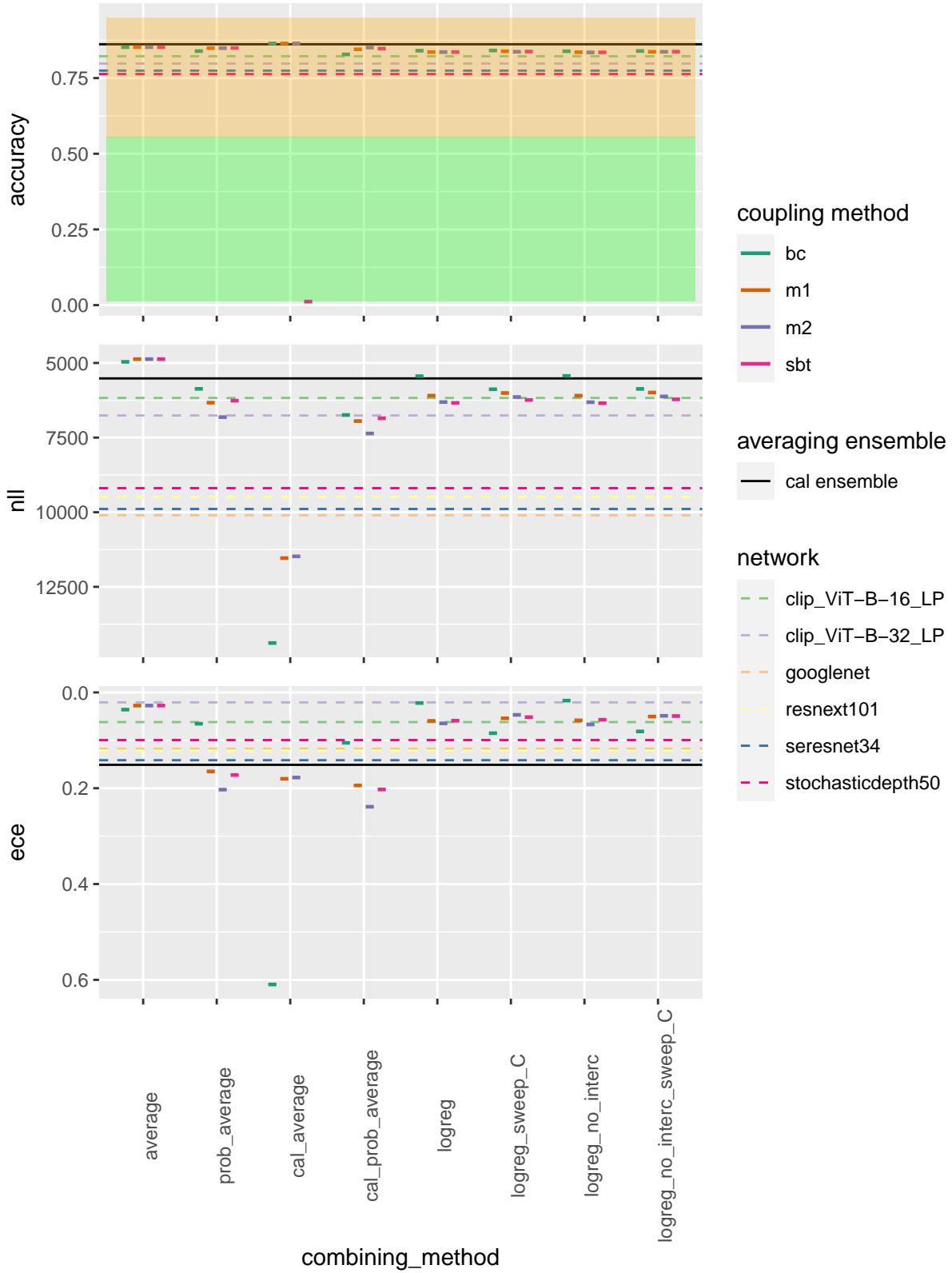
Ensemble metrics  
Error inconsistency 0.486699998378754



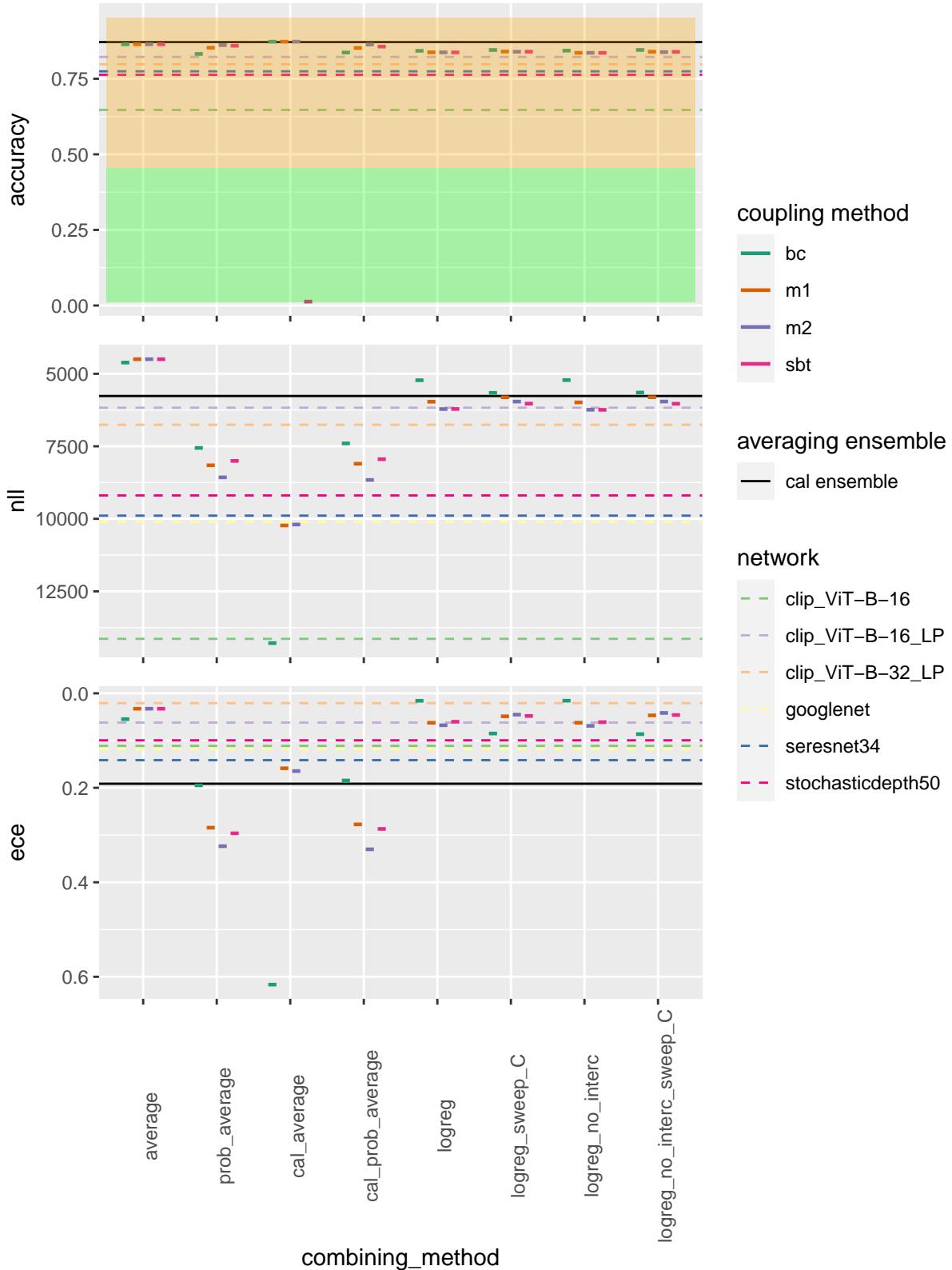
Ensemble metrics  
Error inconsistency 0.496099978685379



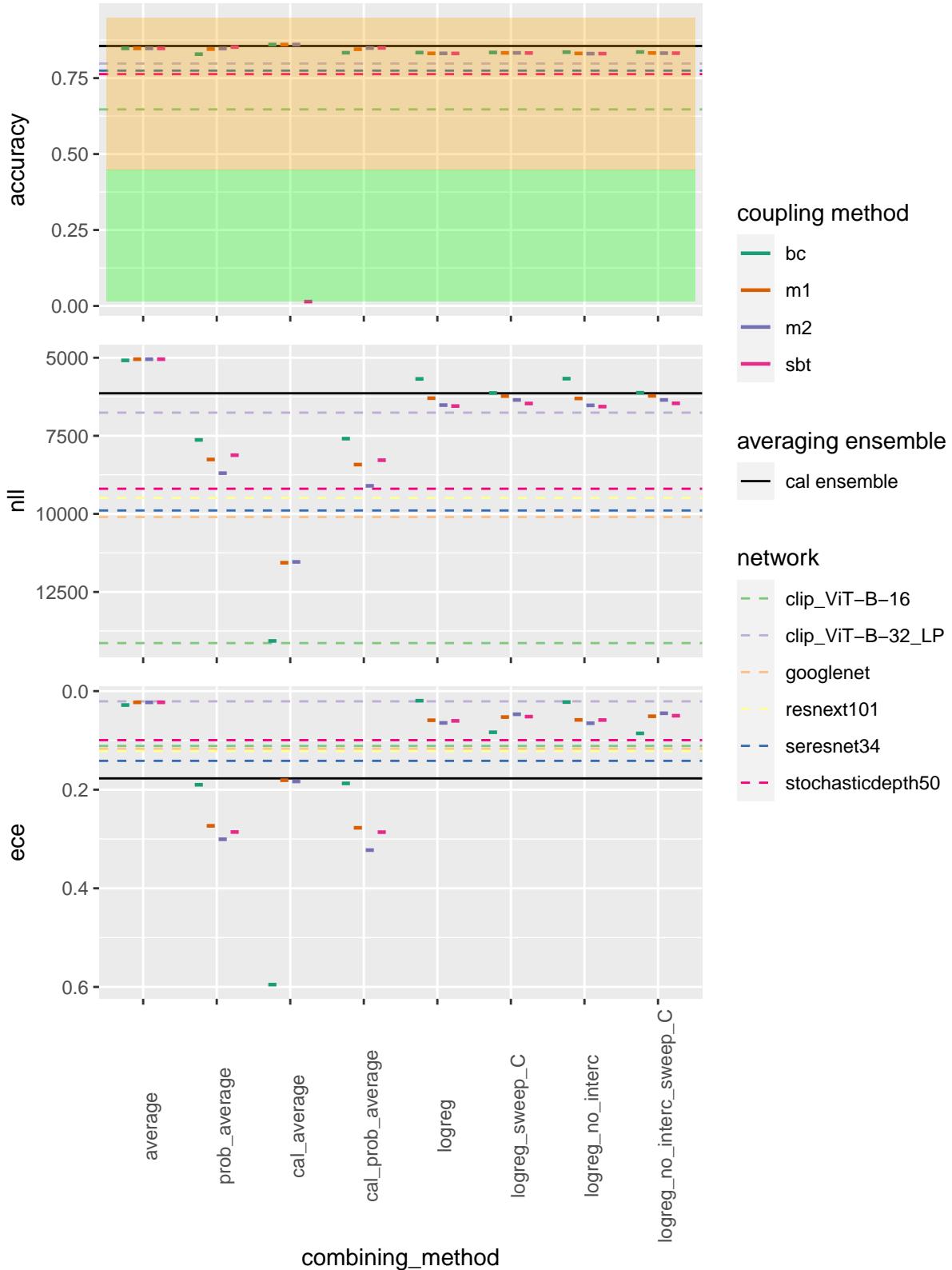
Ensemble metrics  
Error inconsistency 0.394899994134903



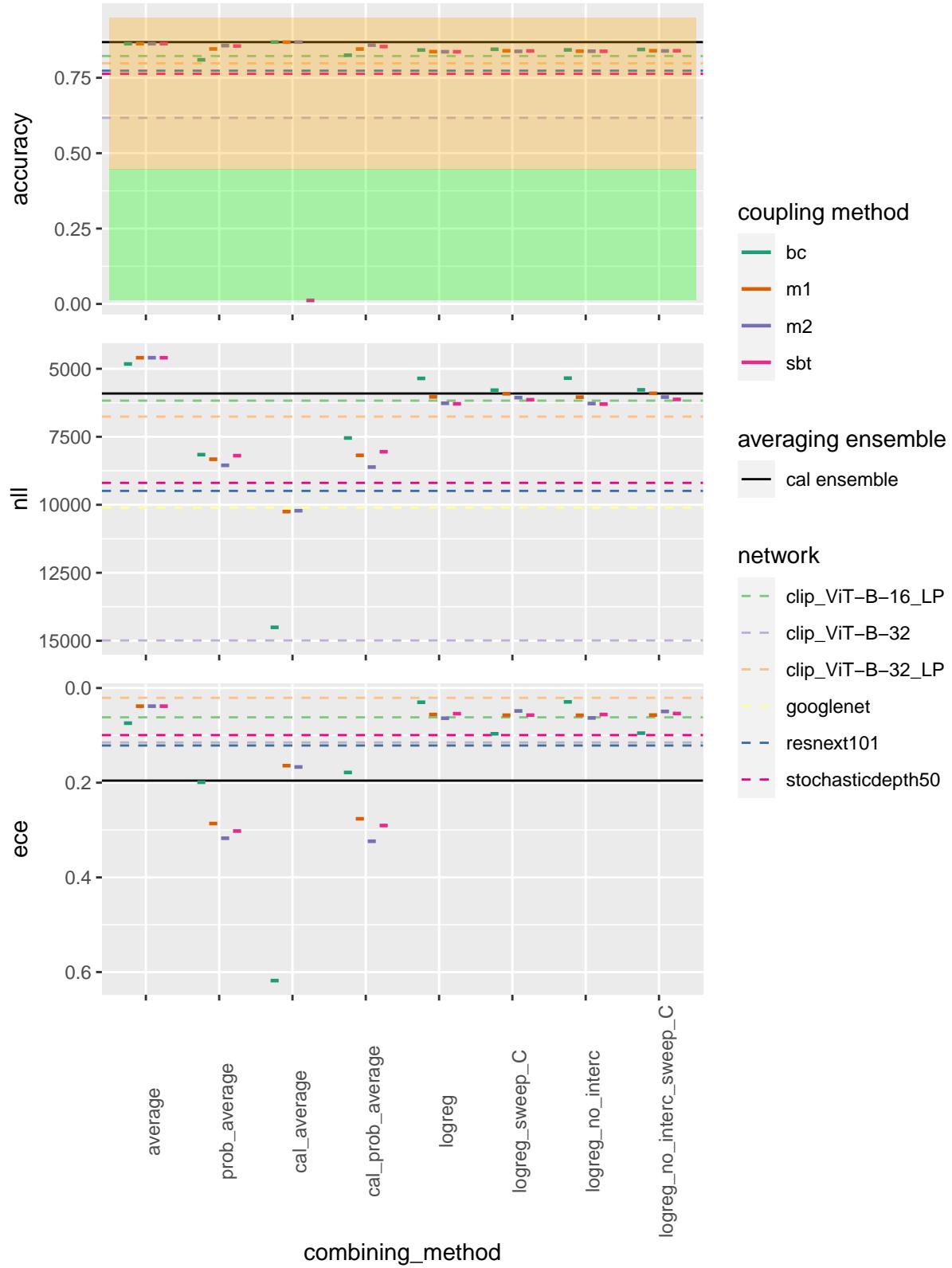
Ensemble metrics  
Error inconsistency 0.496799975633621



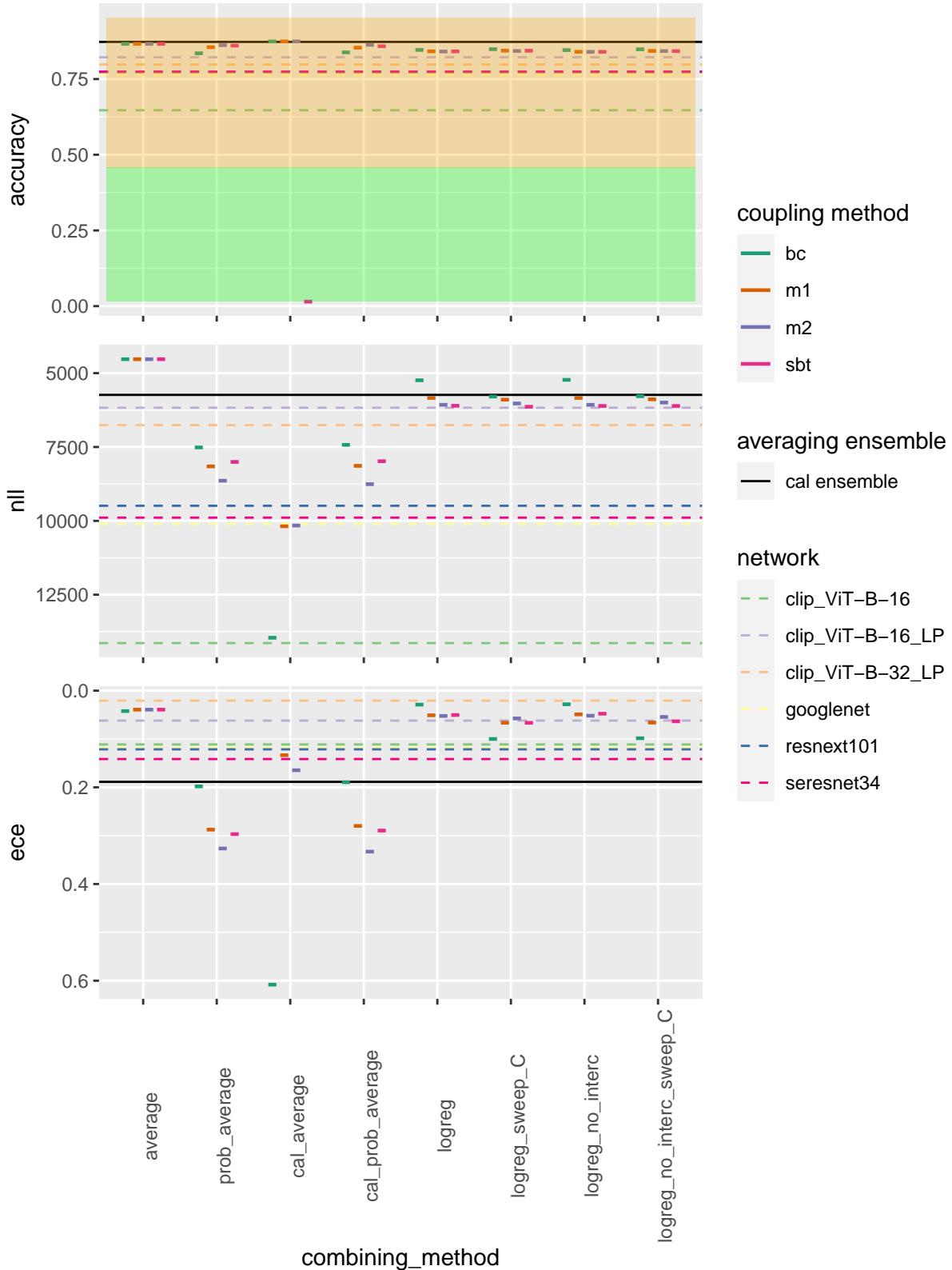
Ensemble metrics  
Error inconsistency 0.502099990844727



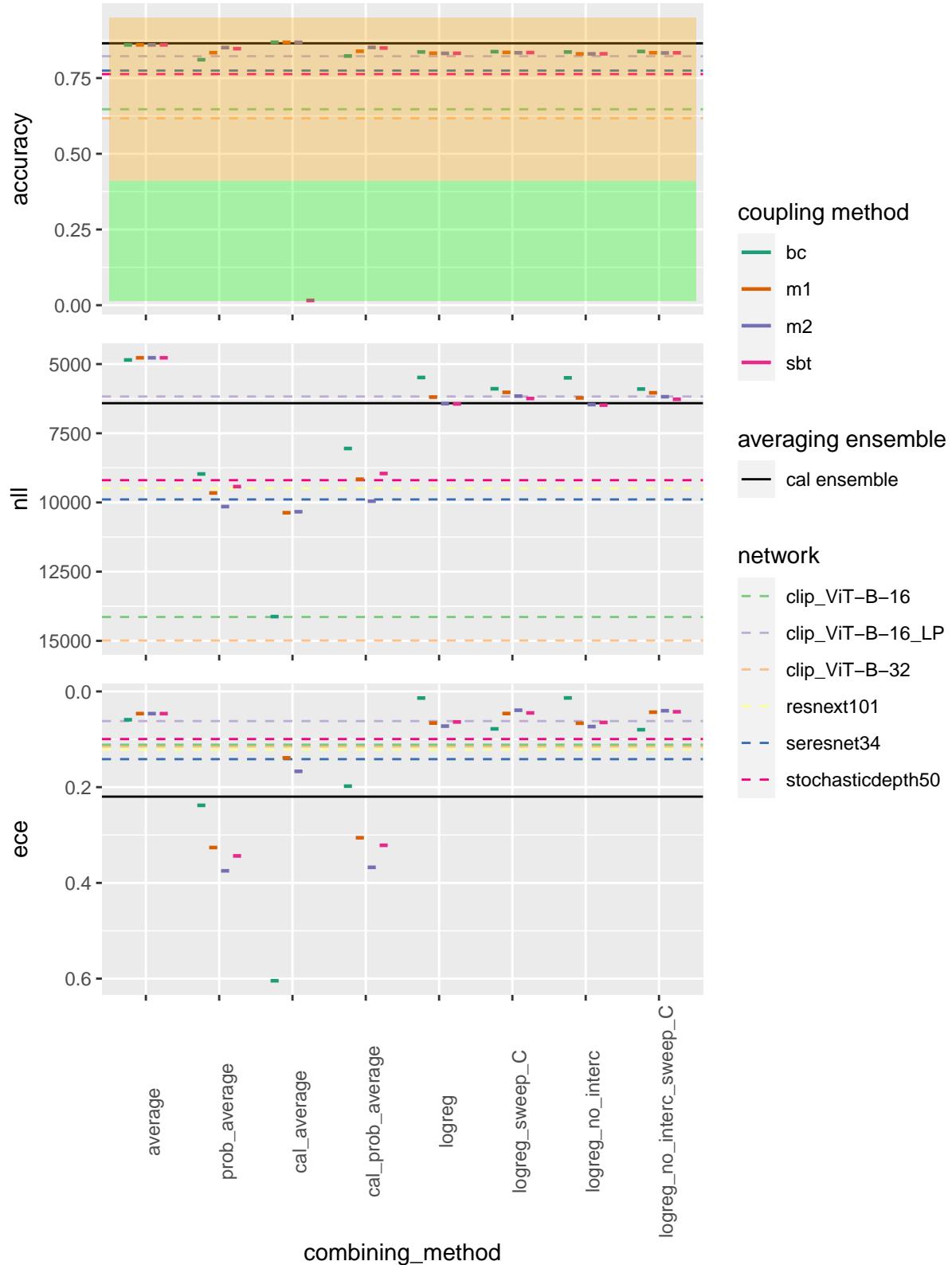
Ensemble metrics  
Error inconsistency 0.503199994564056



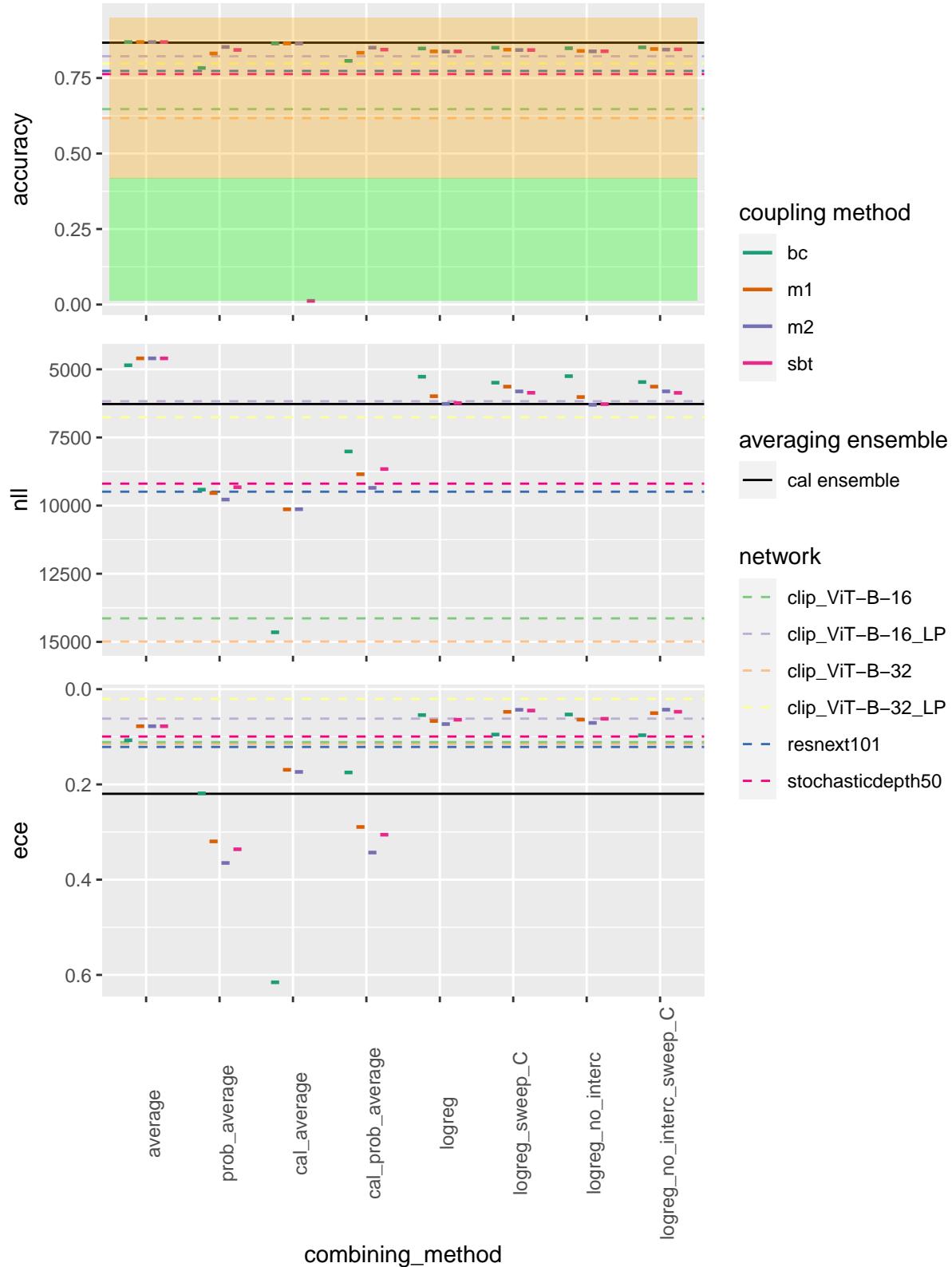
Ensemble metrics  
Error inconsistency 0.494599997997284



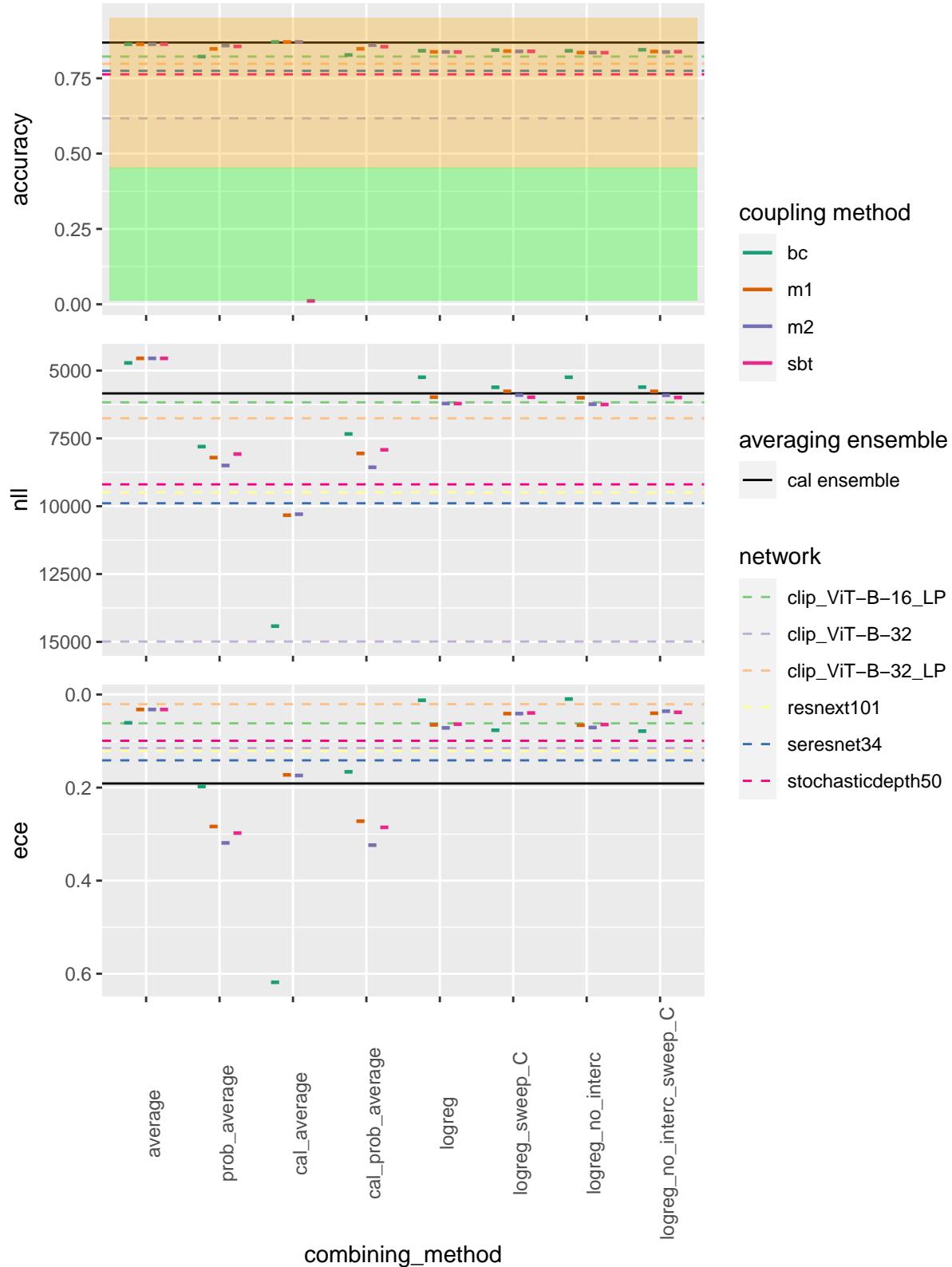
Ensemble metrics  
Error inconsistency 0.538999974727631



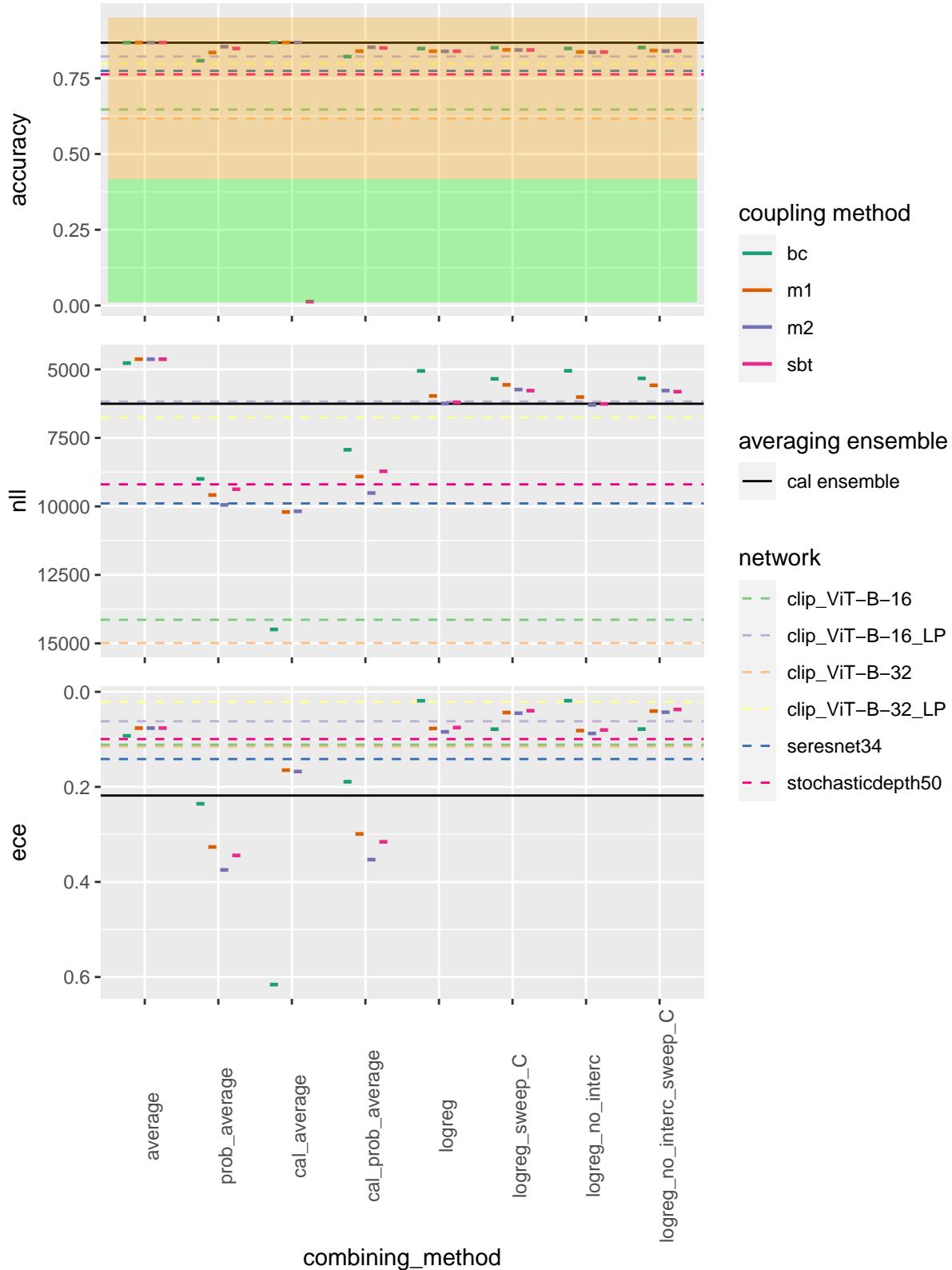
Ensemble metrics  
Error inconsistency 0.532899975776672



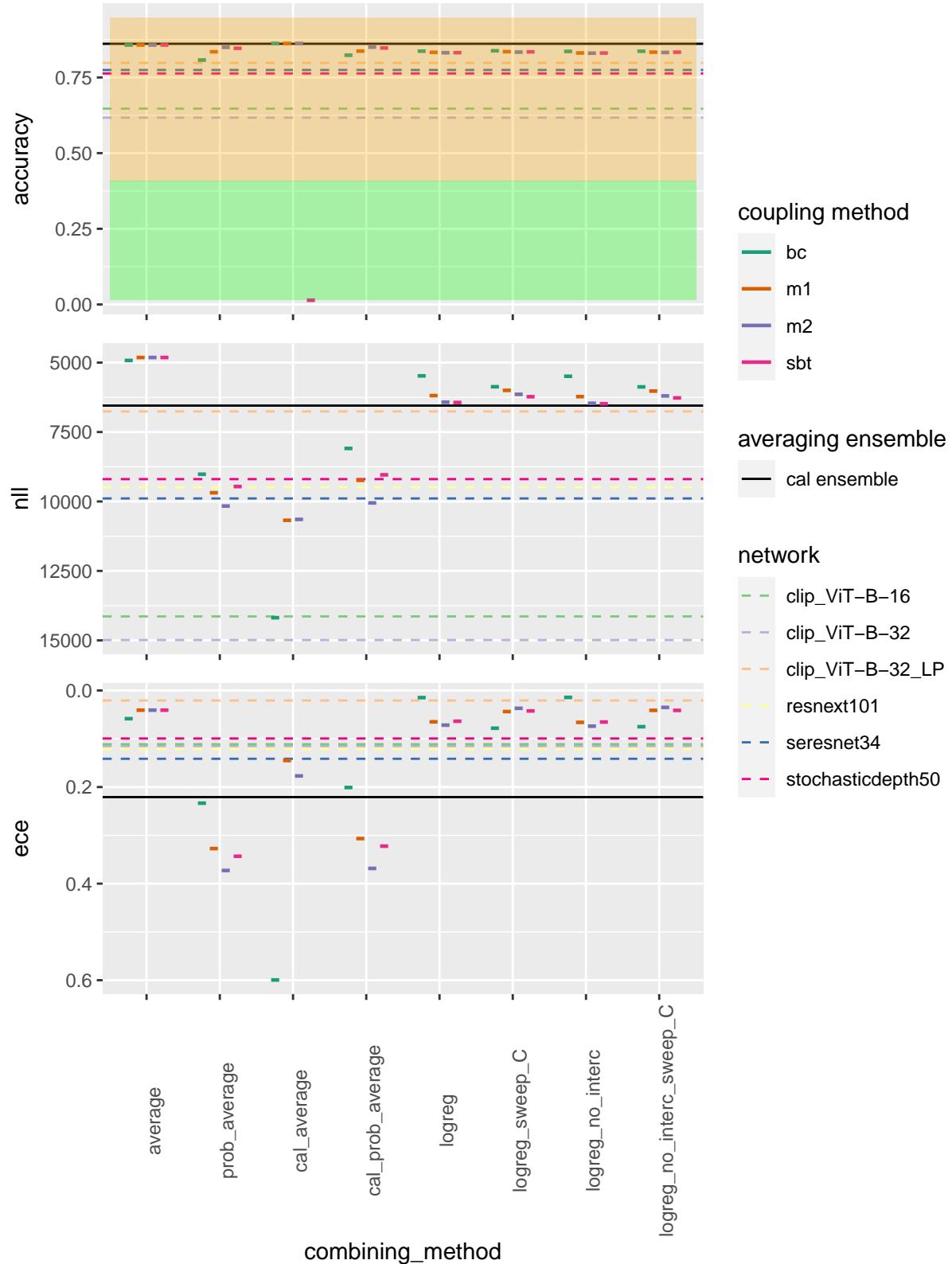
Ensemble metrics  
Error inconsistency 0.499199986457825



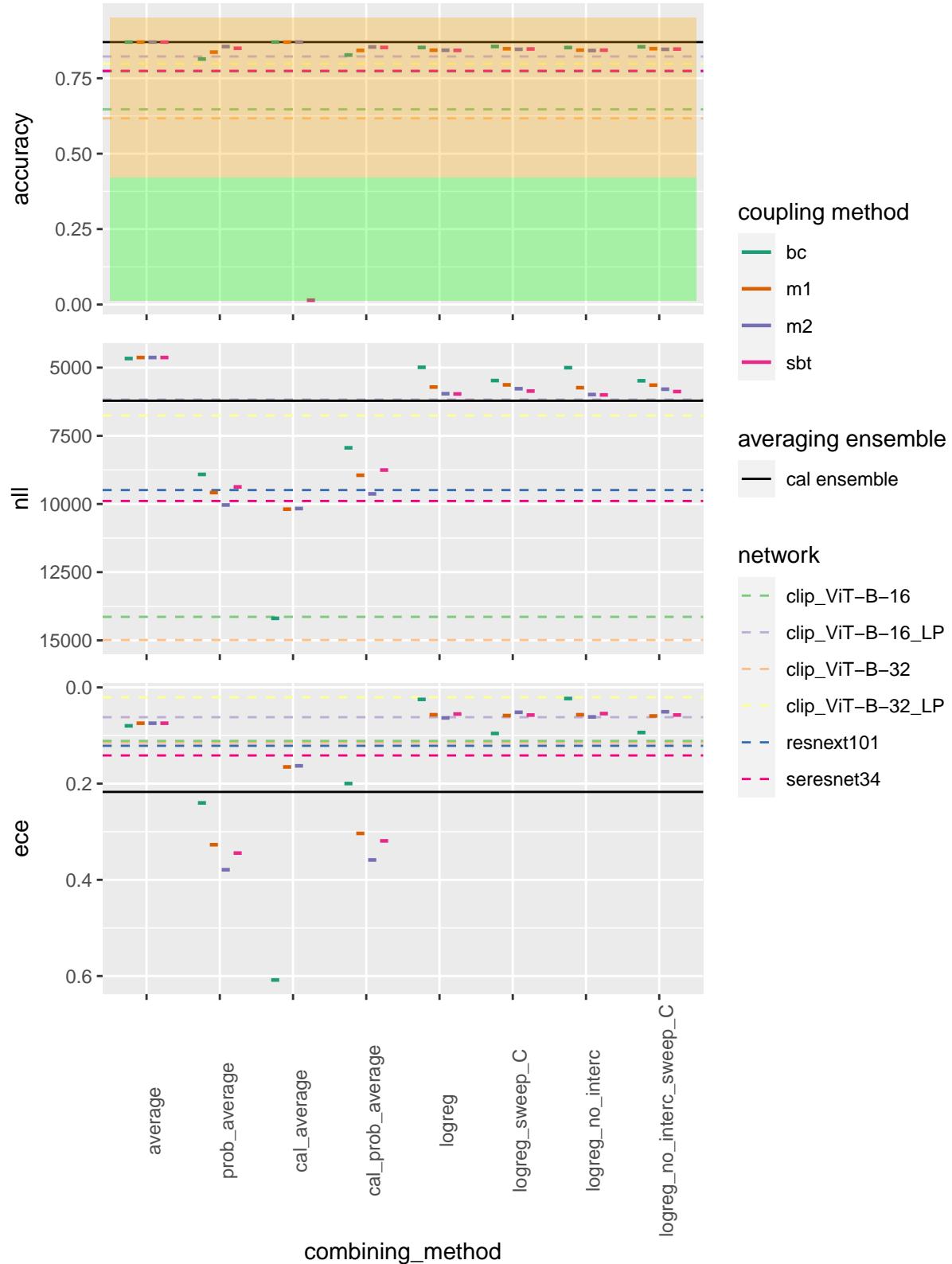
Ensemble metrics  
Error inconsistency 0.532899975776672



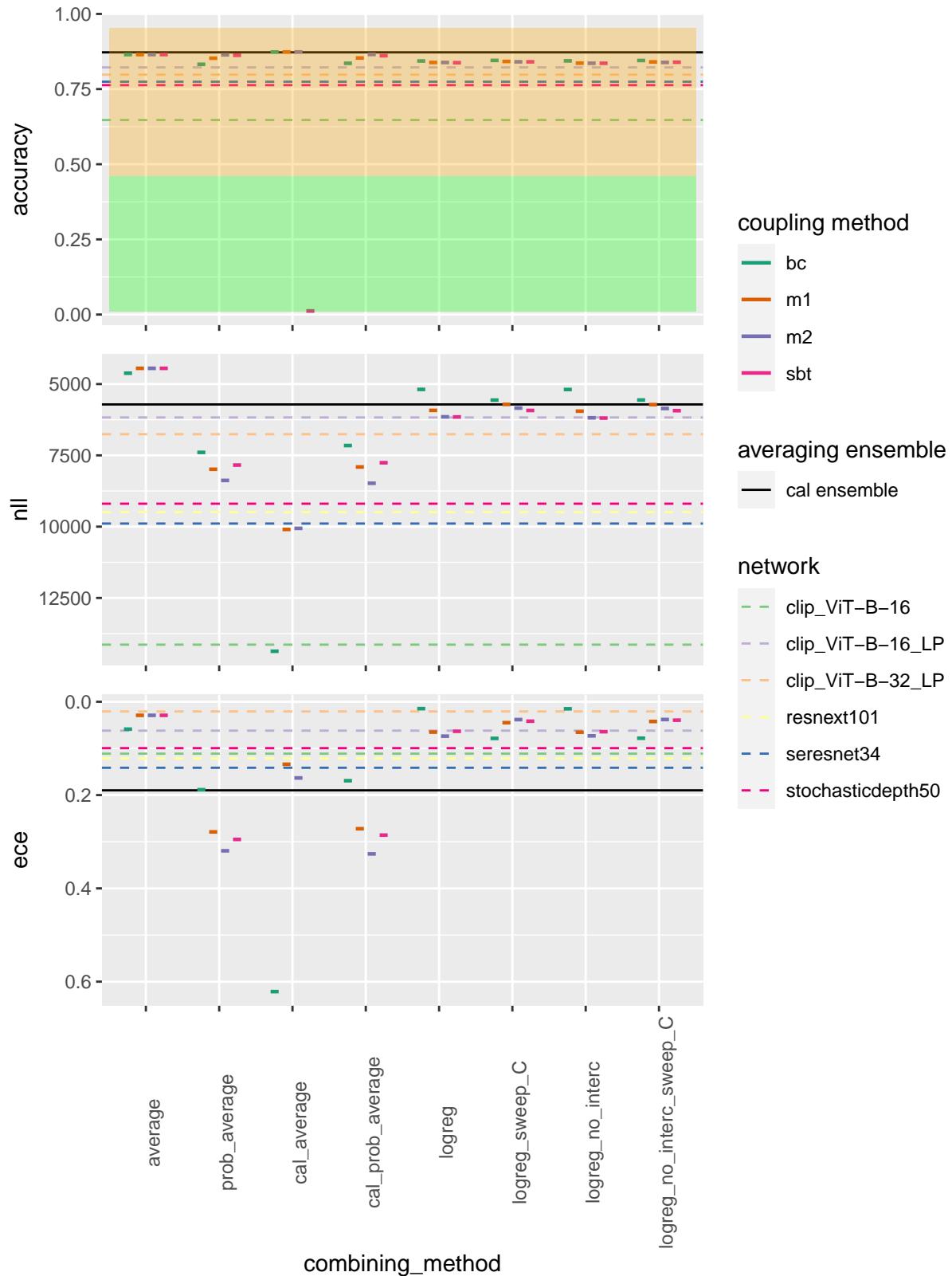
Ensemble metrics  
Error inconsistency 0.541000008583069



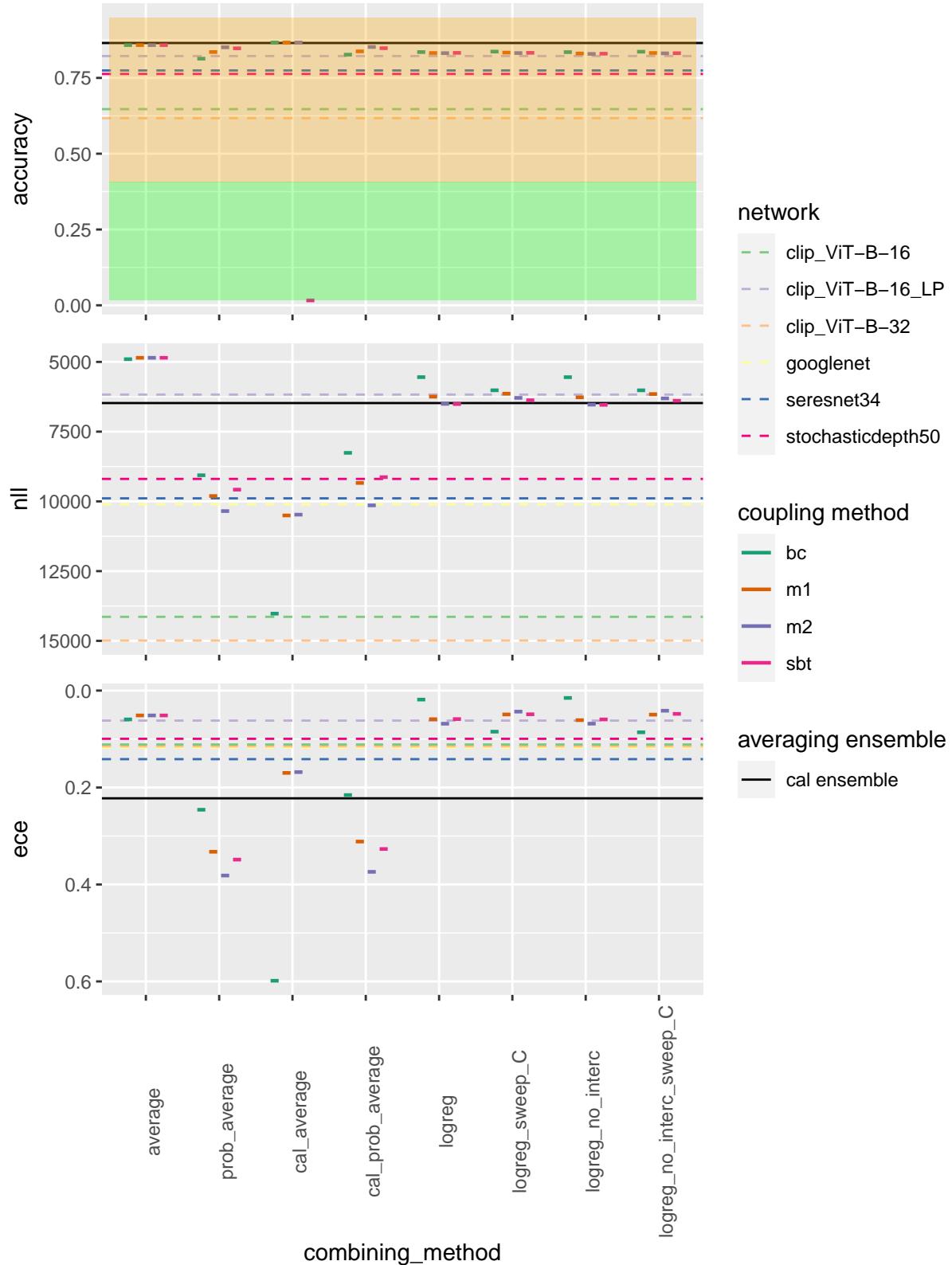
Ensemble metrics  
Error inconsistency 0.529900014400482



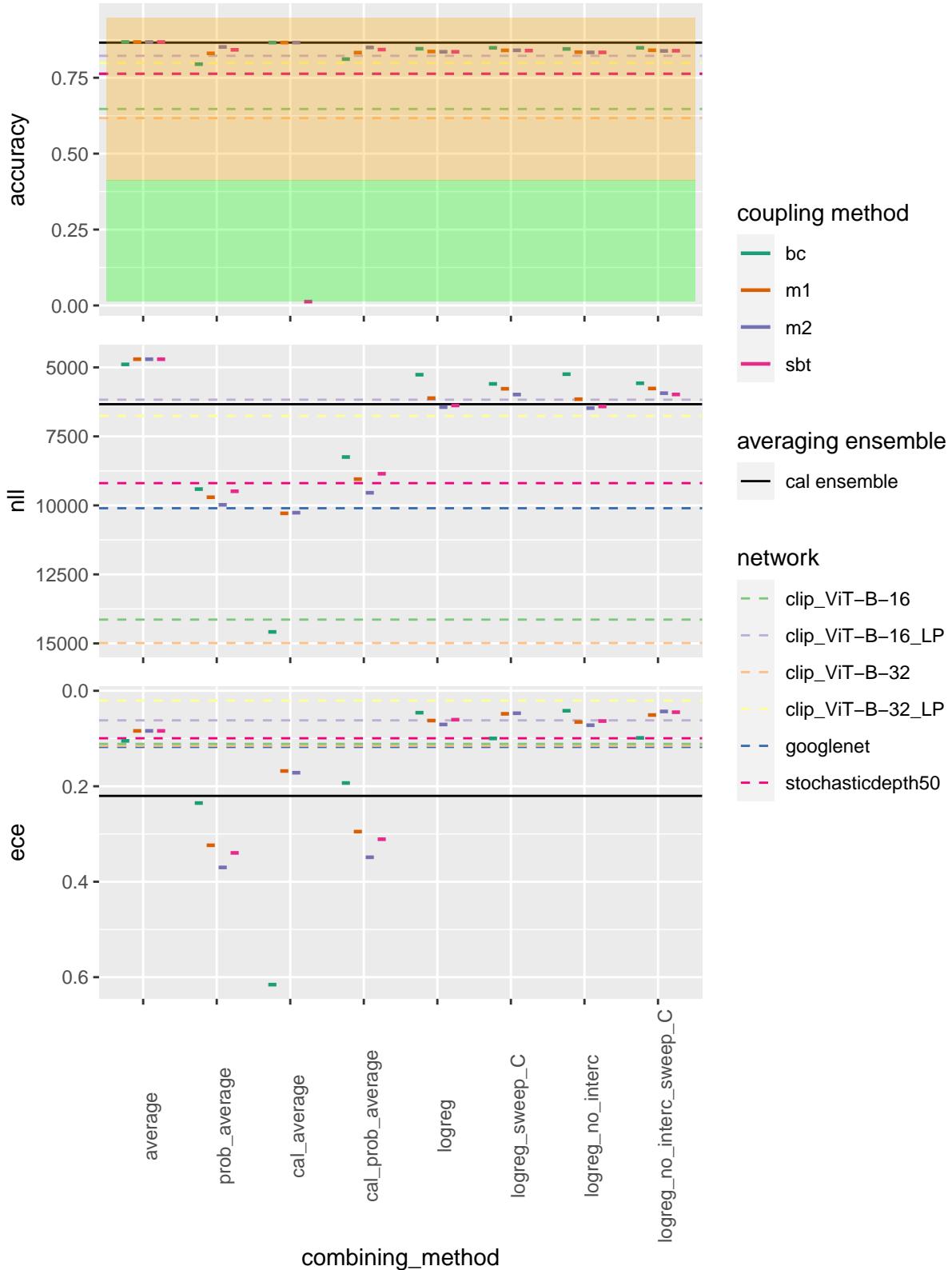
Ensemble metrics  
Error inconsistency 0.493099987506866



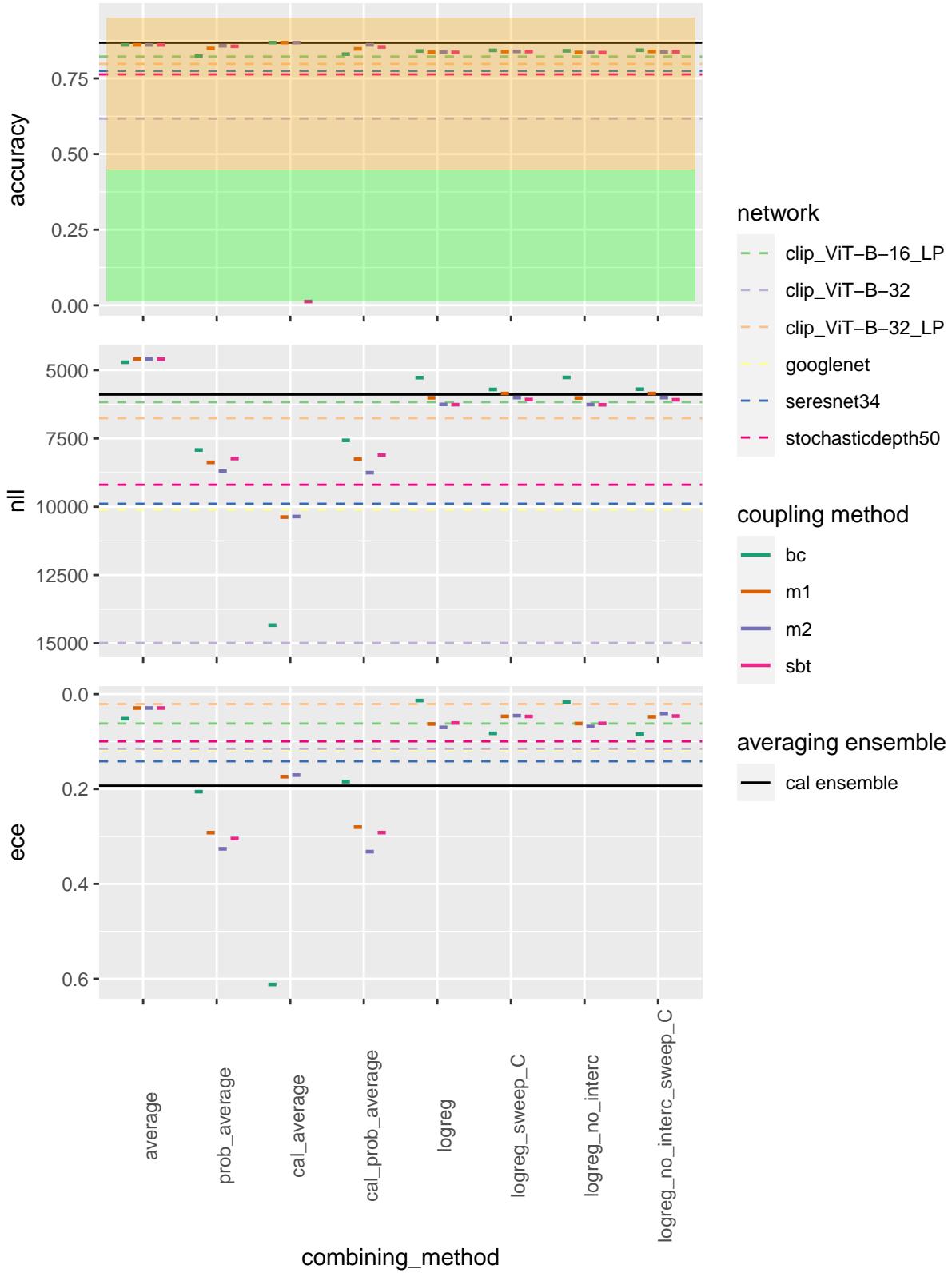
Ensemble metrics  
Error inconsistency 0.542599976062775



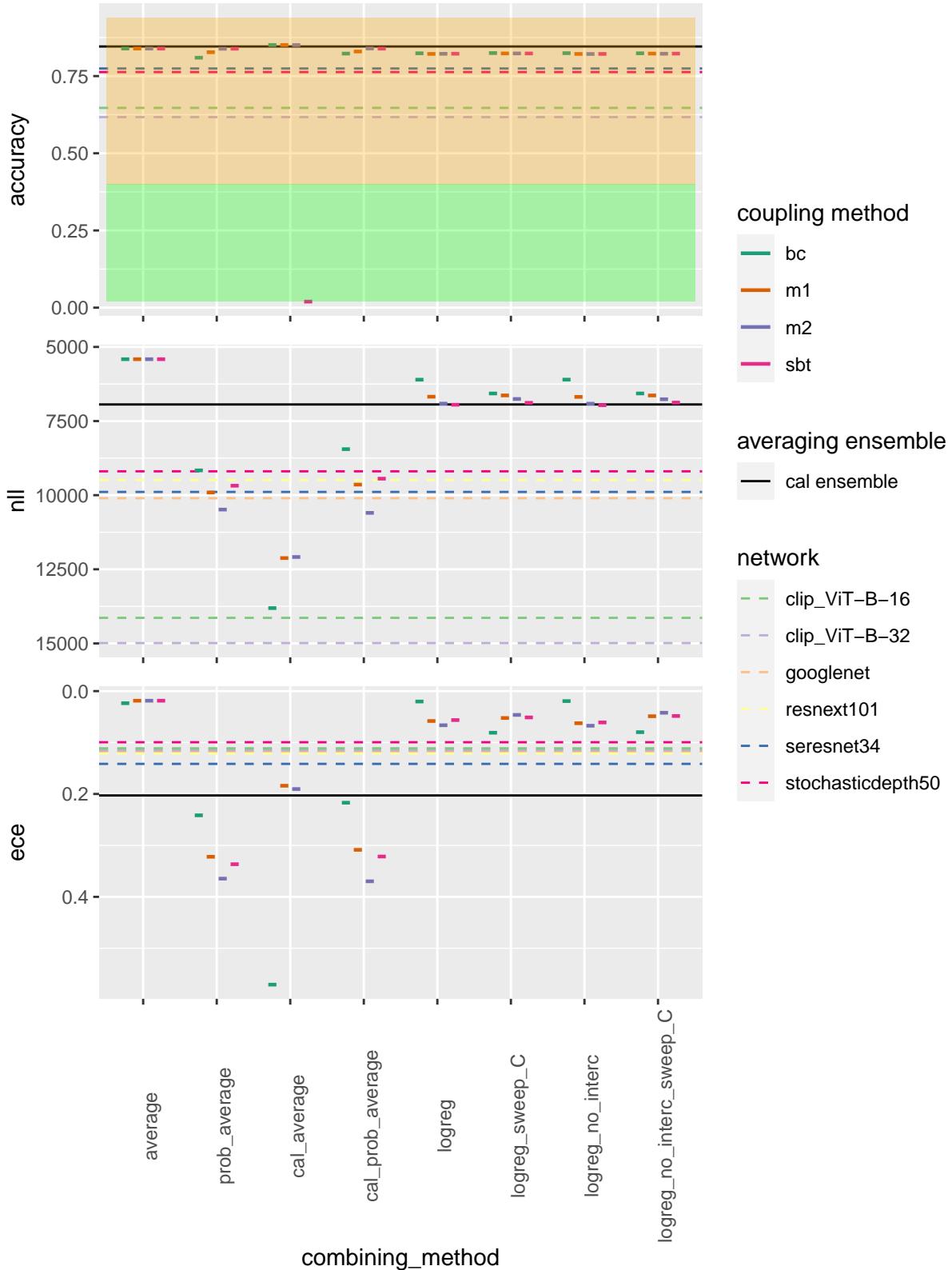
Ensemble metrics  
Error inconsistency 0.535099983215332



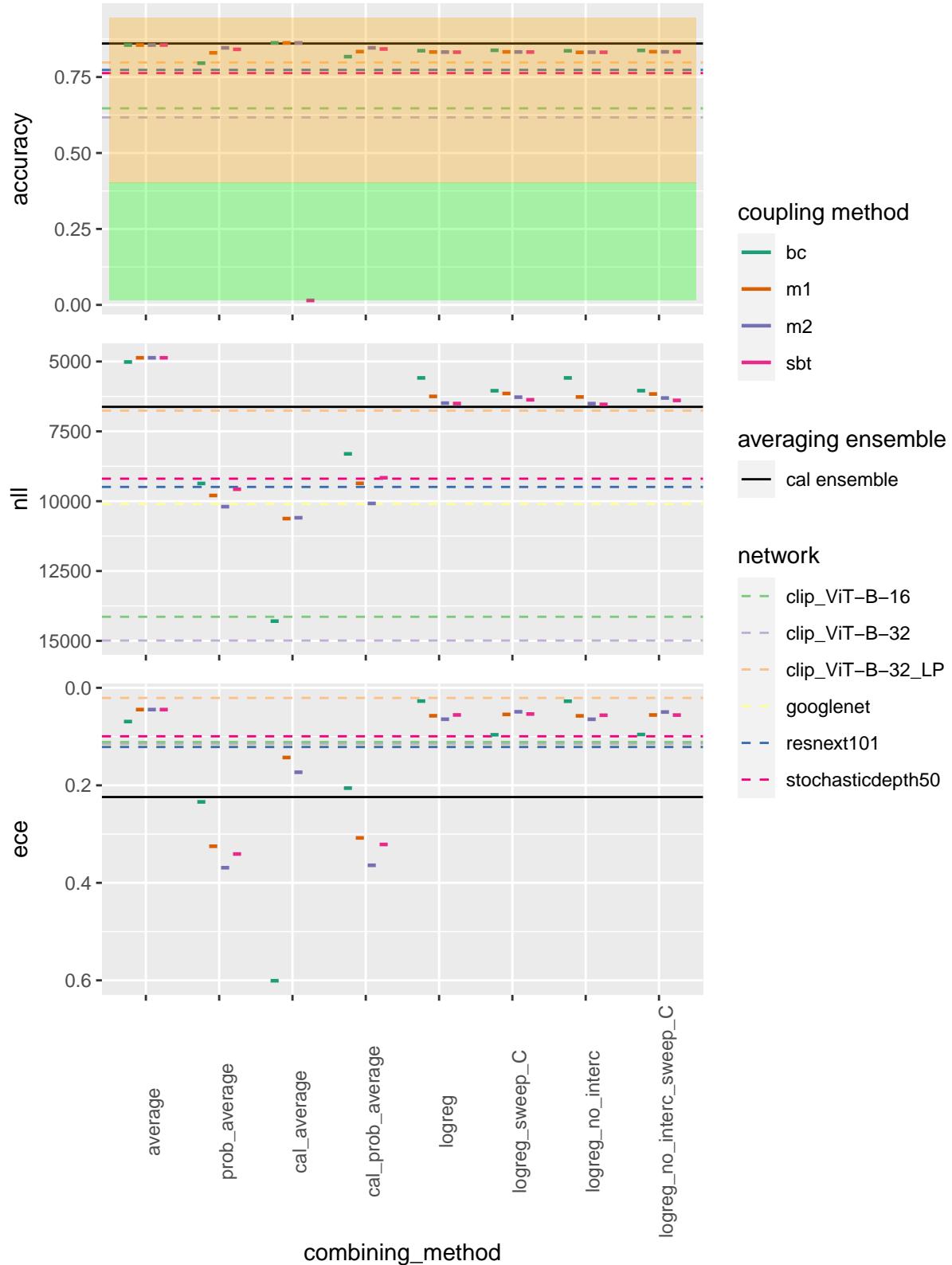
Ensemble metrics  
Error inconsistency 0.503600001335144



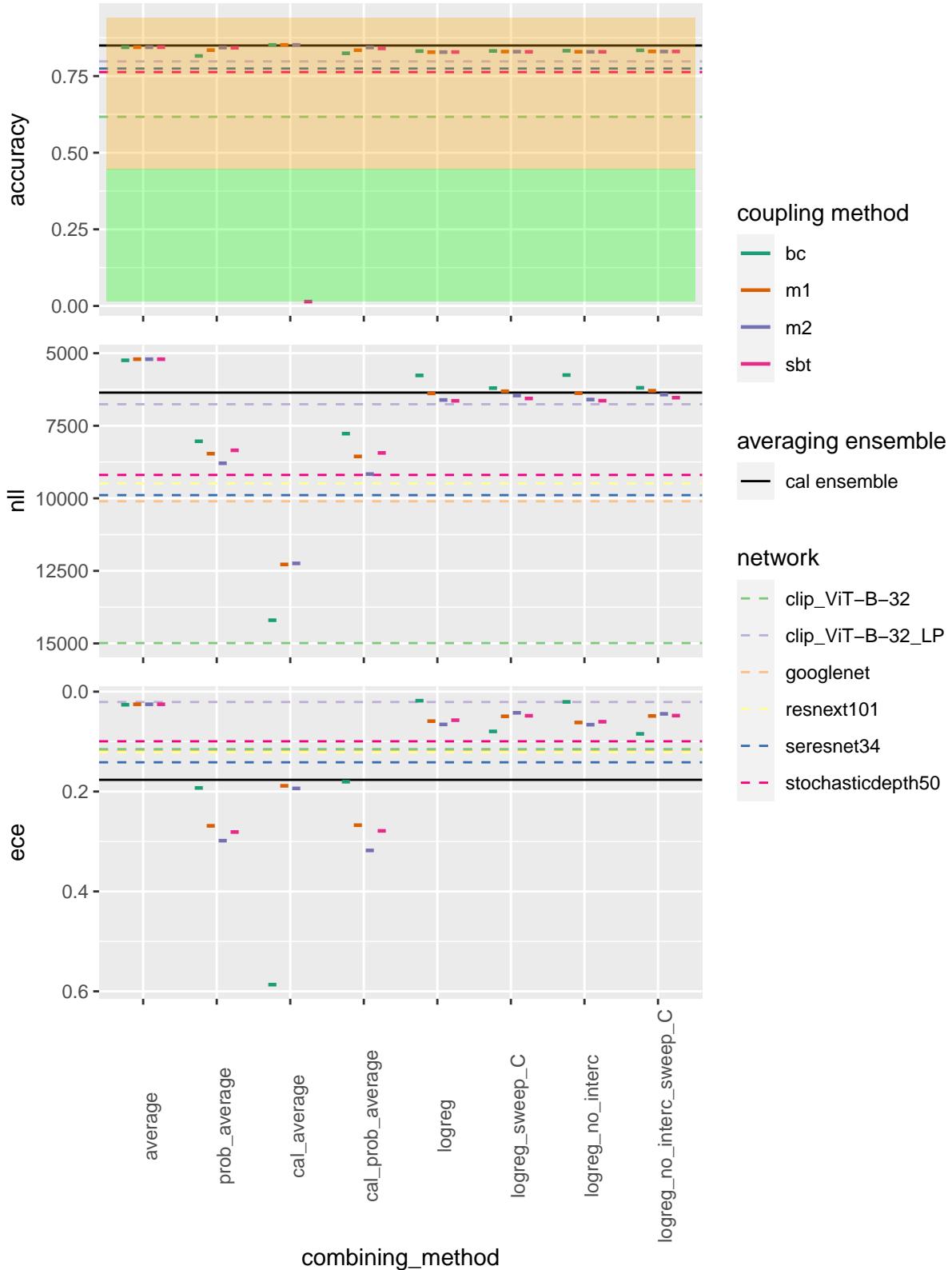
Ensemble metrics  
Error inconsistency 0.539699971675873



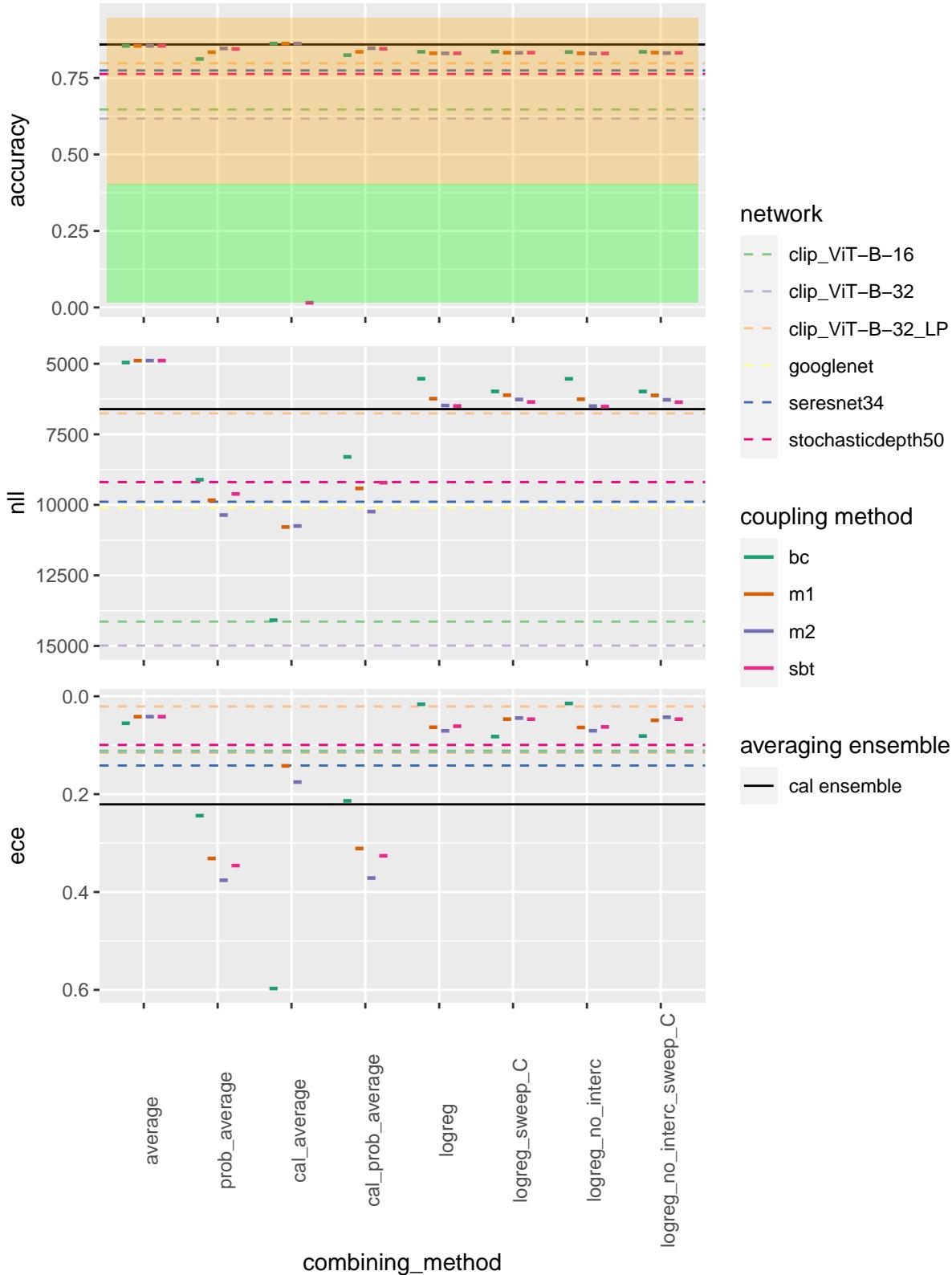
Ensemble metrics  
Error inconsistency 0.543900012969971



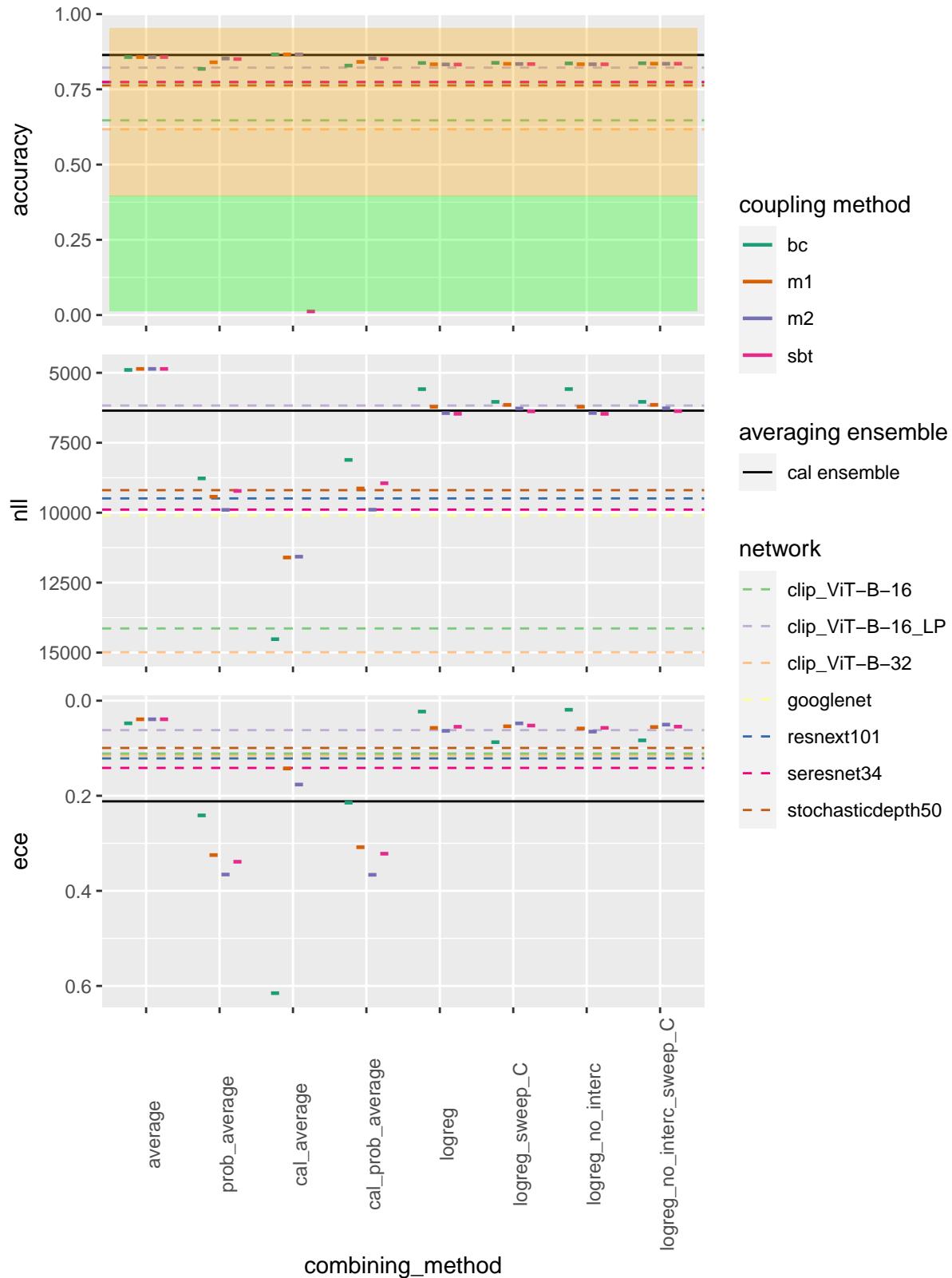
Ensemble metrics  
Error inconsistency 0.494800001382828



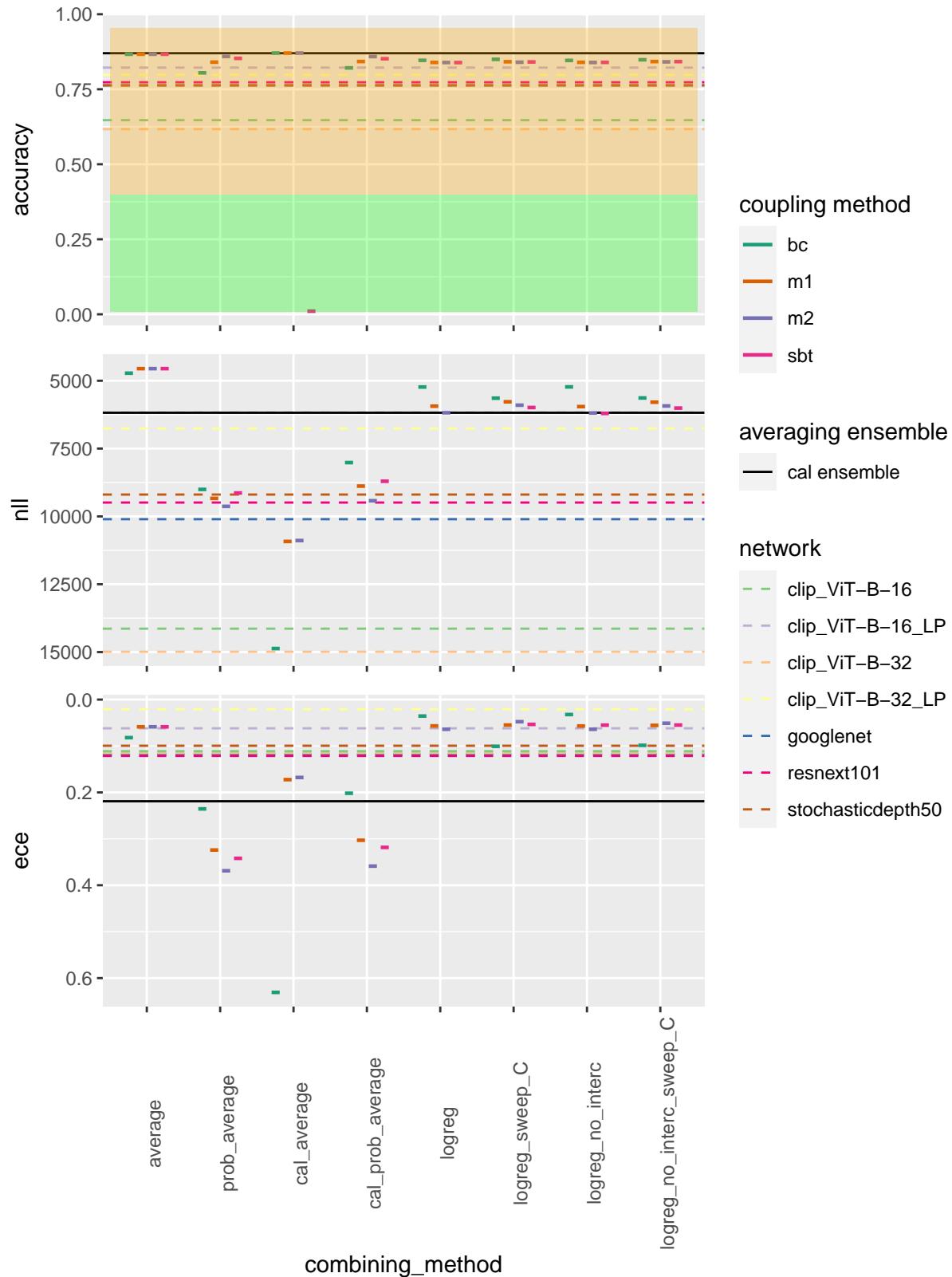
Ensemble metrics  
Error inconsistency 0.544099986553192



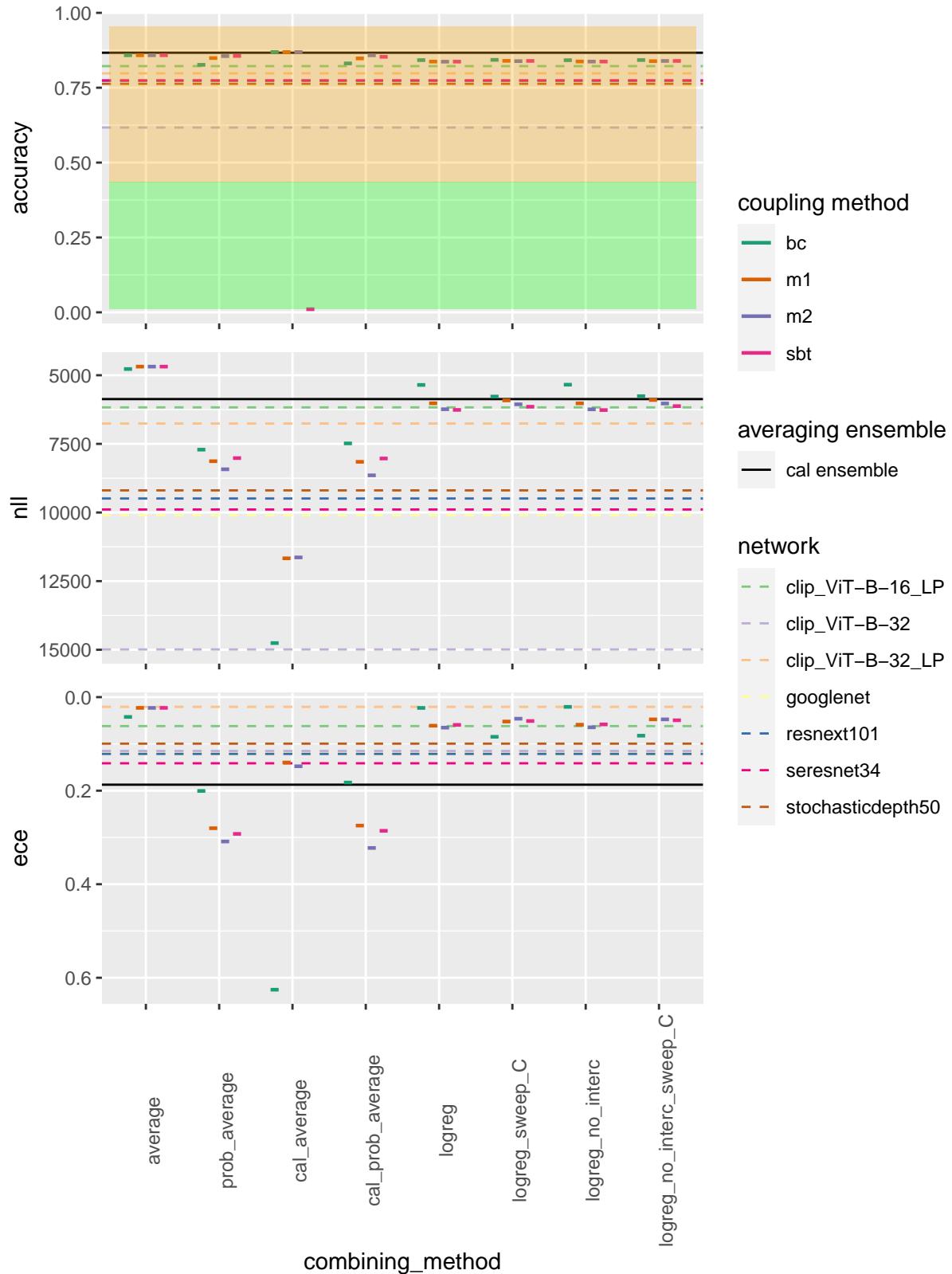
Ensemble metrics  
Error inconsistency 0.557699978351593



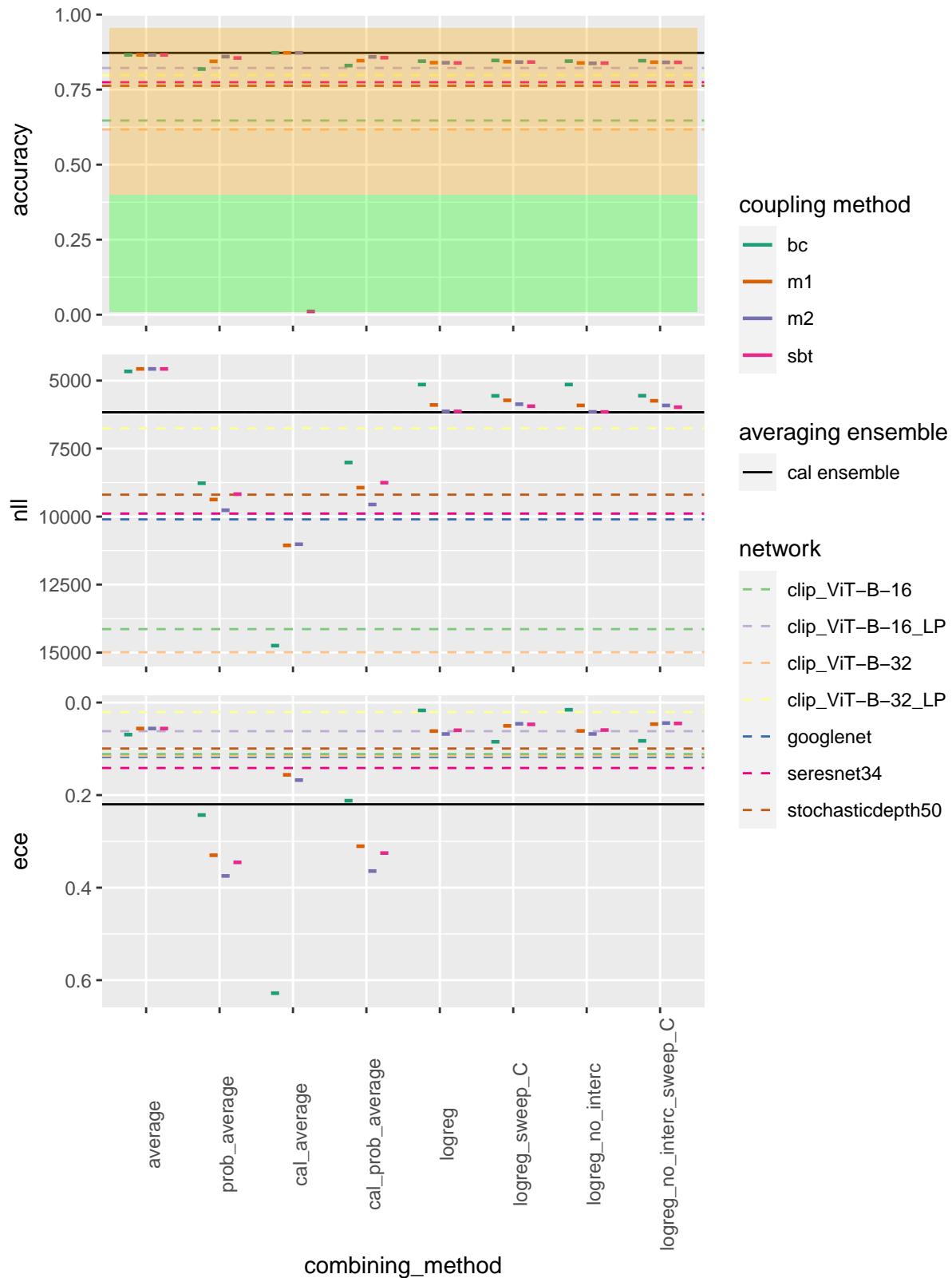
Ensemble metrics  
Error inconsistency 0.555099964141846



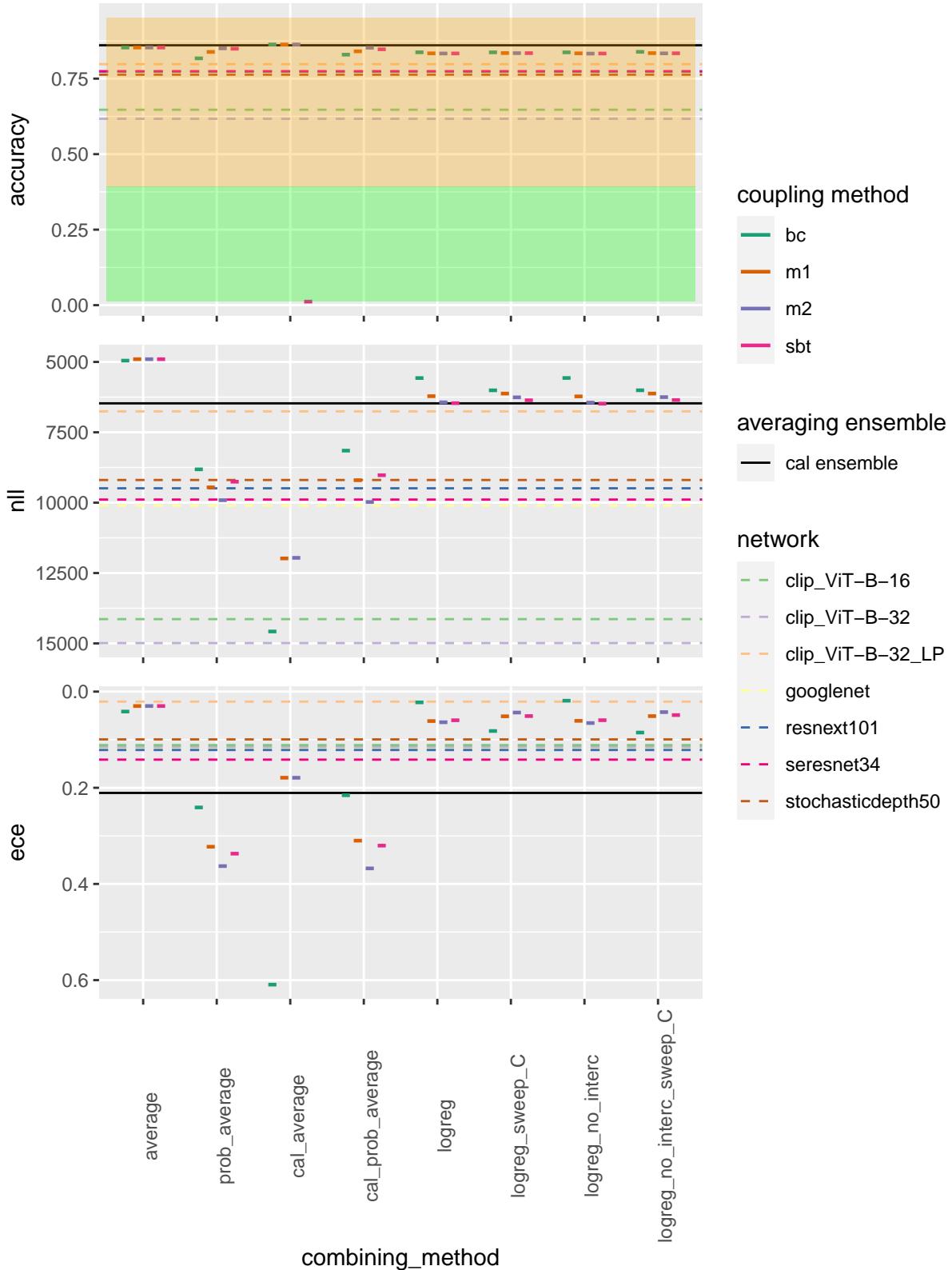
Ensemble metrics  
Error inconsistency 0.518899977207184



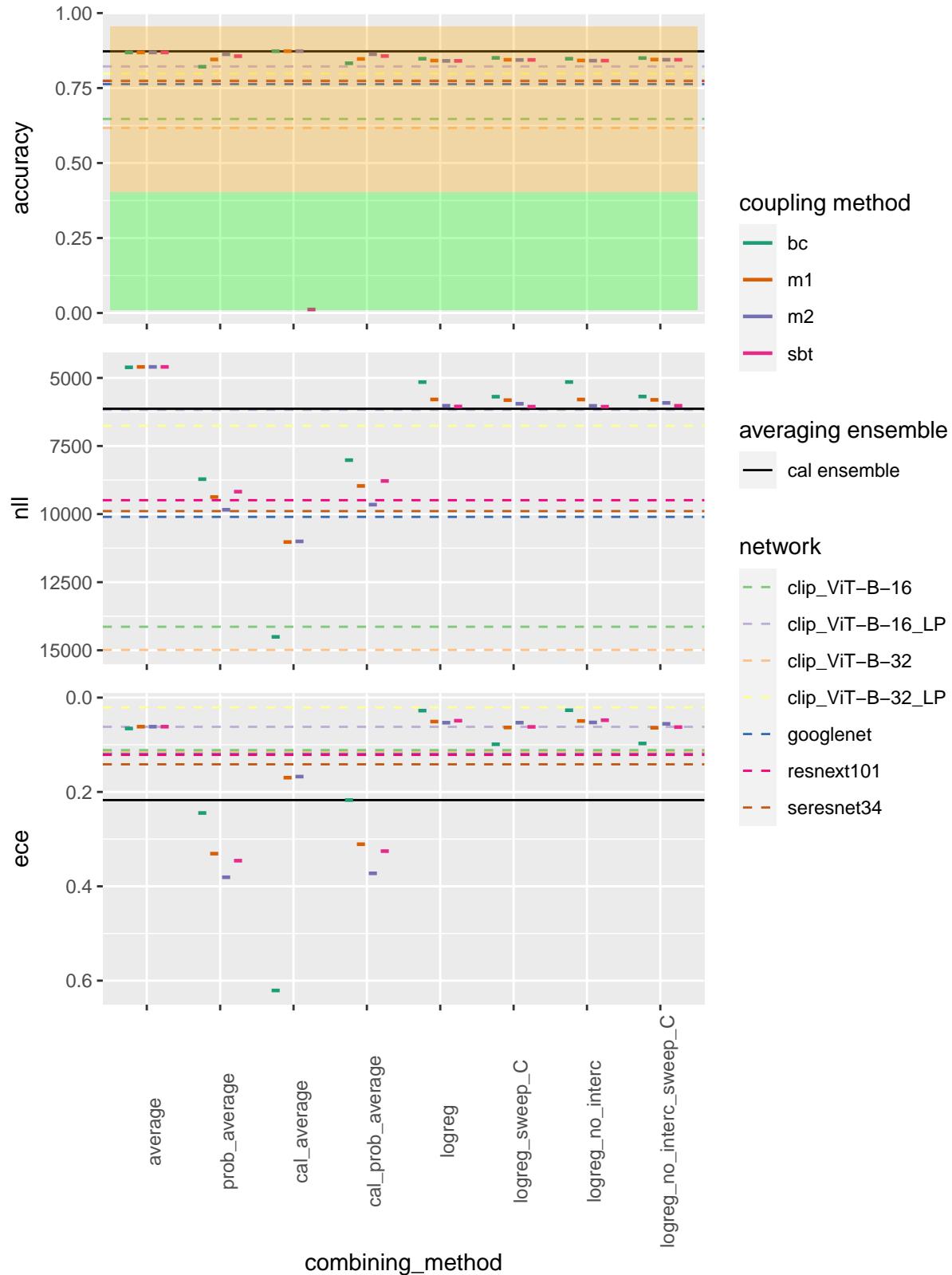
Ensemble metrics  
Error inconsistency 0.555400013923645



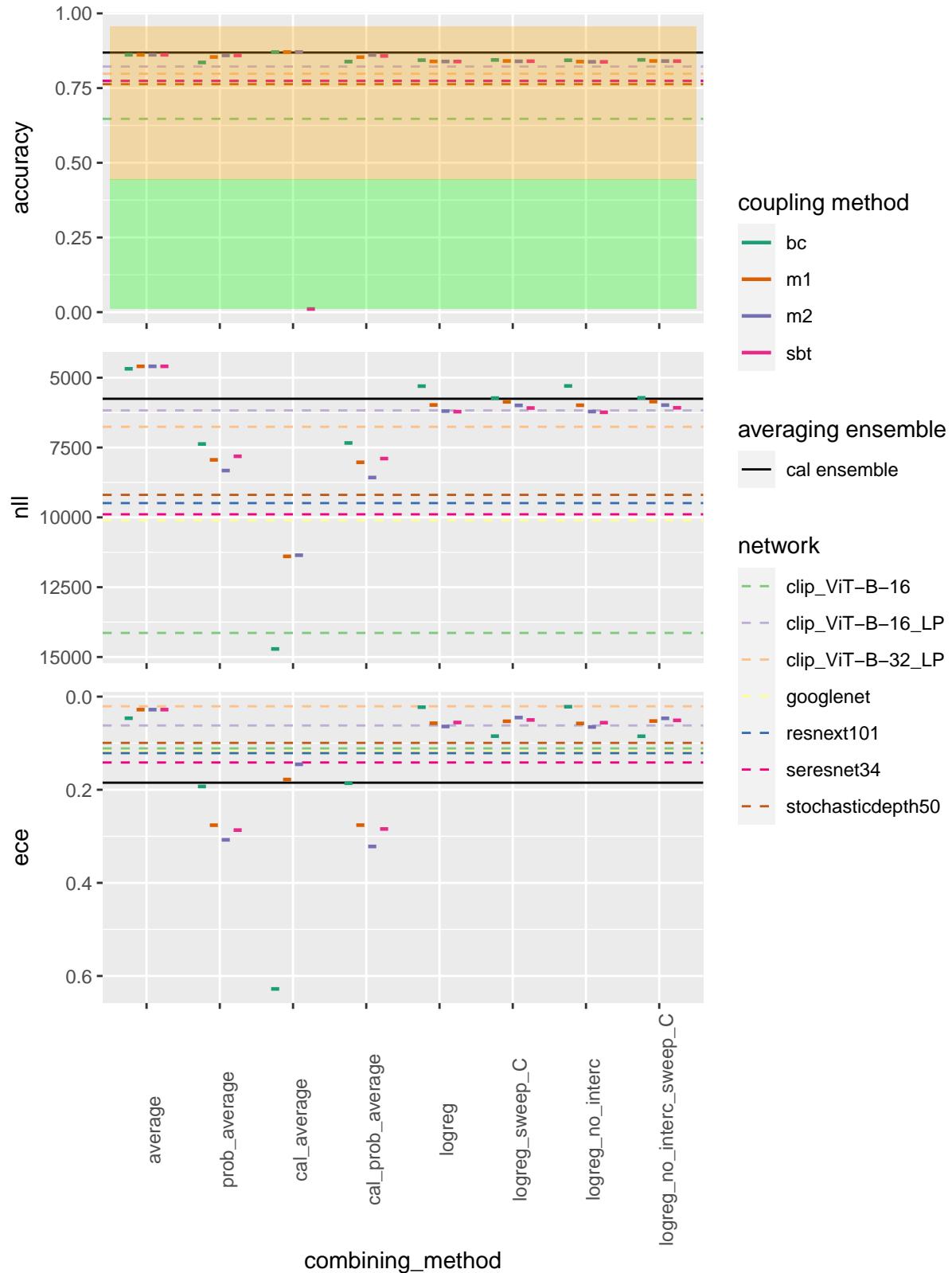
Ensemble metrics  
Error inconsistency 0.559099972248077



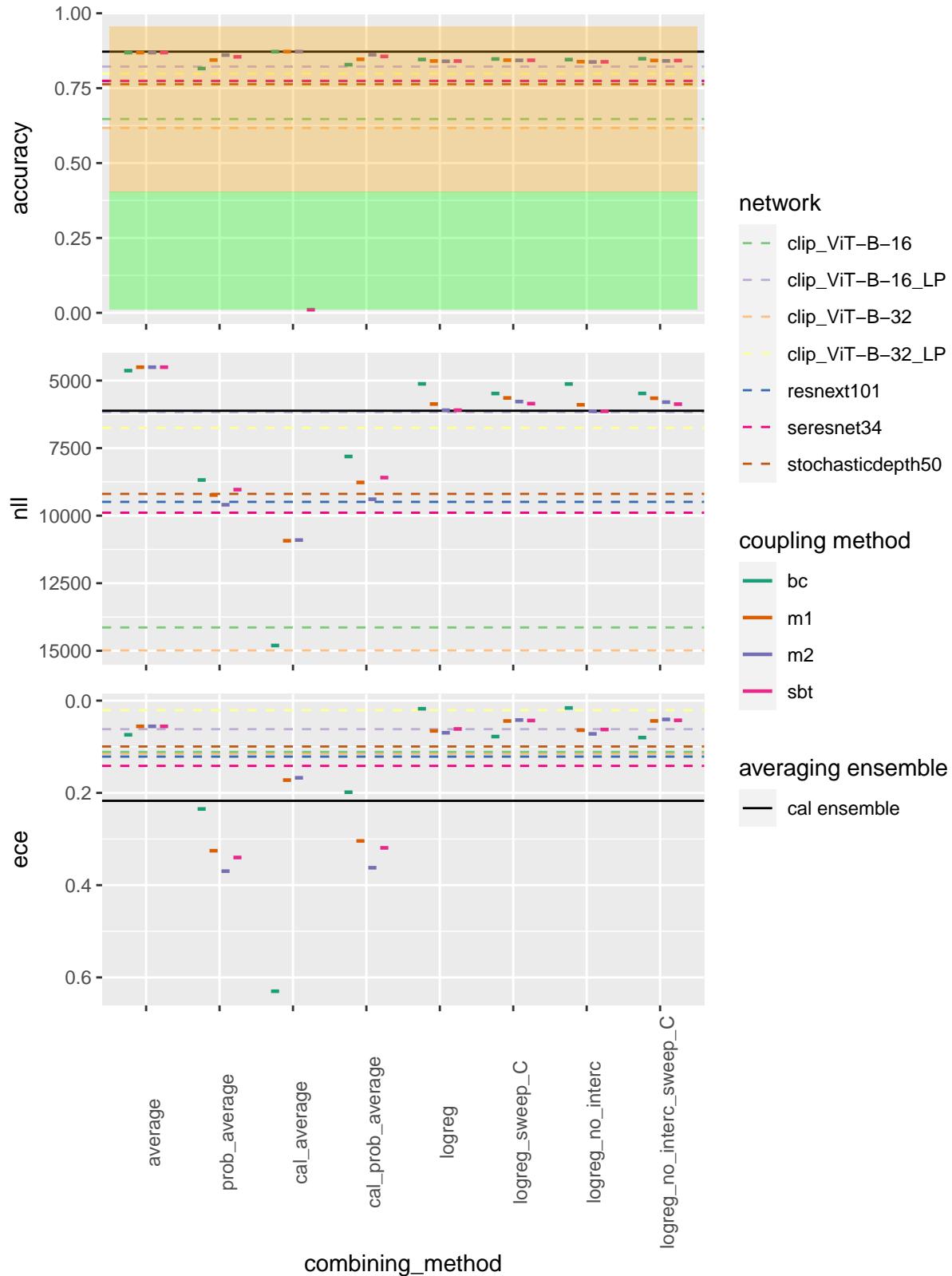
Ensemble metrics  
Error inconsistency 0.552599966526031



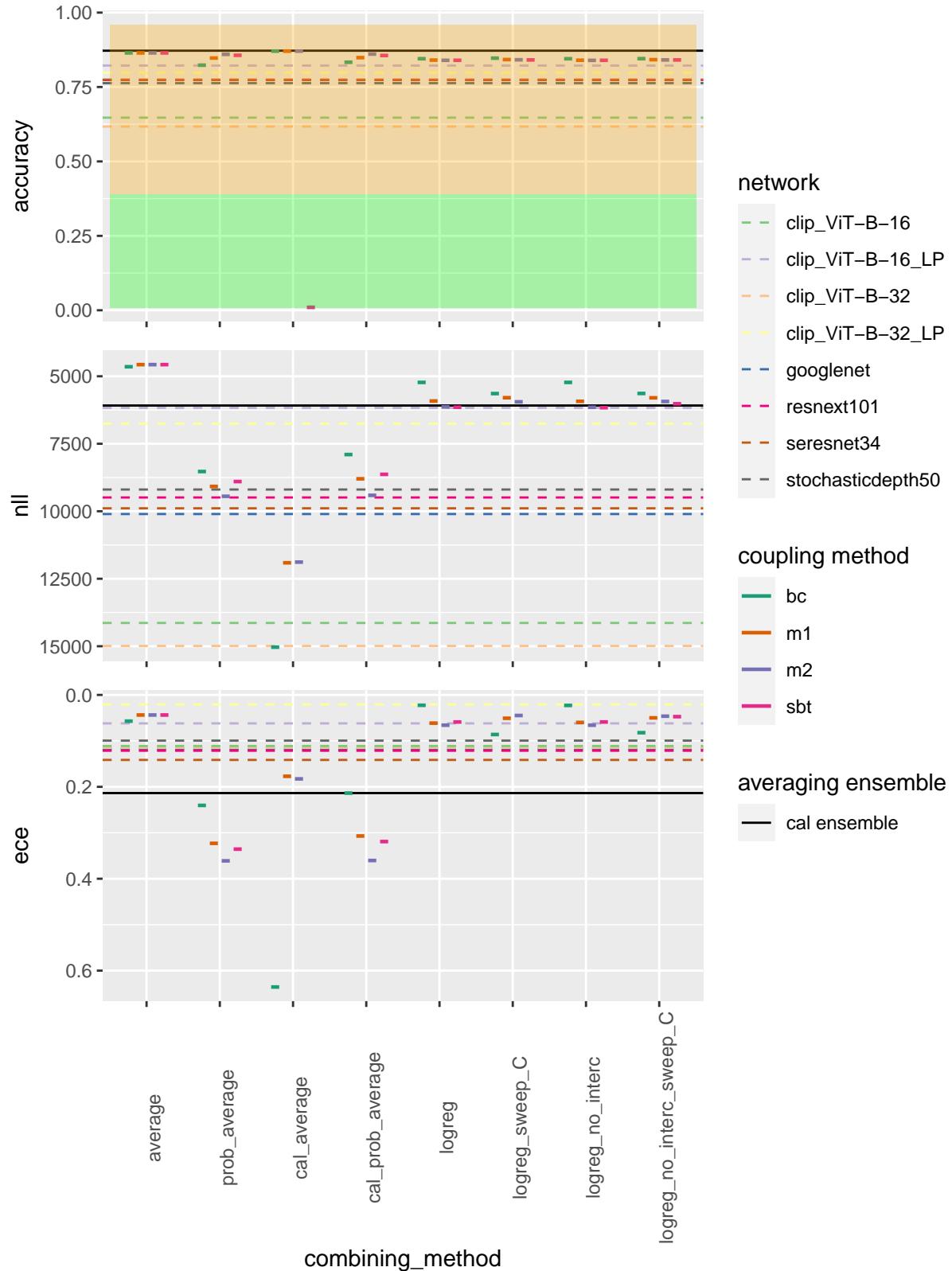
Ensemble metrics  
Error inconsistency 0.513399958610535



Ensemble metrics  
Error inconsistency 0.551899969577789



Ensemble metrics  
Error inconsistency 0.568899989128113

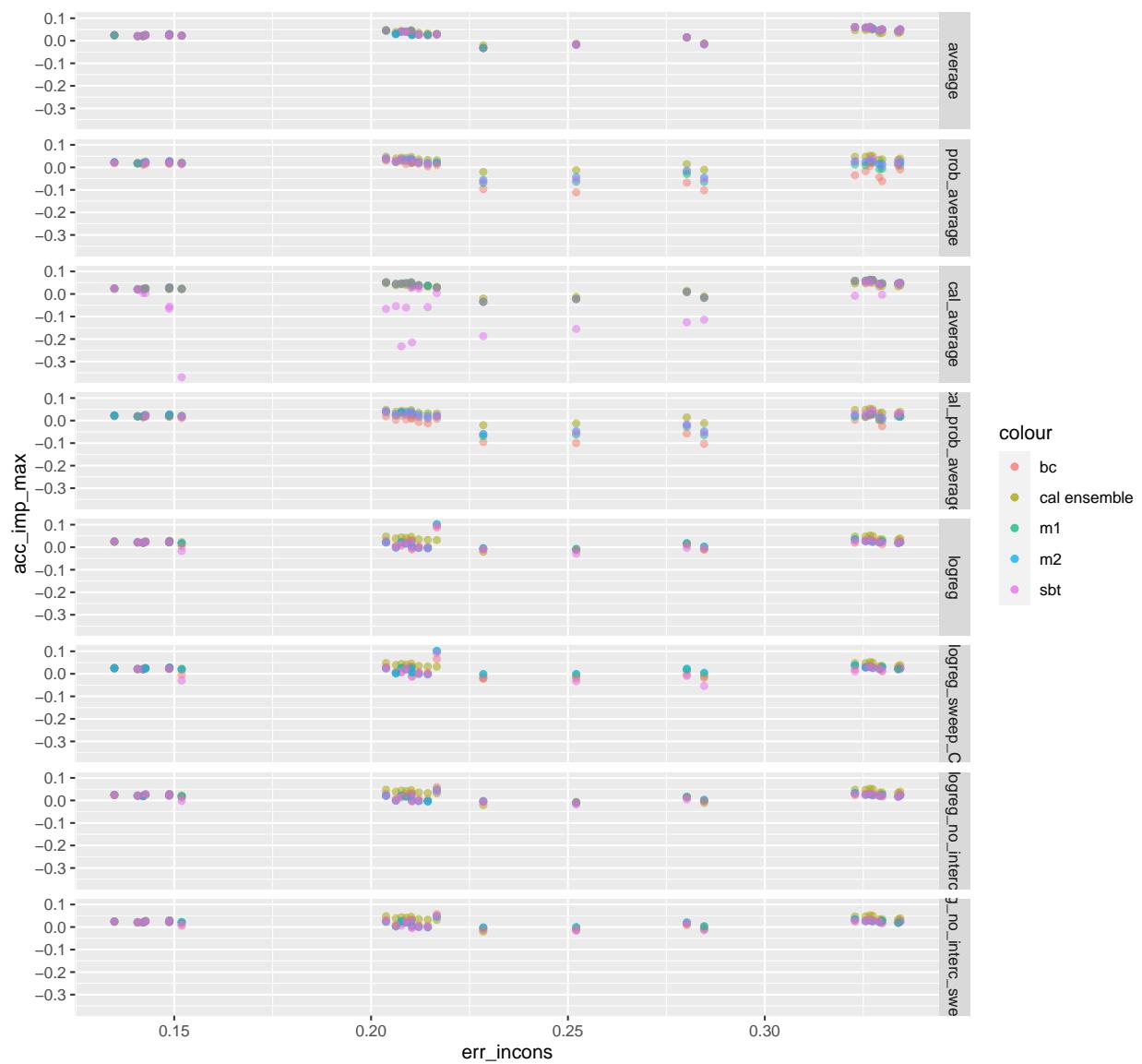


```

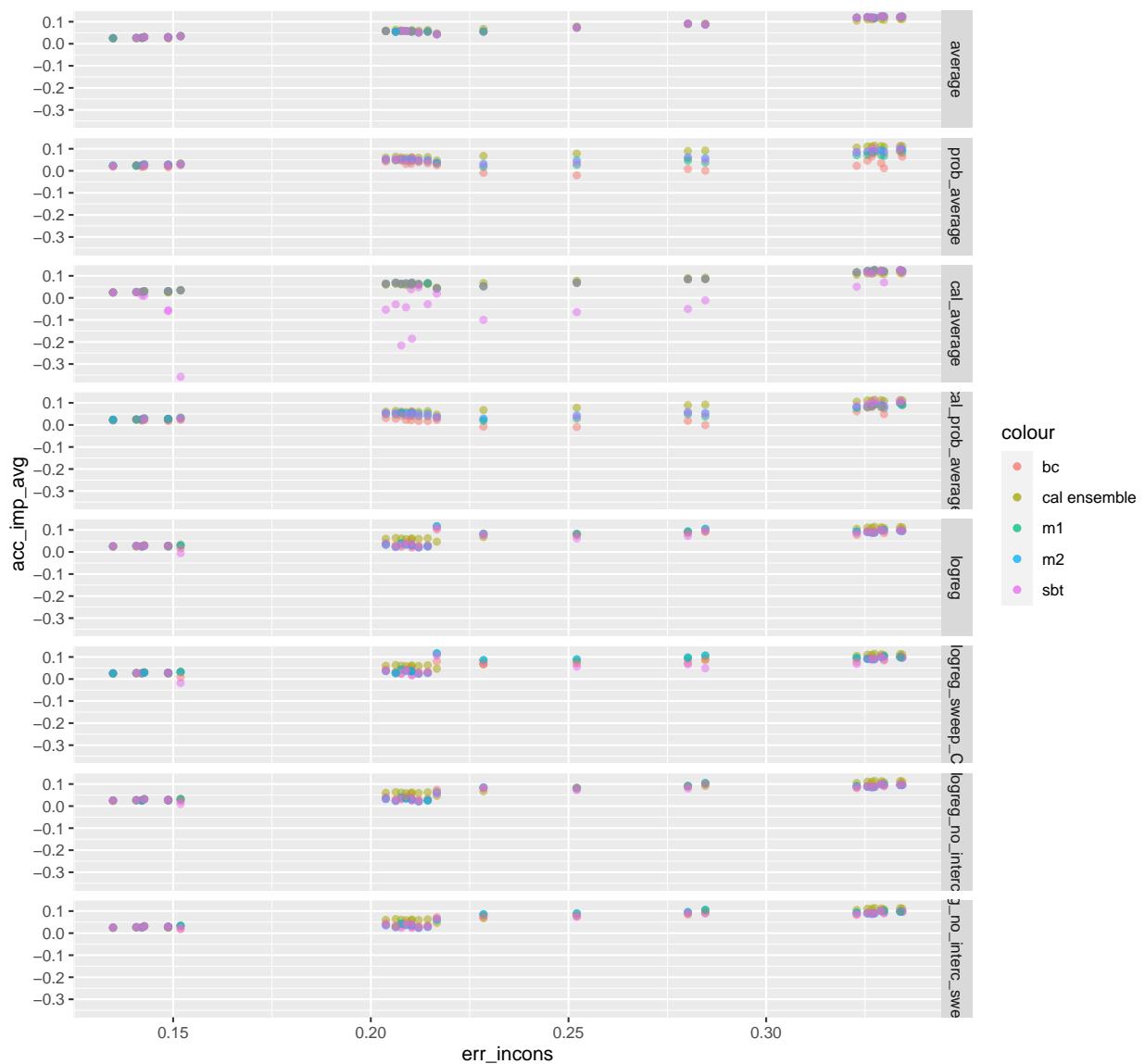
xax <- c(
  "err_incons",
  "nll_var", "nll_min", "nll_max", "nll_avg",
  "acc_var", "acc_min", "acc_max", "acc_avg",
  "ece_var", "ece_min", "ece_max", "ece_avg"
)
yax <- c("acc_imp_max", "acc_imp_avg")
for (sss in unique(ens_cal_plt_df$combination_size))
{
  cur_ens_cal_plt_df <- ens_cal_plt_df %>% filter(combination_size == sss)
  cur_ens_pwc_plt_df <- ens_pwc_plt_df %>% filter(combination_size == sss)
  for (xa in xax)
  {
    for (ya in yax)
    {
      cur_plot <- ggplot() +
        geom_point(
          data = cur_ens_cal_plt_df,
          mapping = aes_string(x = xa, y = ya, color = shQuote("cal ensemble")),
          alpha = 0.5
        ) +
        geom_point(
          data = cur_ens_pwc_plt_df,
          mapping = aes_string(x = xa, y = ya, color = "coupling_method"),
          alpha = 0.5
        ) +
        facet_grid(rows = vars(combining_method)) +
        ggtitle(sprintf(
          "Accuracy improvement of ensemble over
          the %s of networks vs %s.
          Ensemble size %s",
          if (ya == "acc_imp_max") "best" else "average", xa, sss
        ))
      print(cur_plot)
    }
  }
}

```

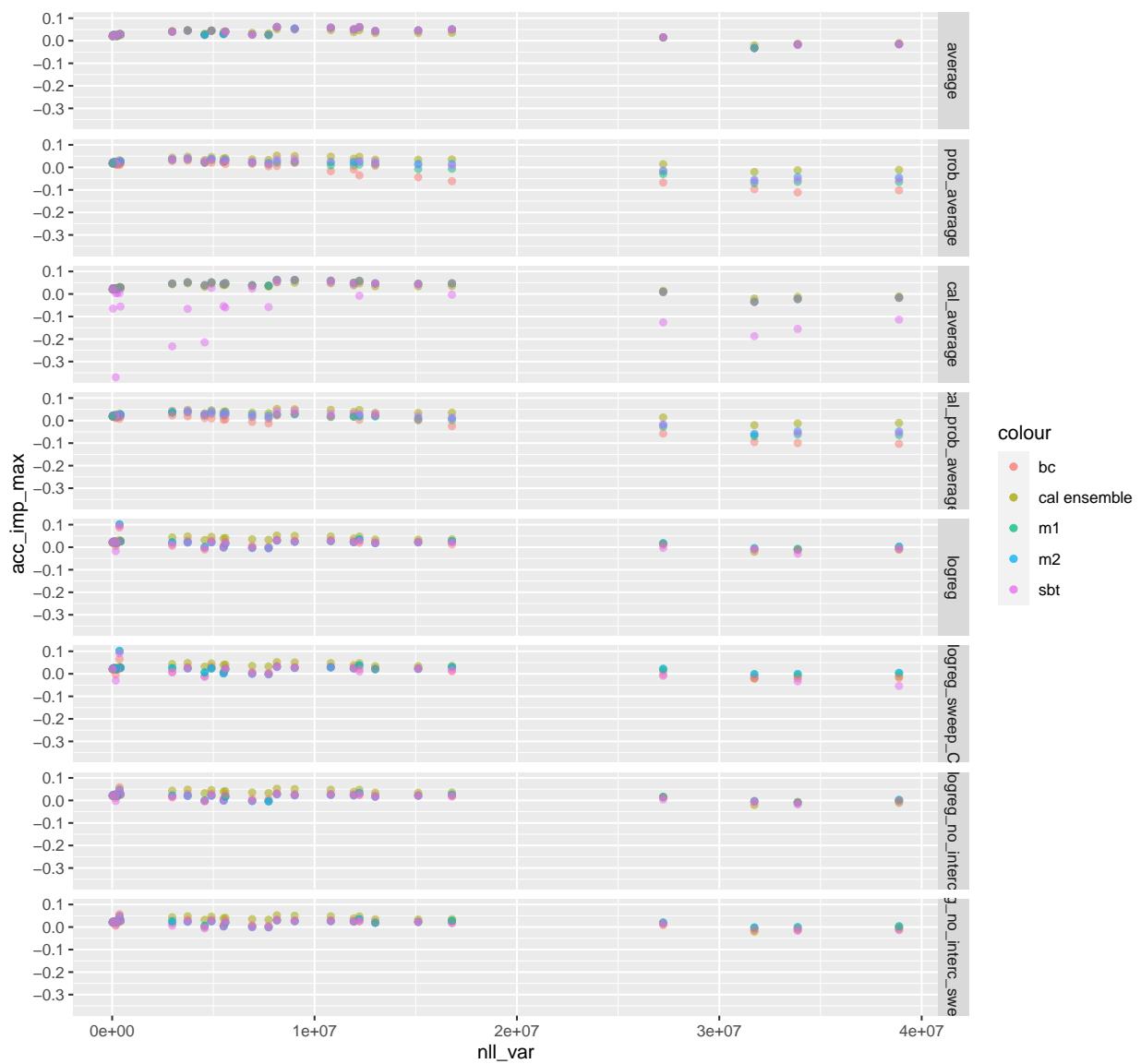
Accuracy improvement of ensemble over  
the best of networks vs err\_incons.  
Ensemble size 2



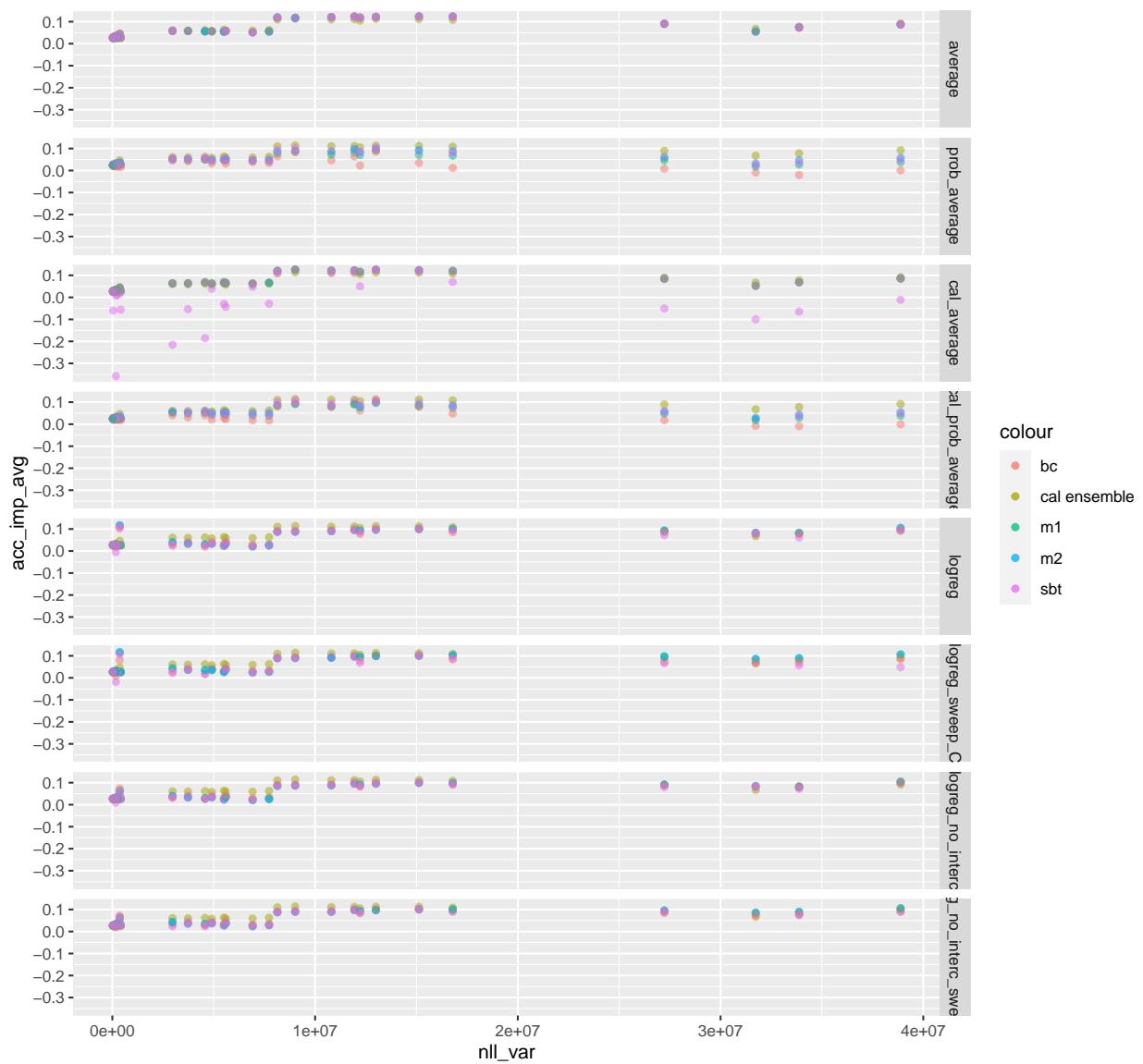
Accuracy improvement of ensemble over  
the average of networks vs err\_incons.  
Ensemble size 2



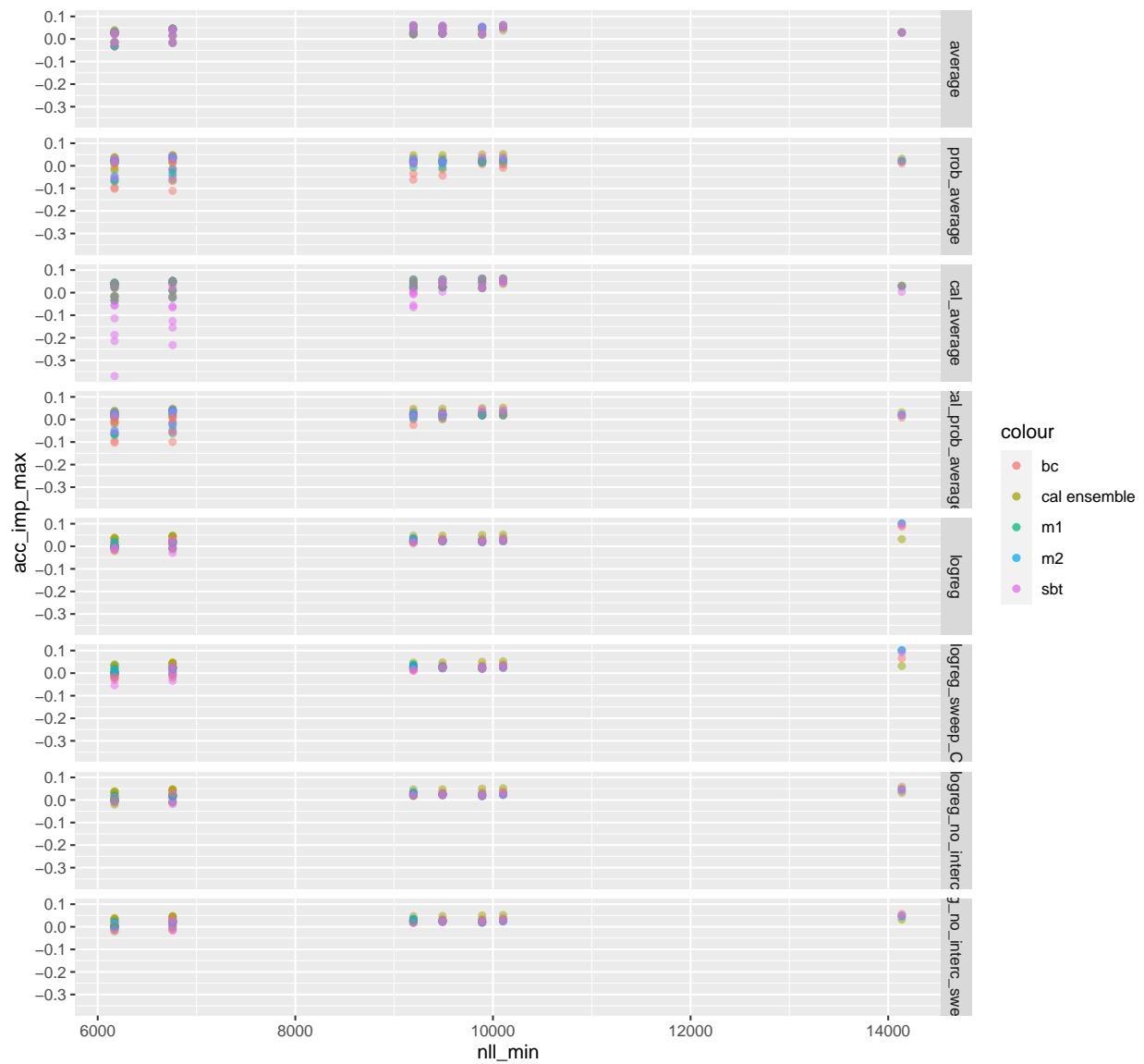
Accuracy improvement of ensemble over  
the best of networks vs nll\_var.  
Ensemble size 2



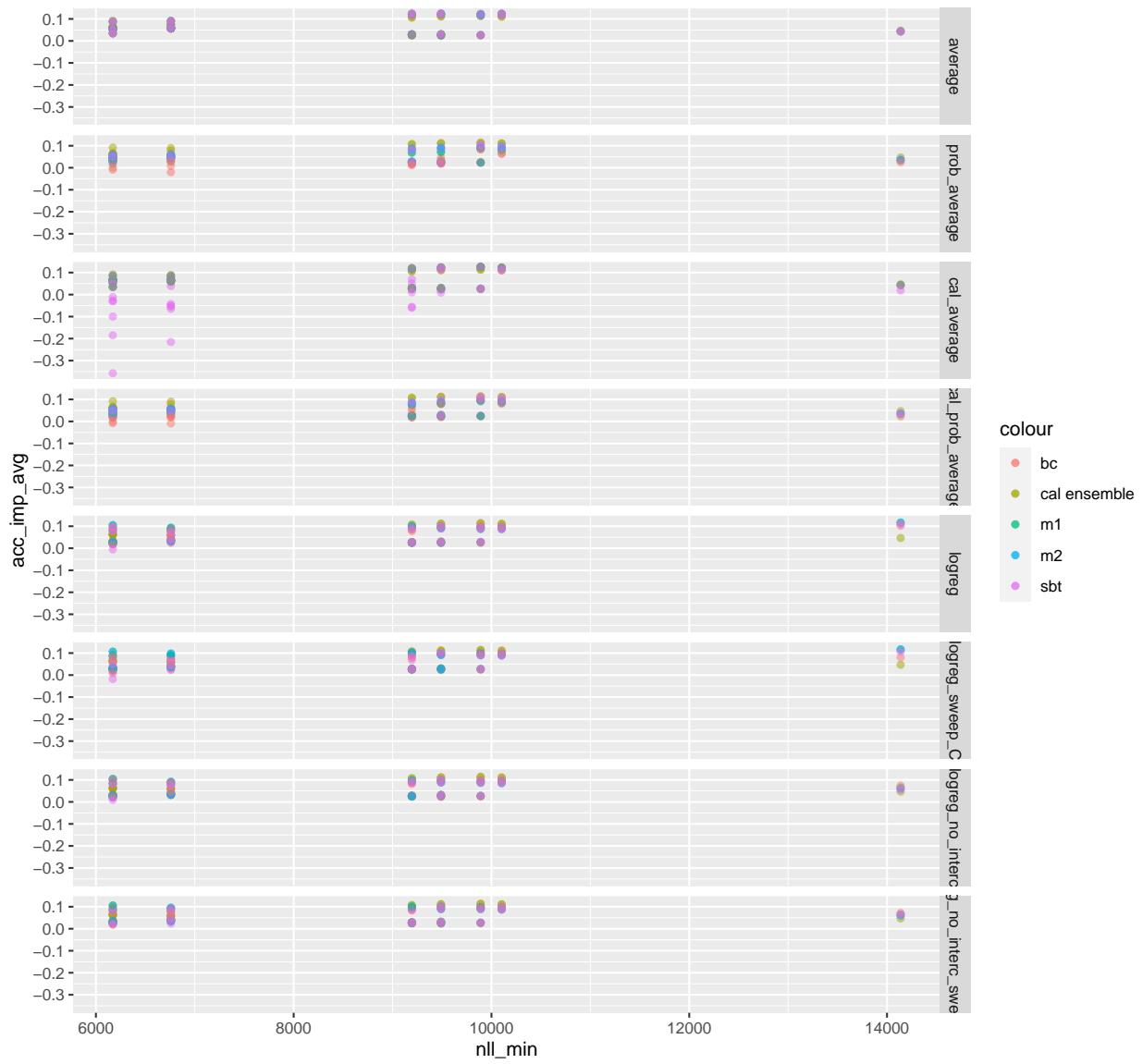
Accuracy improvement of ensemble over  
the average of networks vs nll\_var.  
Ensemble size 2



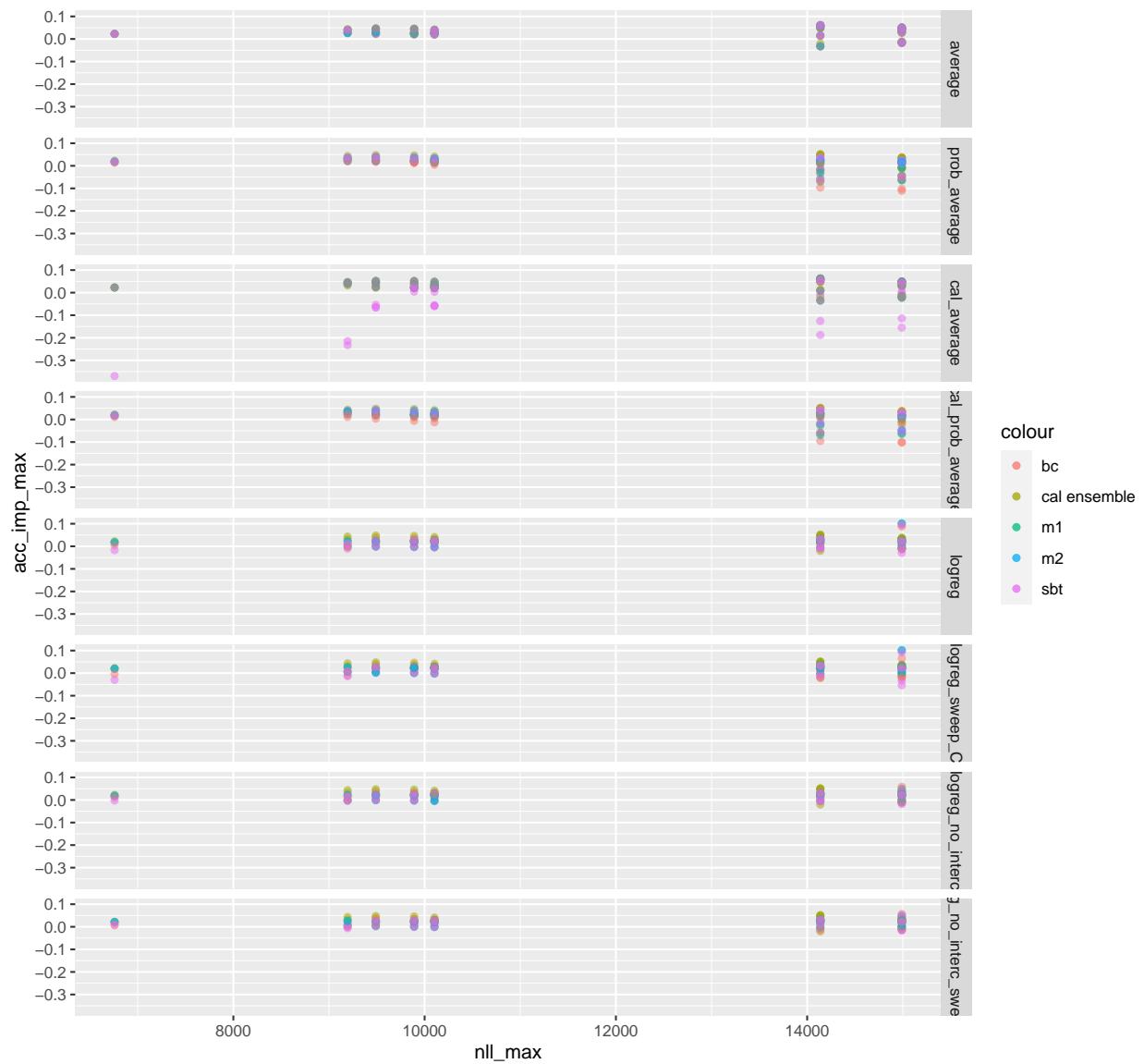
Accuracy improvement of ensemble over  
the best of networks vs nll\_min.  
Ensemble size 2



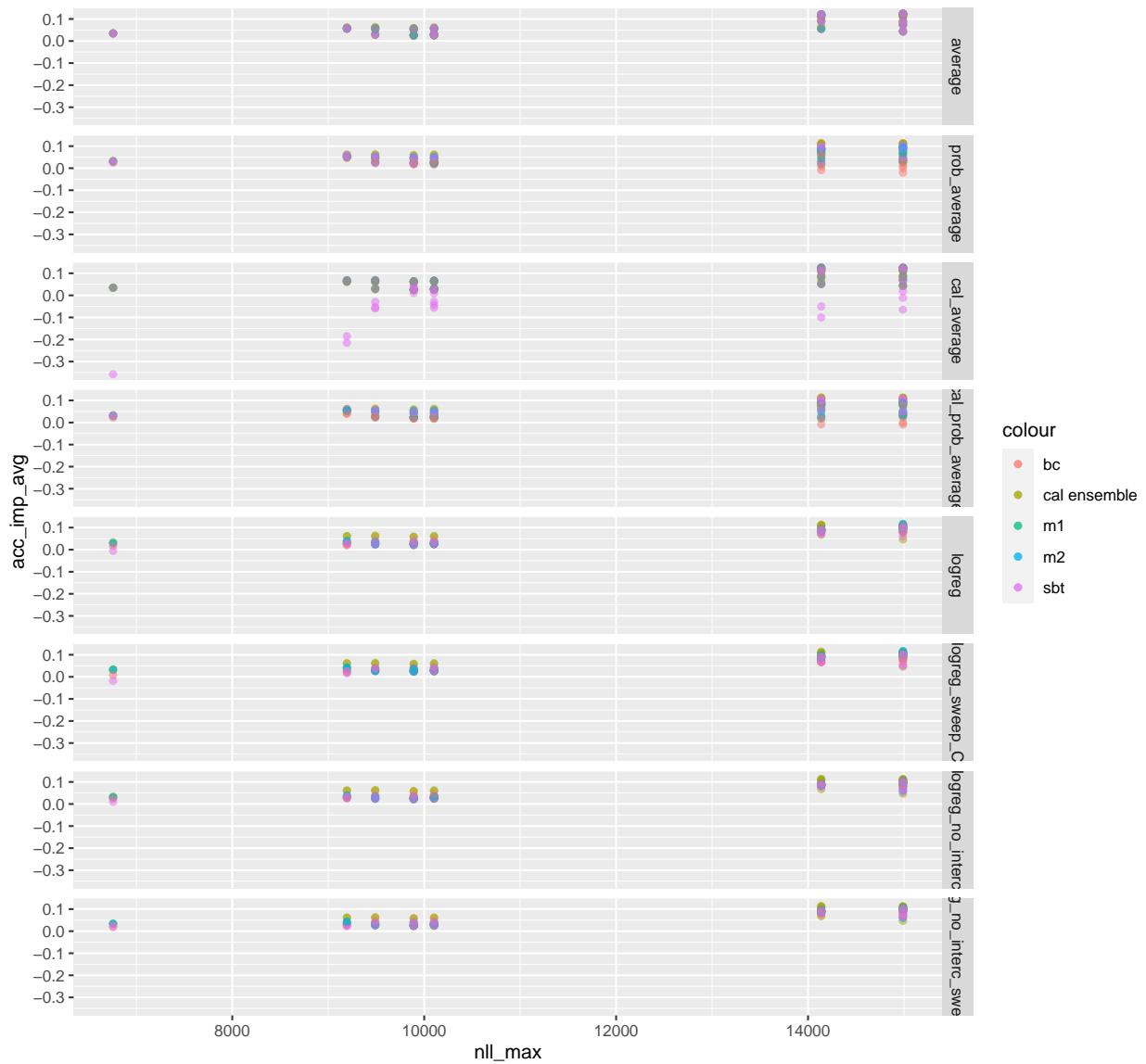
Accuracy improvement of ensemble over  
the average of networks vs nll\_min.  
Ensemble size 2



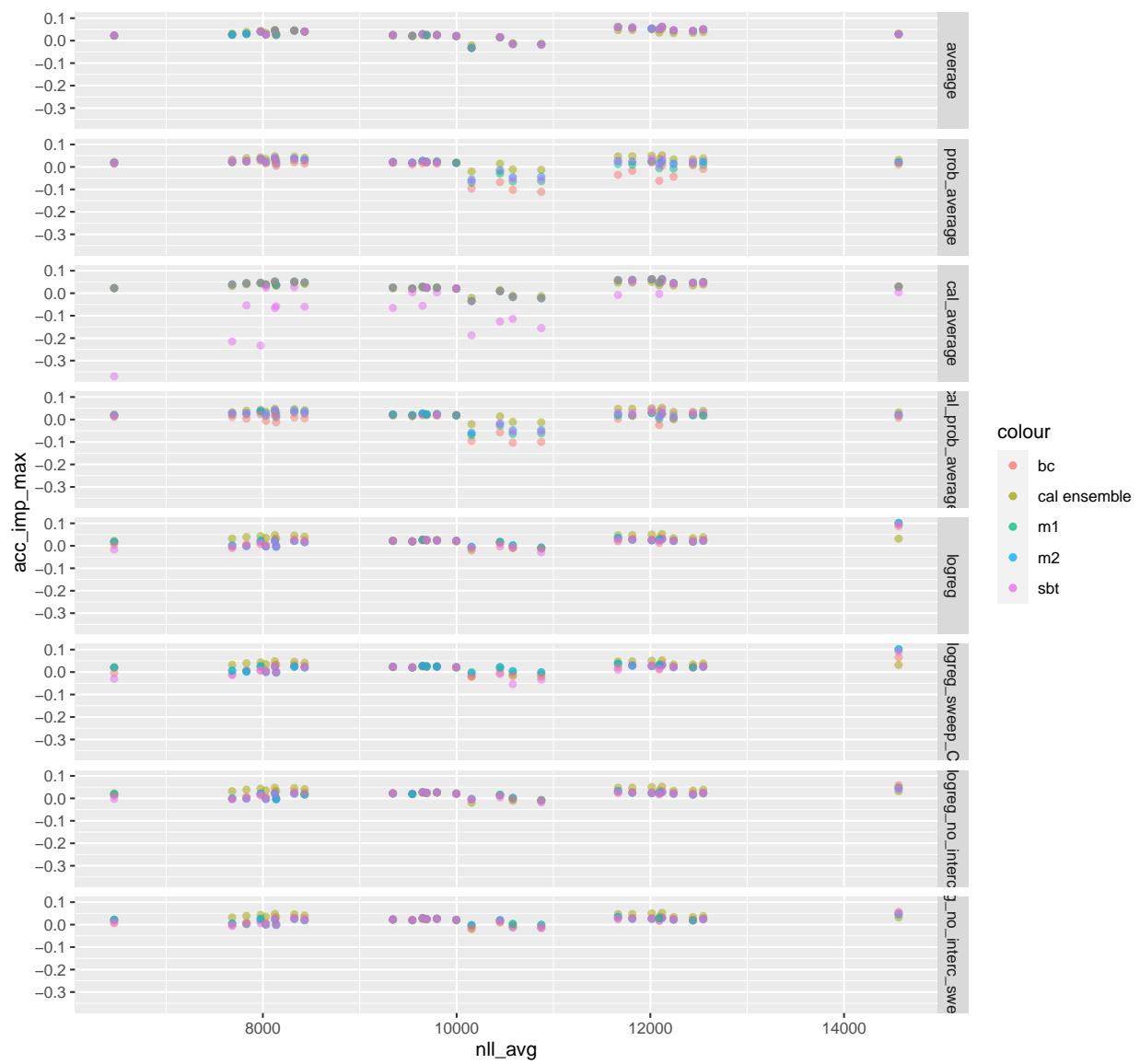
Accuracy improvement of ensemble over  
the best of networks vs nll\_max.  
Ensemble size 2



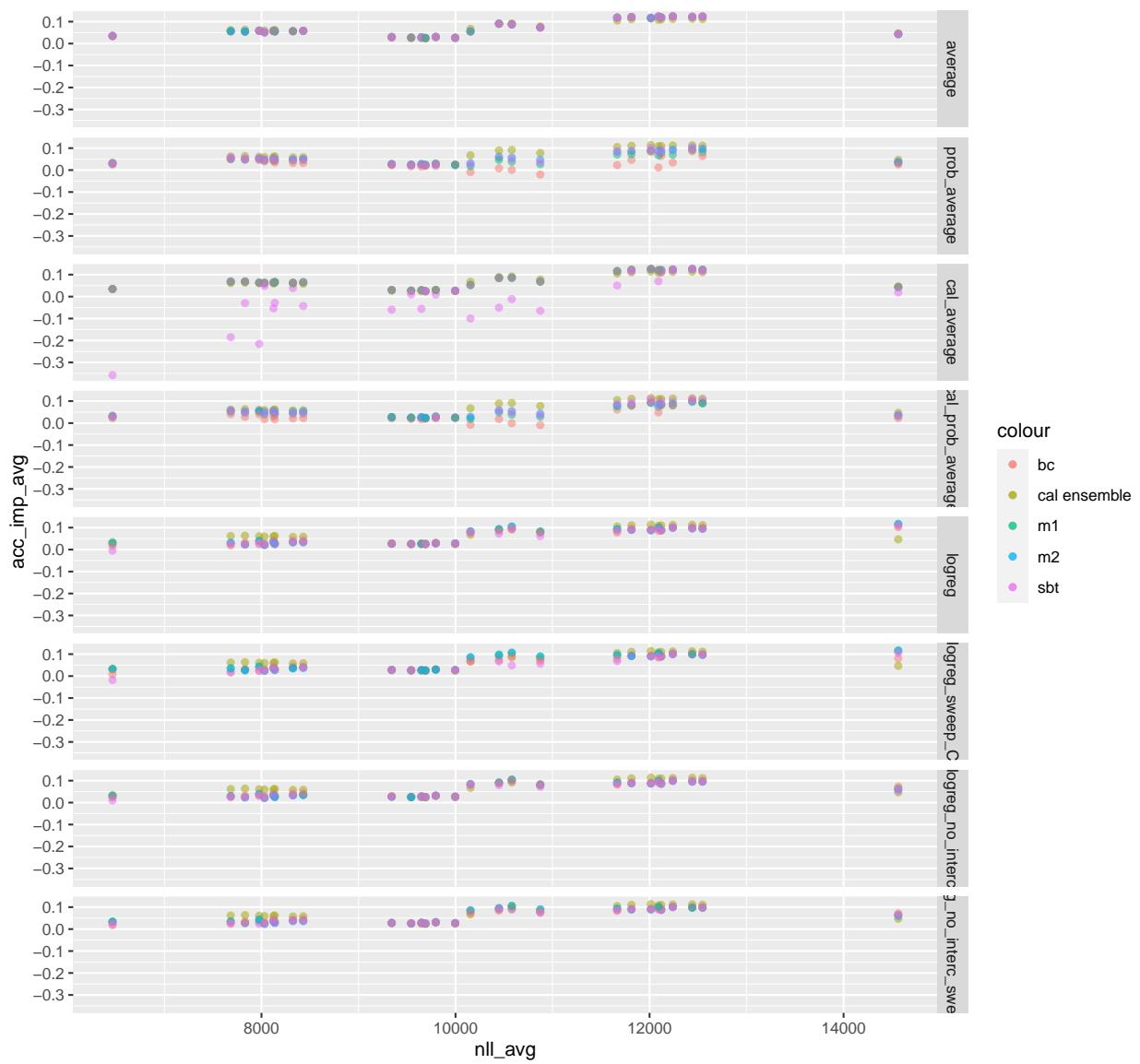
Accuracy improvement of ensemble over  
the average of networks vs nll\_max.  
Ensemble size 2



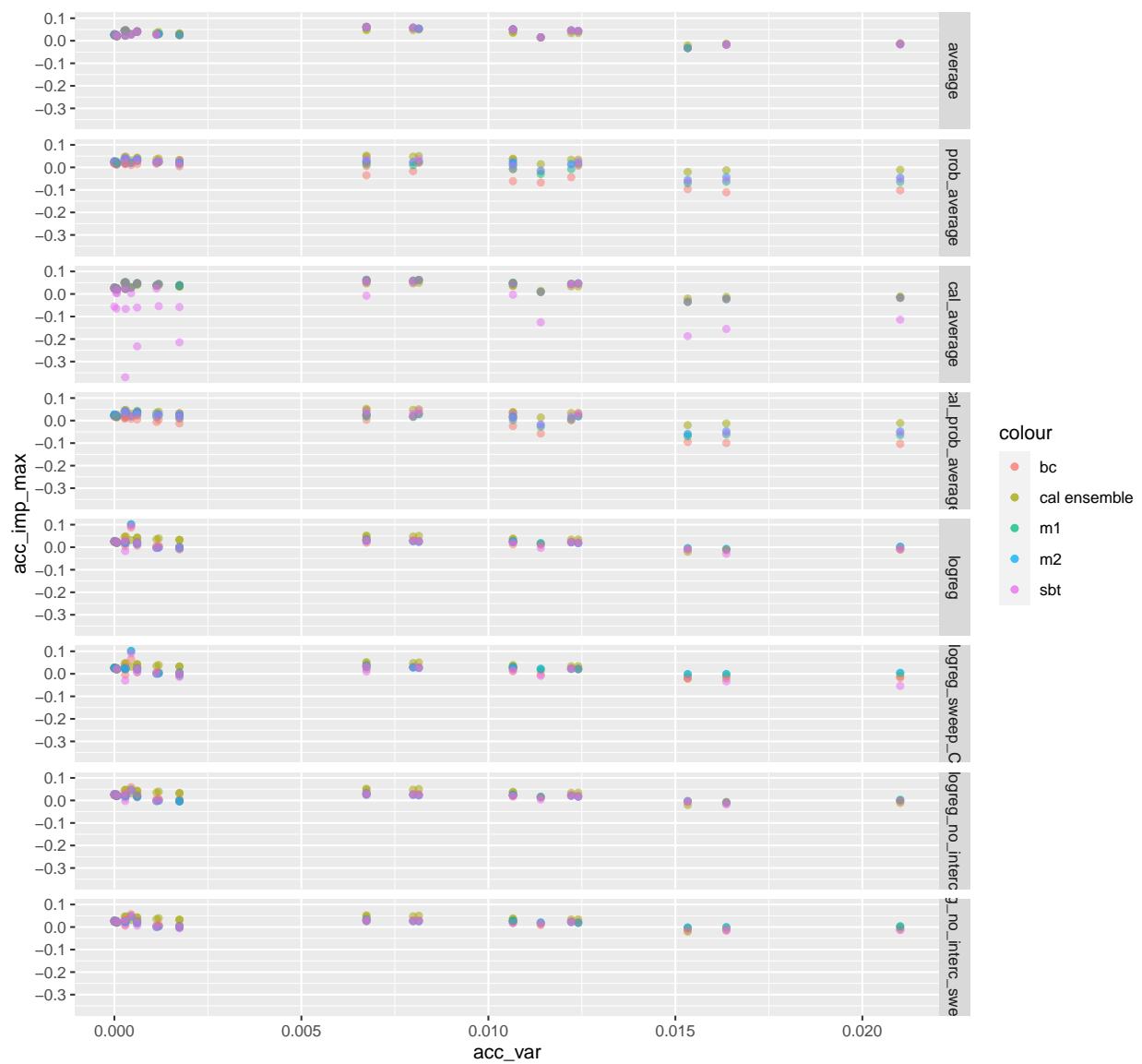
Accuracy improvement of ensemble over  
the best of networks vs nll\_avg.  
Ensemble size 2



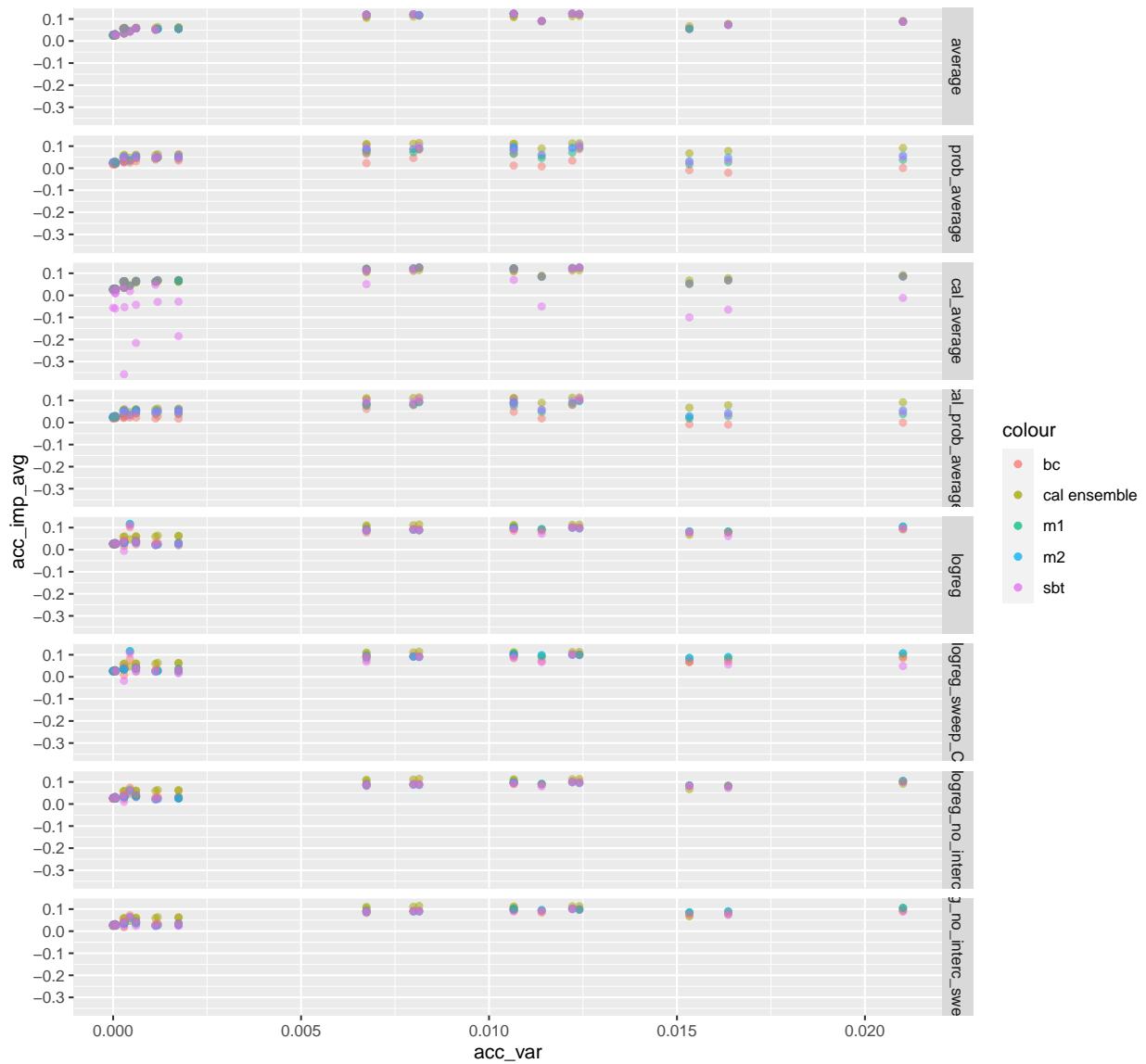
Accuracy improvement of ensemble over  
the average of networks vs nll\_avg.  
Ensemble size 2



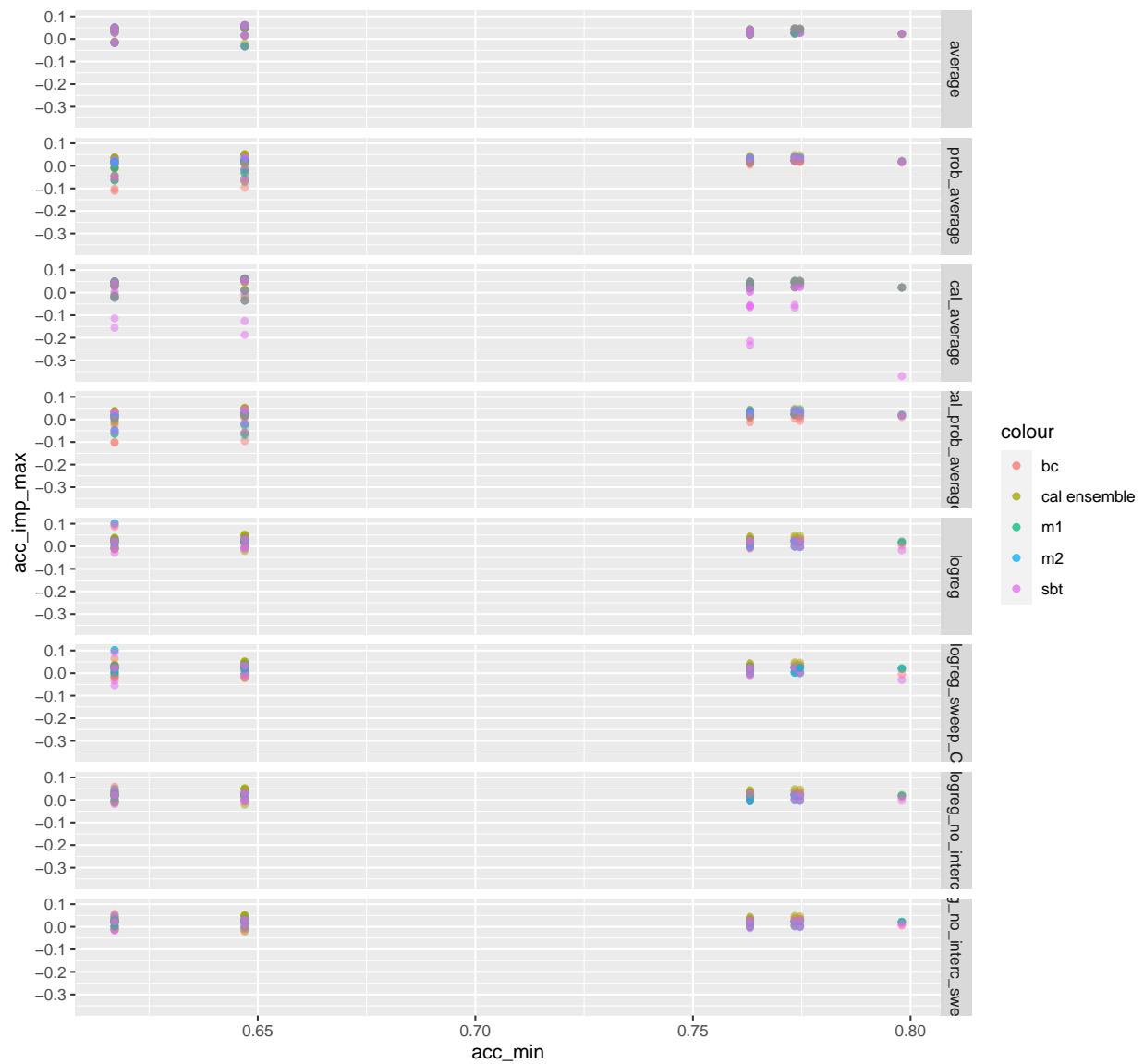
Accuracy improvement of ensemble over  
the best of networks vs acc\_var.  
Ensemble size 2



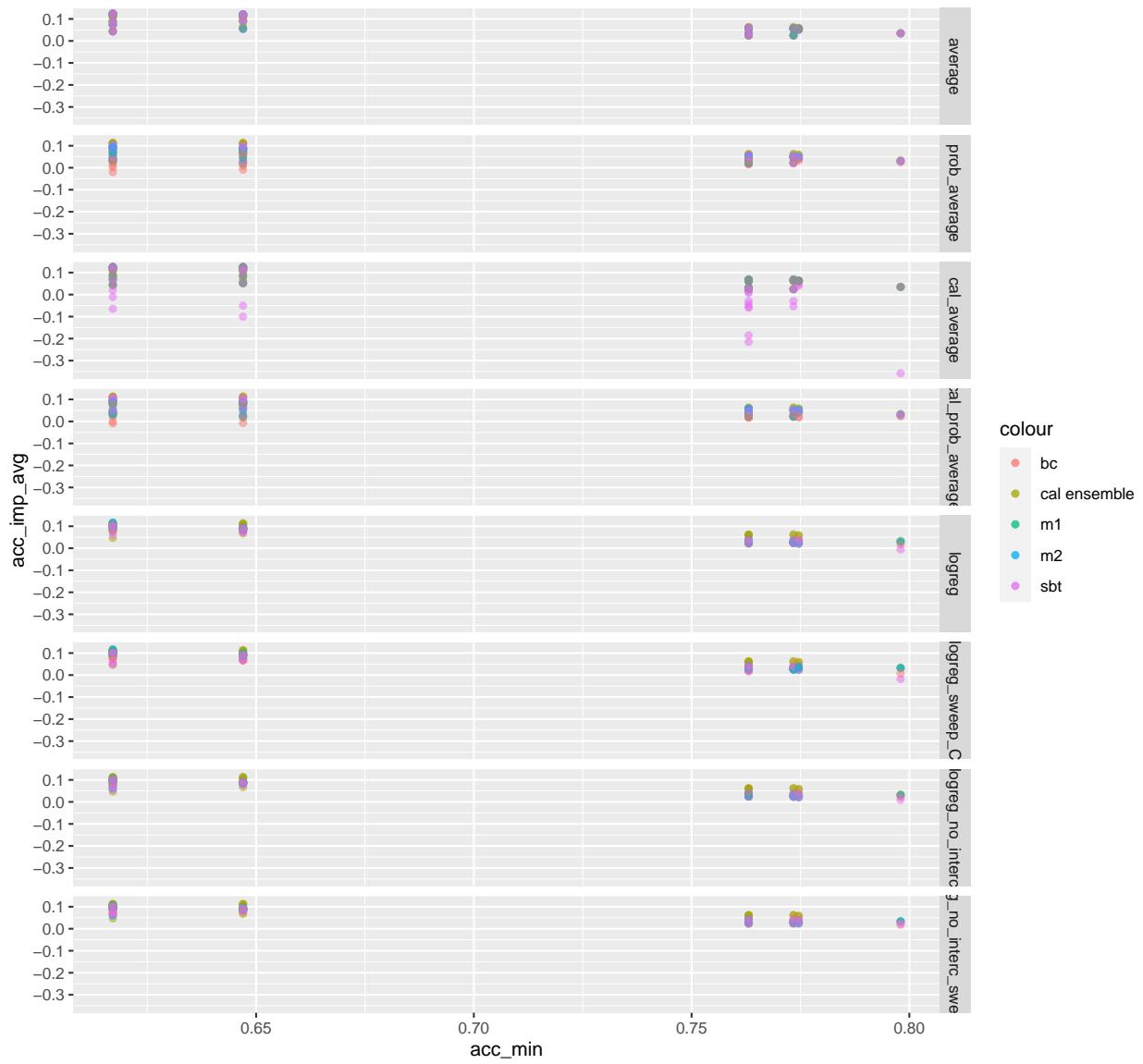
Accuracy improvement of ensemble over  
the average of networks vs acc\_var.  
Ensemble size 2



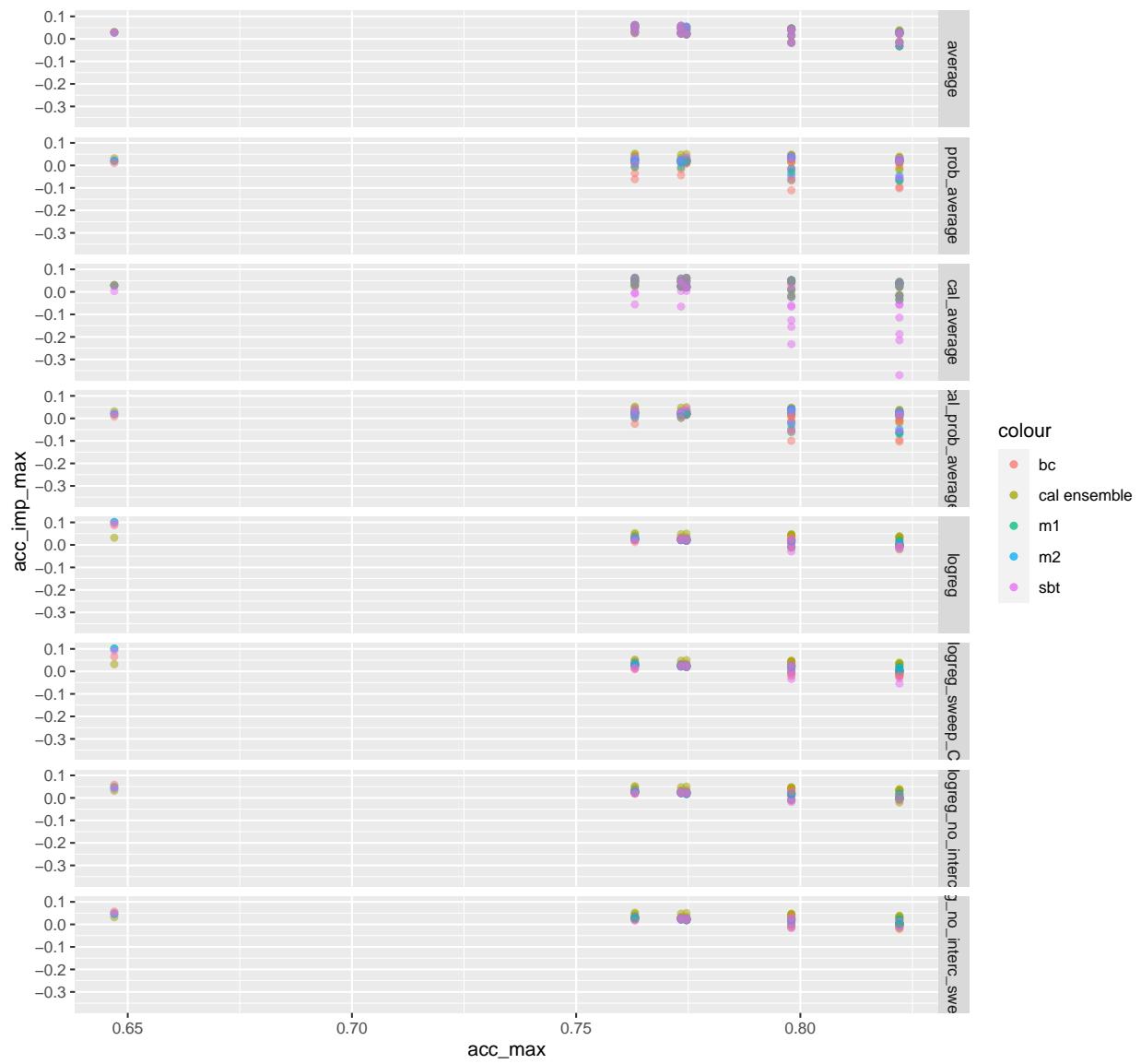
Accuracy improvement of ensemble over  
the best of networks vs acc\_min.  
Ensemble size 2



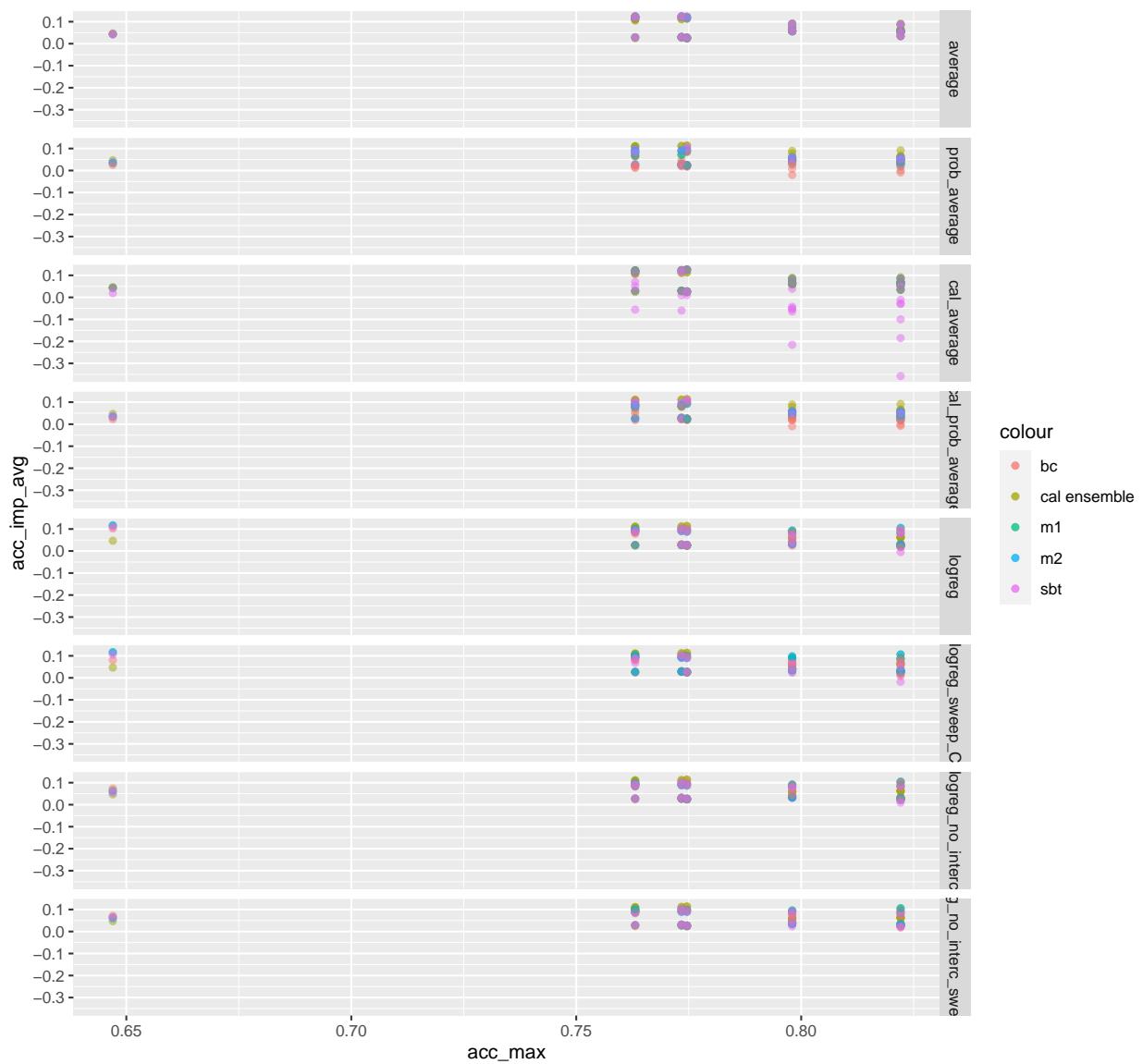
Accuracy improvement of ensemble over  
the average of networks vs acc\_min.  
Ensemble size 2



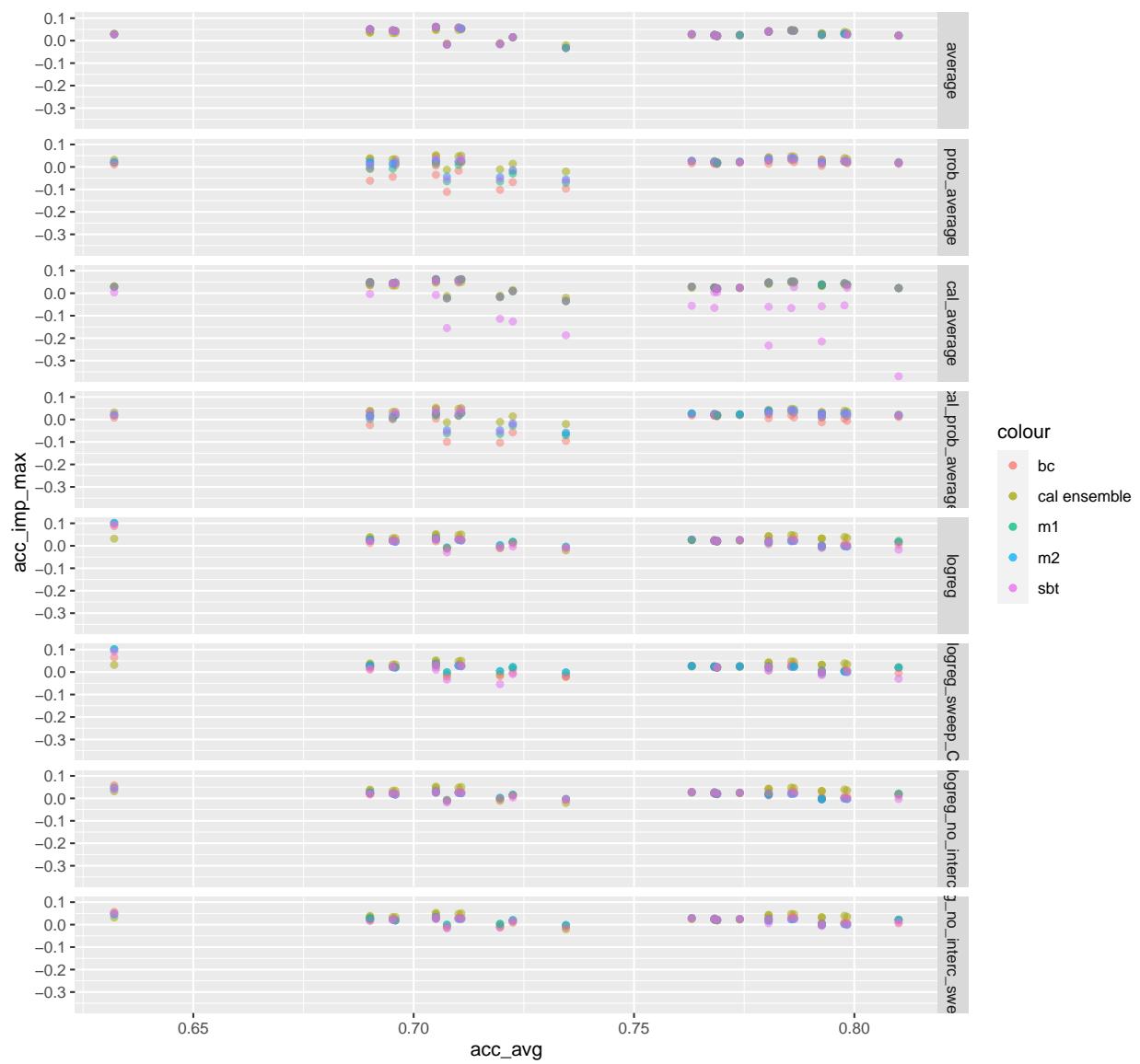
Accuracy improvement of ensemble over  
the best of networks vs acc\_max.  
Ensemble size 2



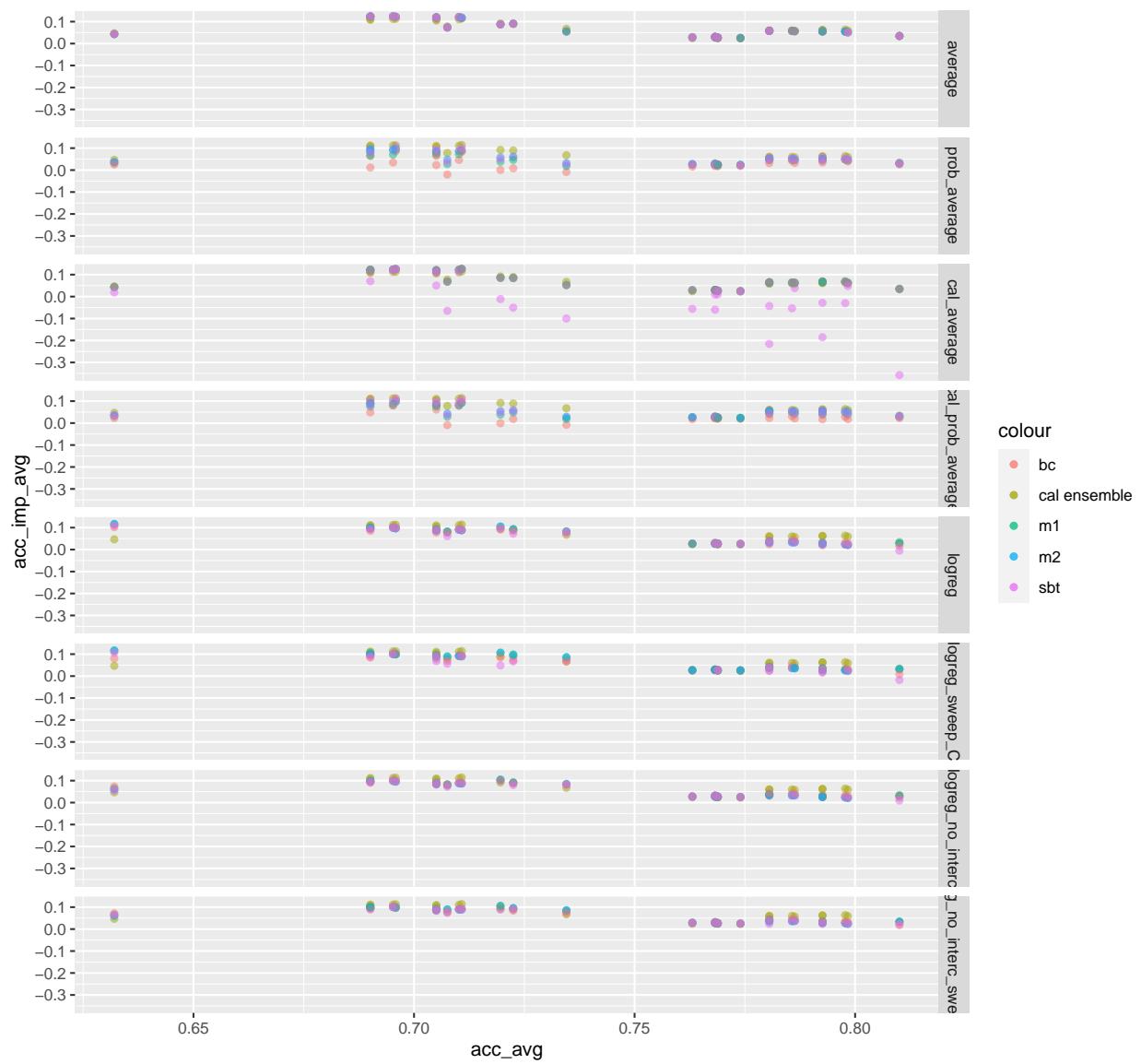
Accuracy improvement of ensemble over  
the average of networks vs acc\_max.  
Ensemble size 2



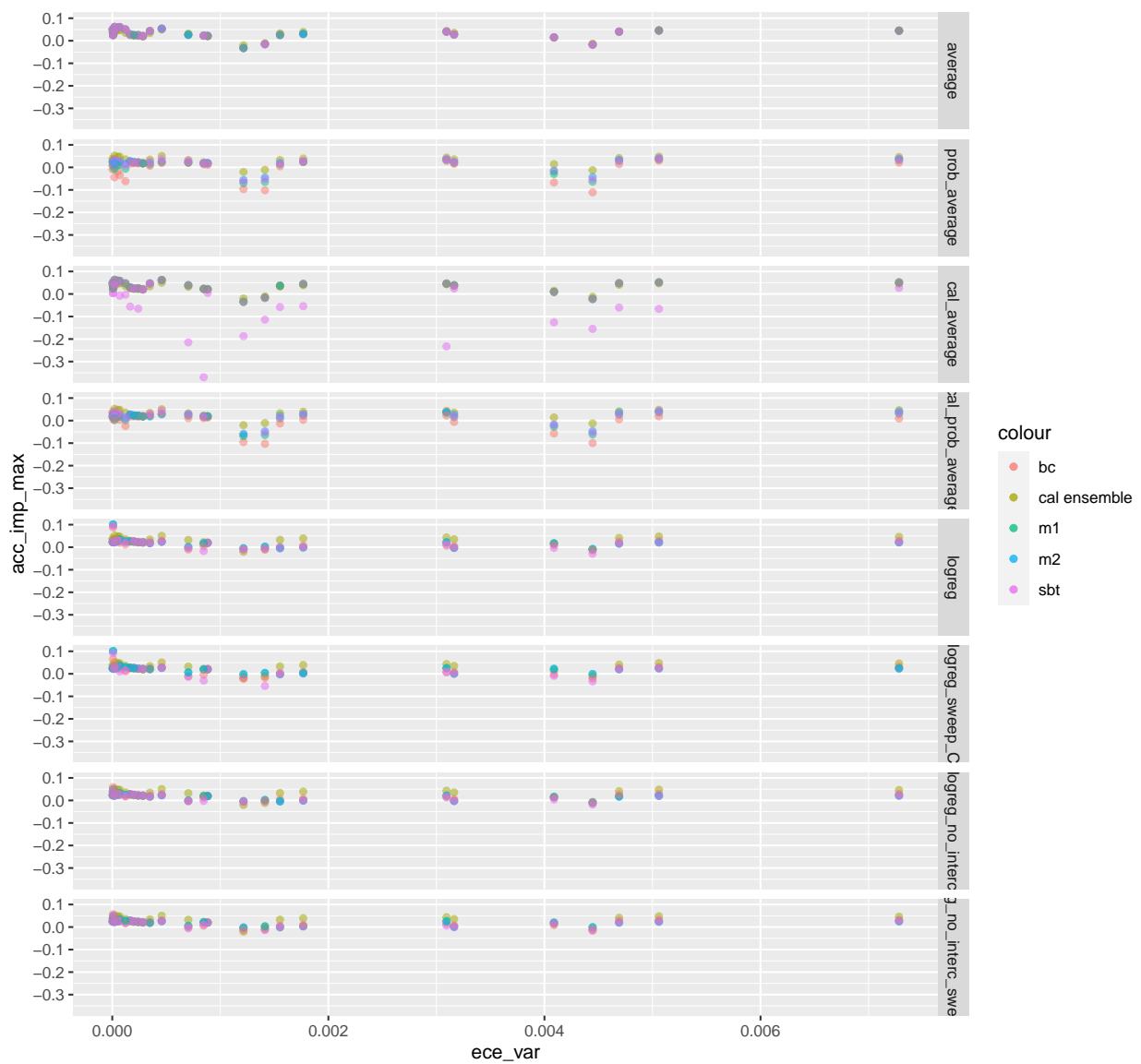
Accuracy improvement of ensemble over  
the best of networks vs acc\_avg.  
Ensemble size 2



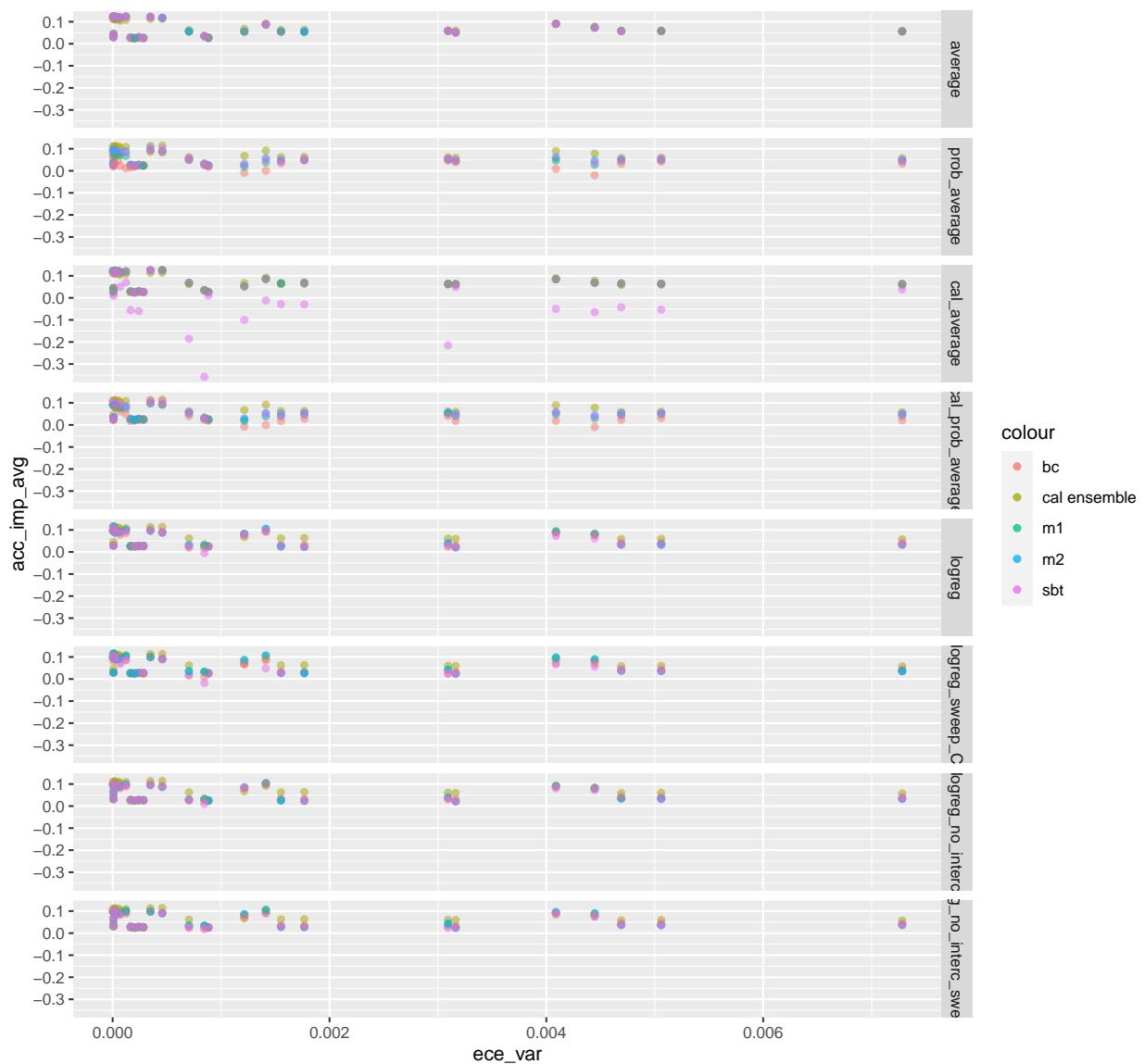
Accuracy improvement of ensemble over  
the average of networks vs acc\_avg.  
Ensemble size 2



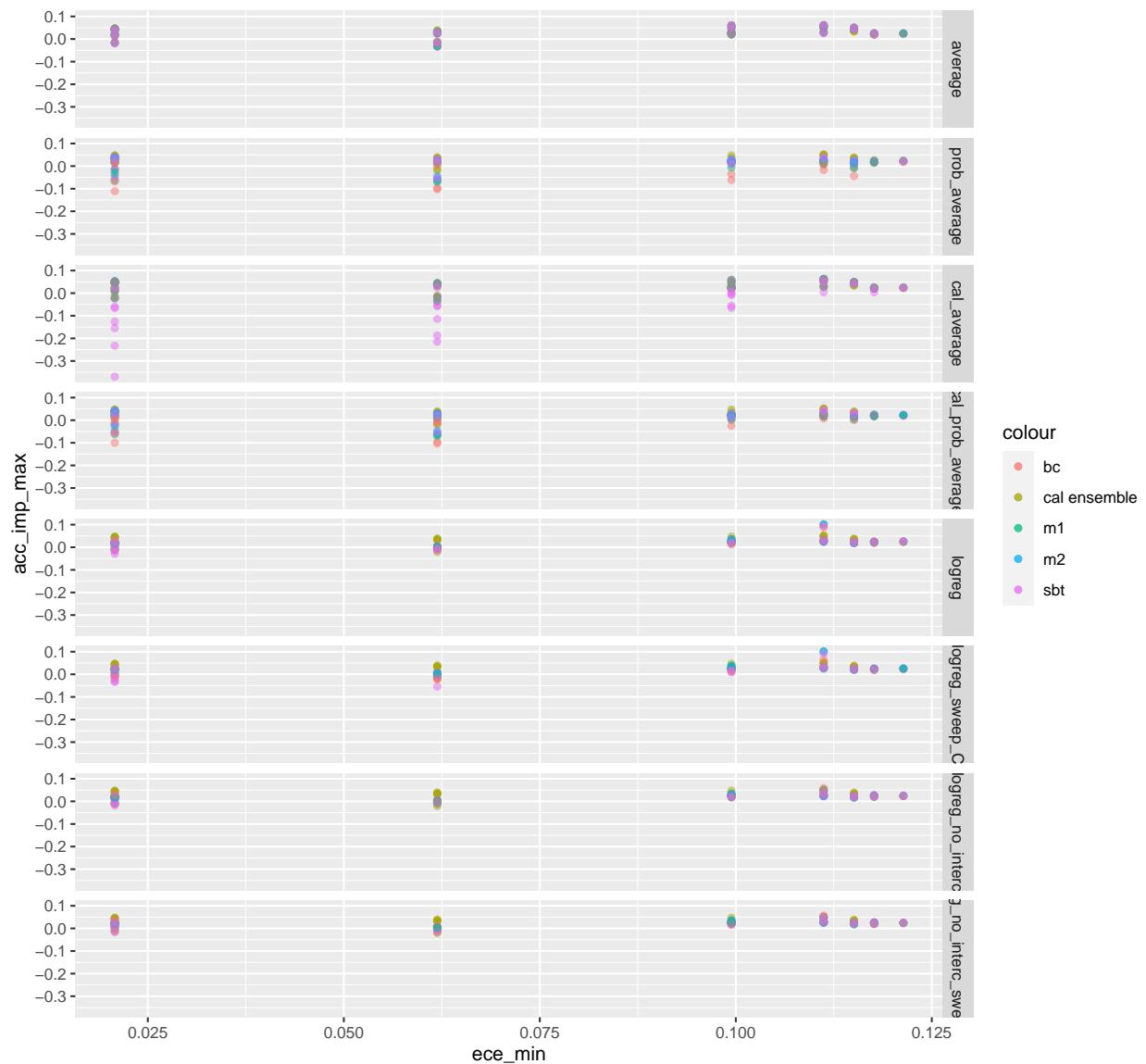
Accuracy improvement of ensemble over  
the best of networks vs ece\_var.  
Ensemble size 2



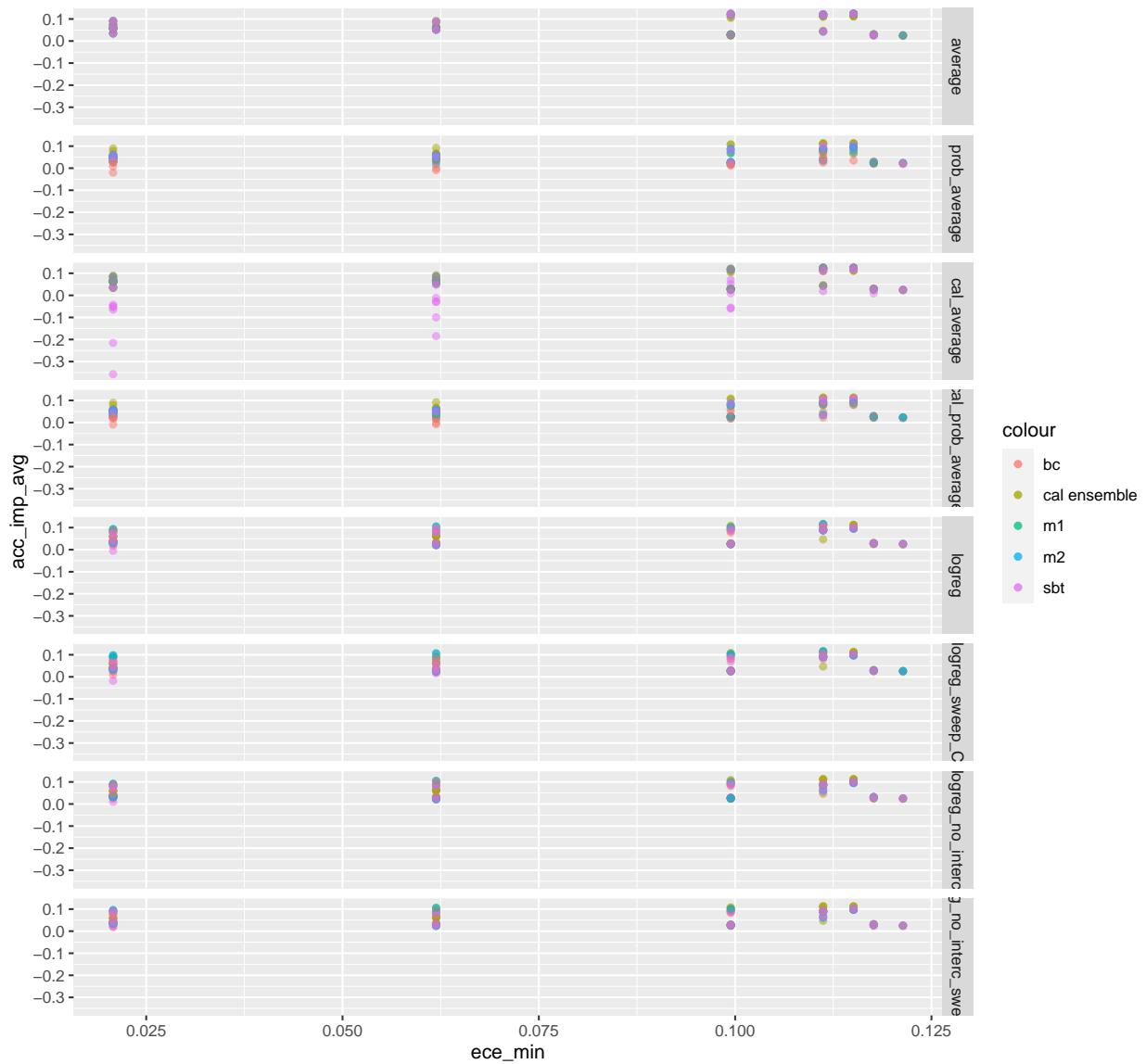
Accuracy improvement of ensemble over  
the average of networks vs ece\_var.  
Ensemble size 2



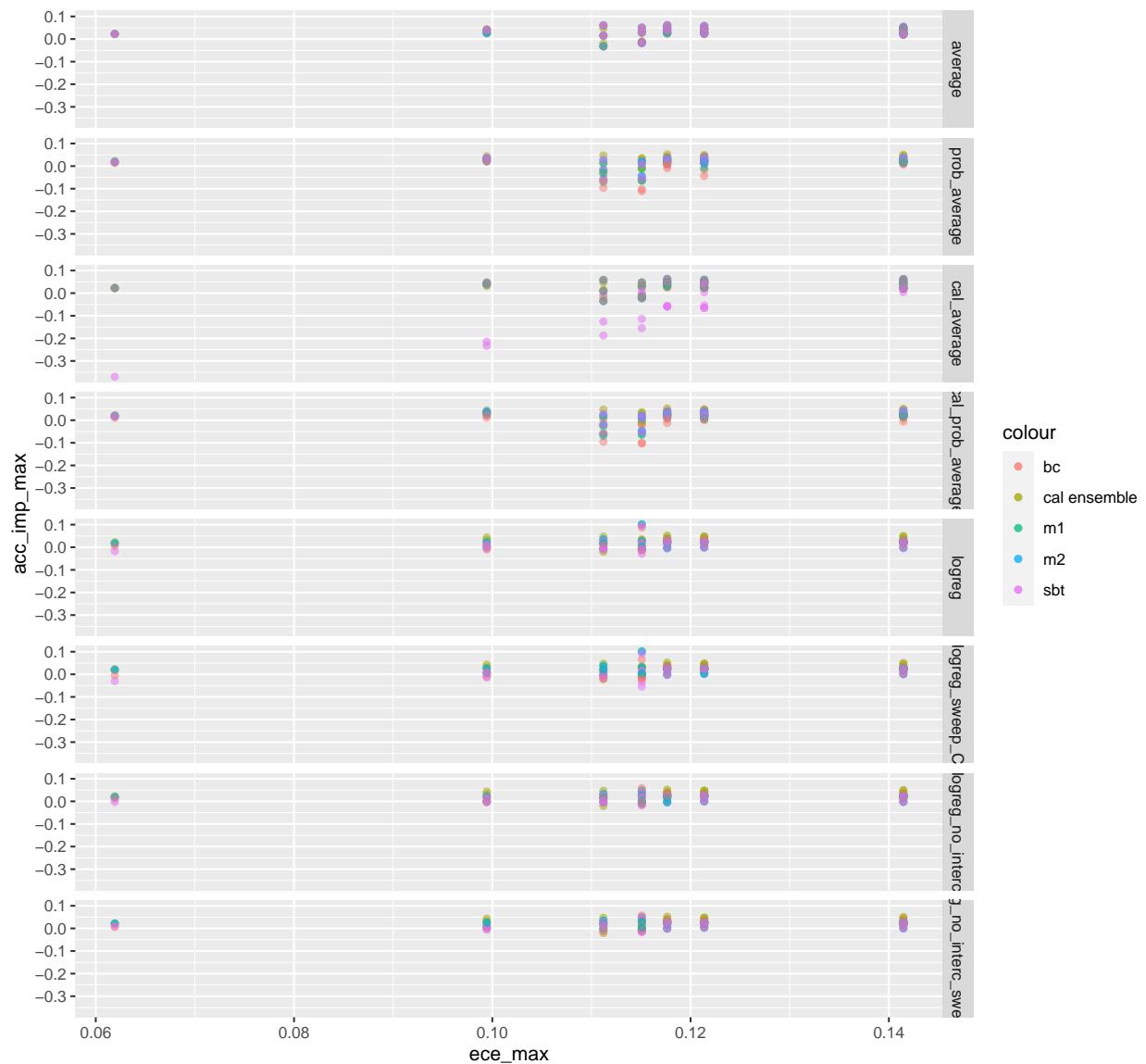
Accuracy improvement of ensemble over  
the best of networks vs ece\_min.  
Ensemble size 2



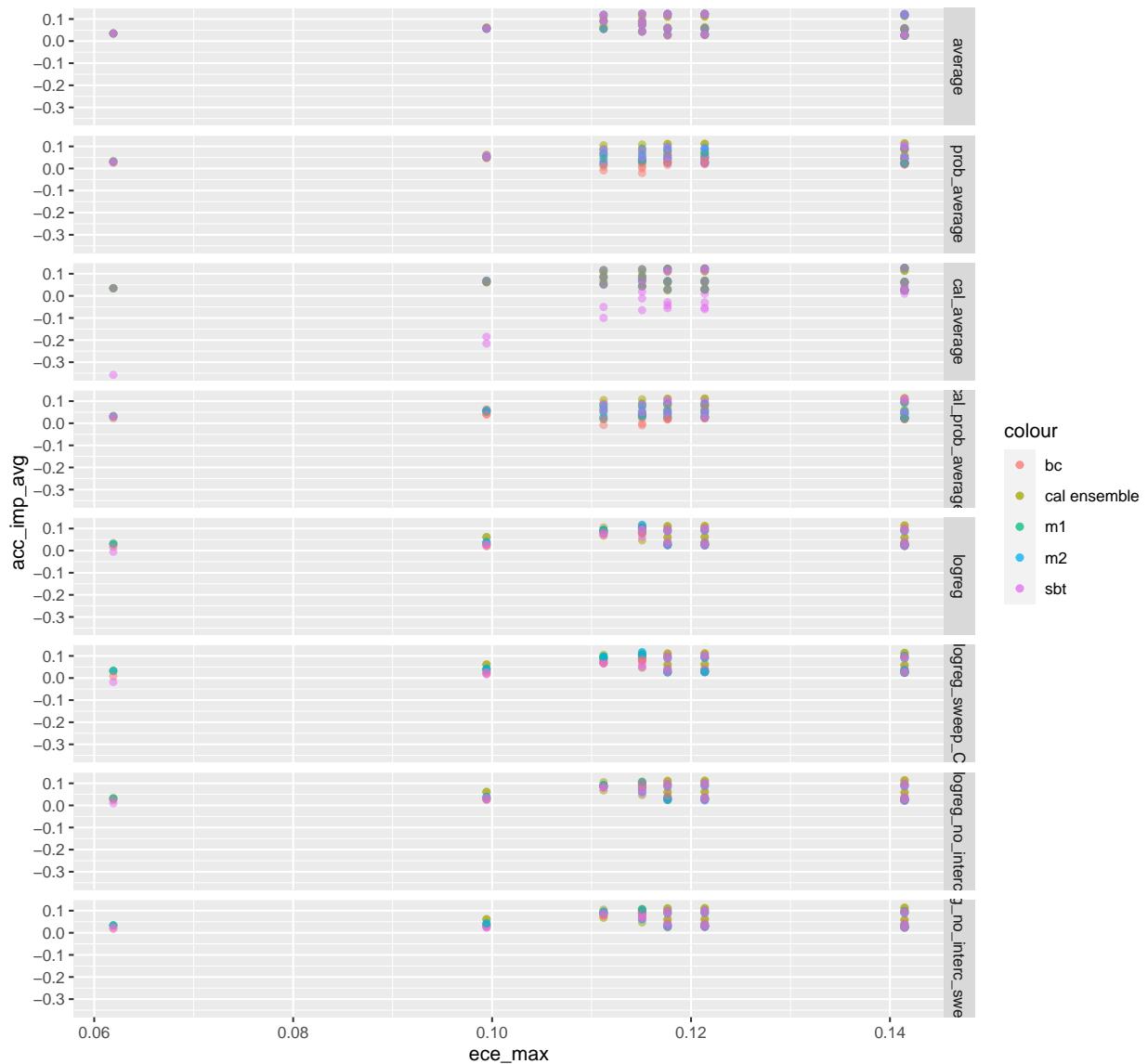
Accuracy improvement of ensemble over  
the average of networks vs ece\_min.  
Ensemble size 2



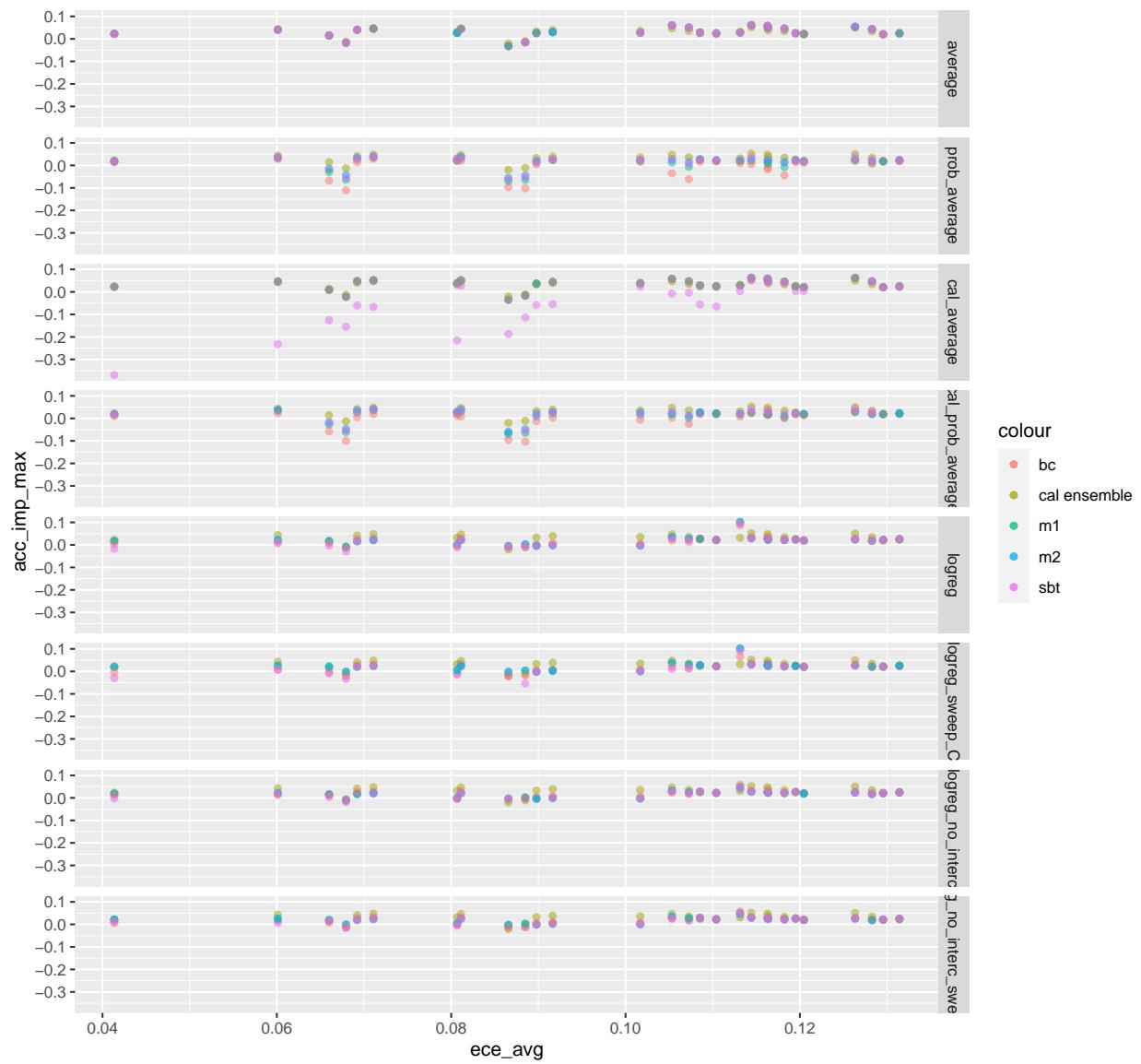
Accuracy improvement of ensemble over  
the best of networks vs ece\_max.  
Ensemble size 2



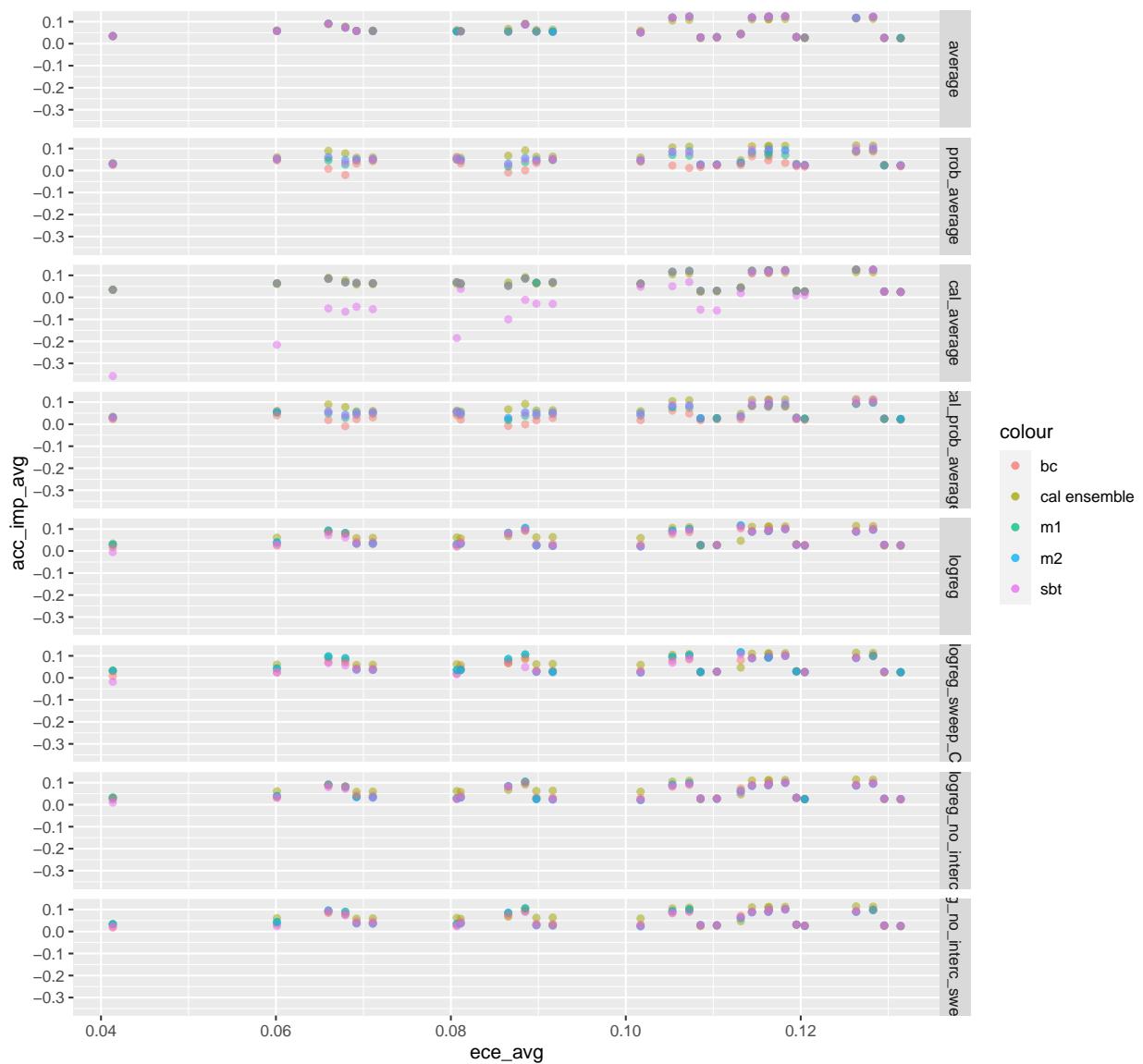
Accuracy improvement of ensemble over  
the average of networks vs ece\_max.  
Ensemble size 2



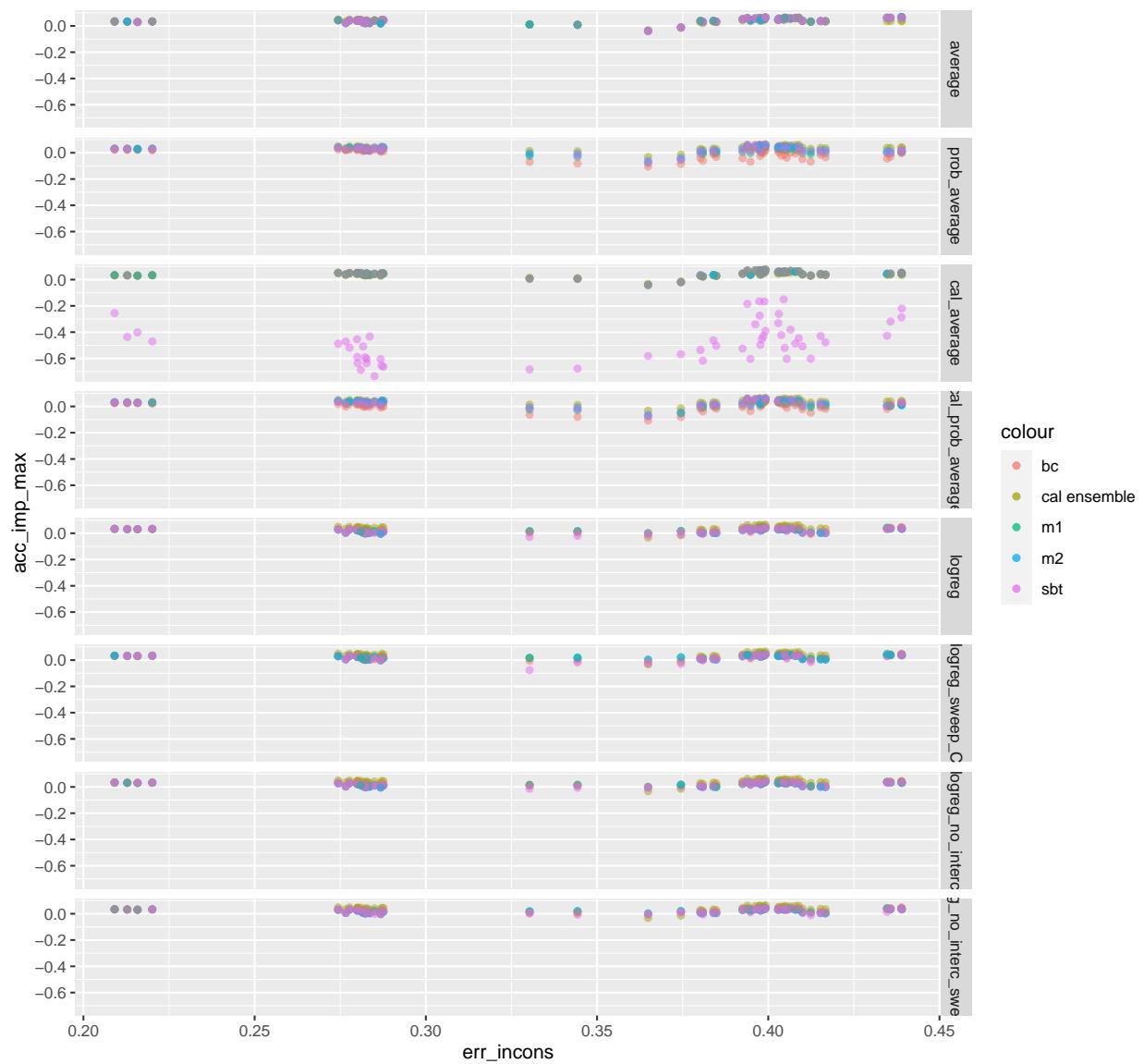
Accuracy improvement of ensemble over  
the best of networks vs ece\_avg.  
Ensemble size 2



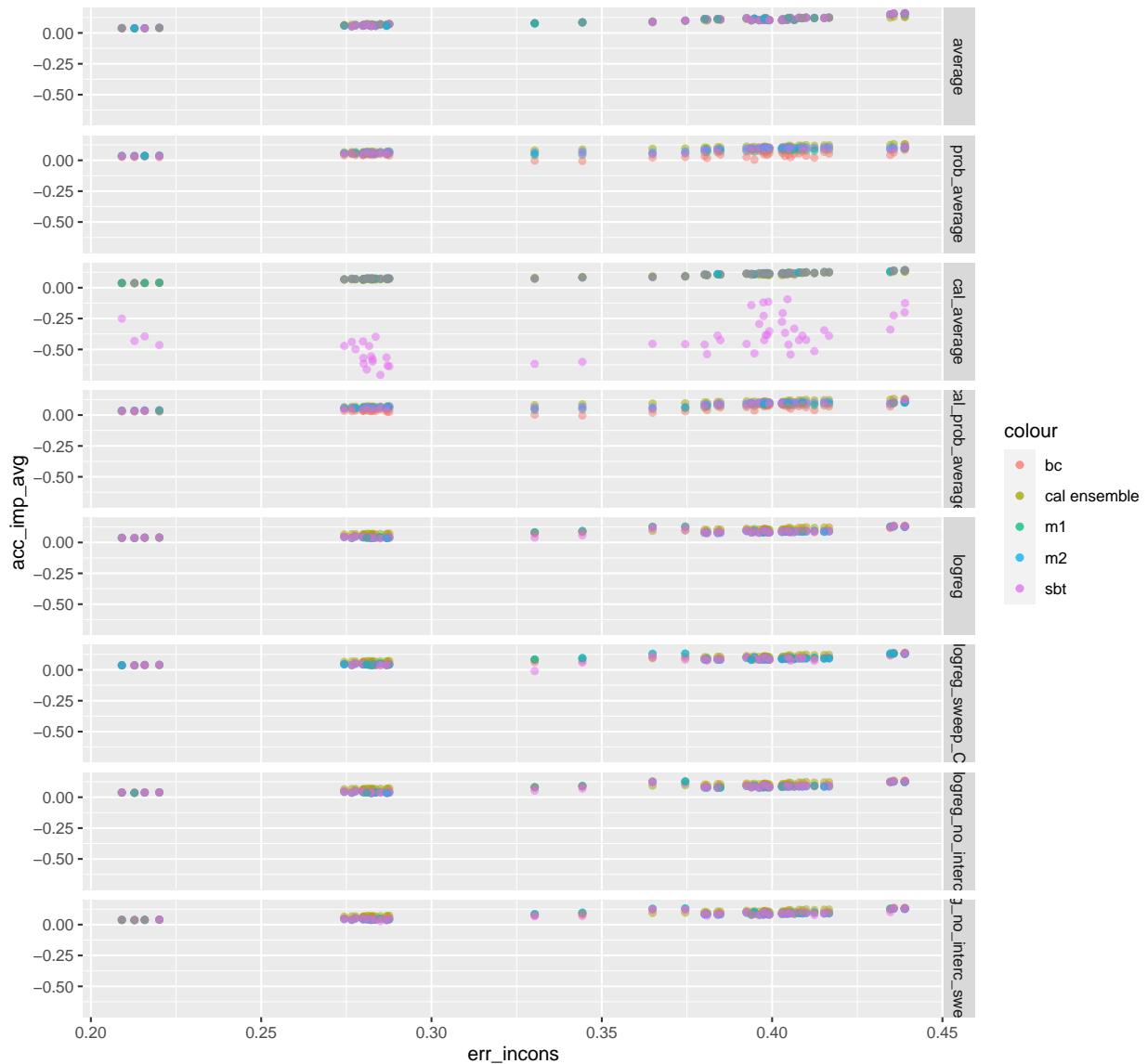
Accuracy improvement of ensemble over  
the average of networks vs ece\_avg.  
Ensemble size 2



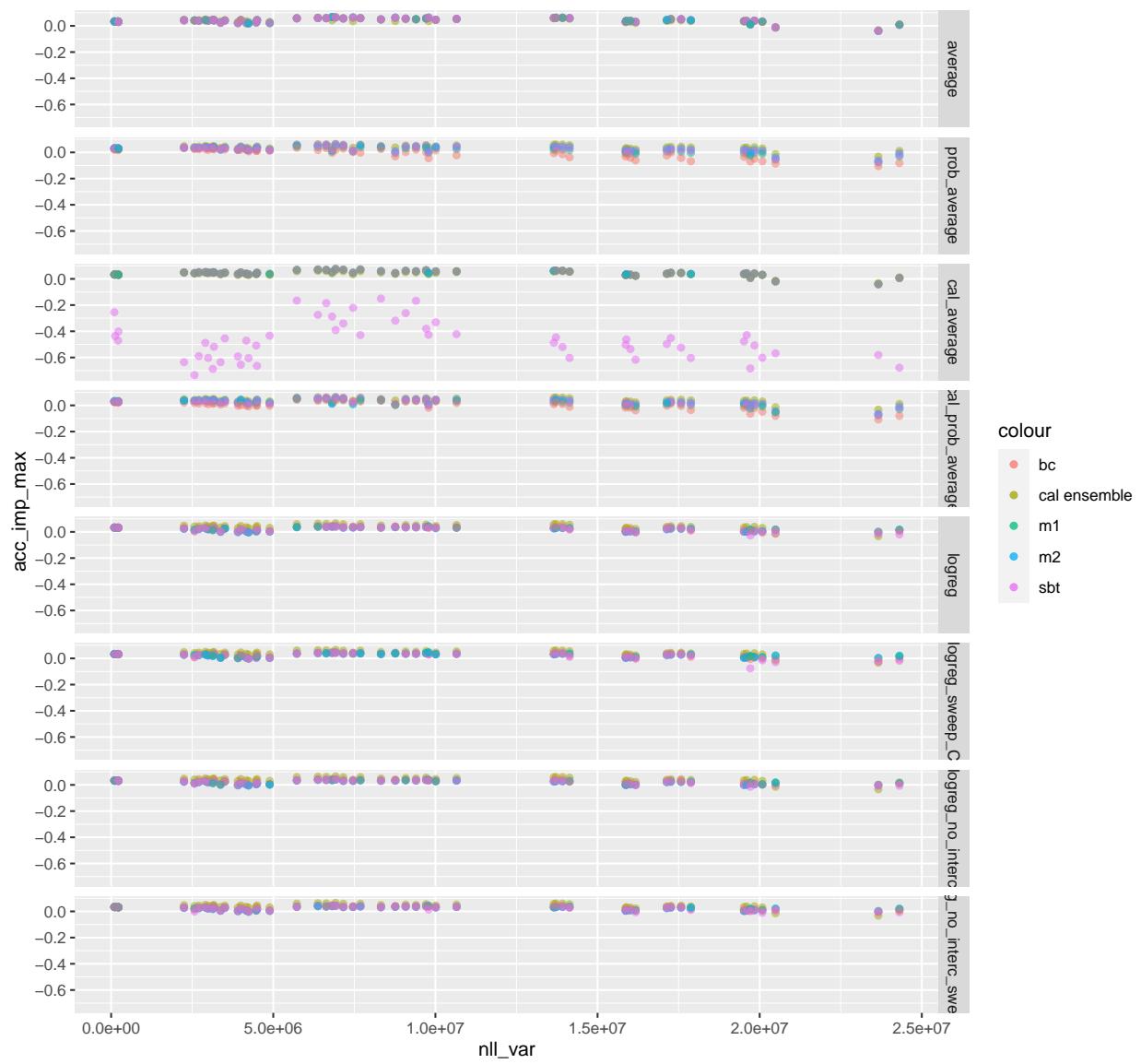
Accuracy improvement of ensemble over  
the best of networks vs err\_incons.  
Ensemble size 3



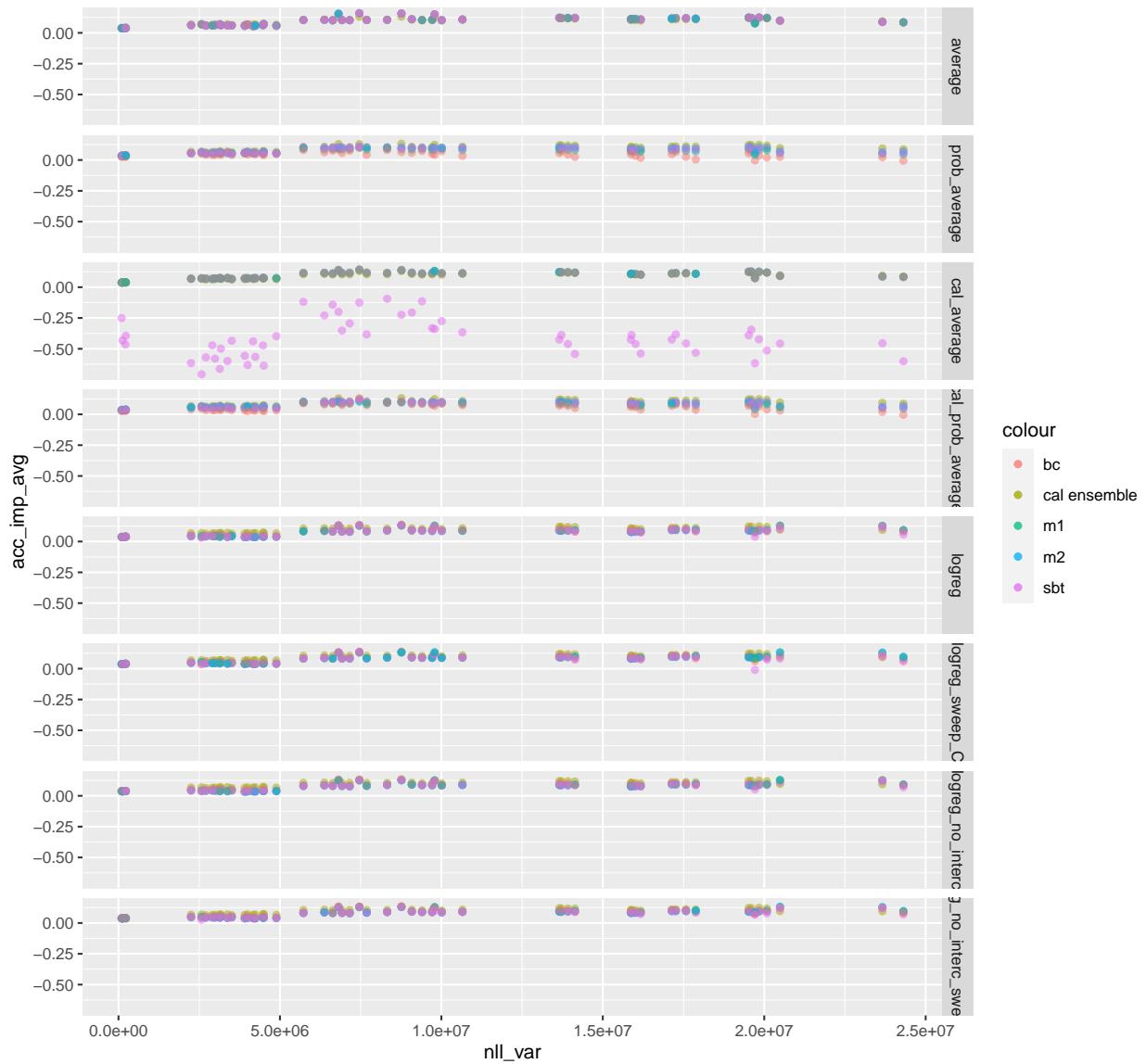
Accuracy improvement of ensemble over  
the average of networks vs err\_incons.  
Ensemble size 3



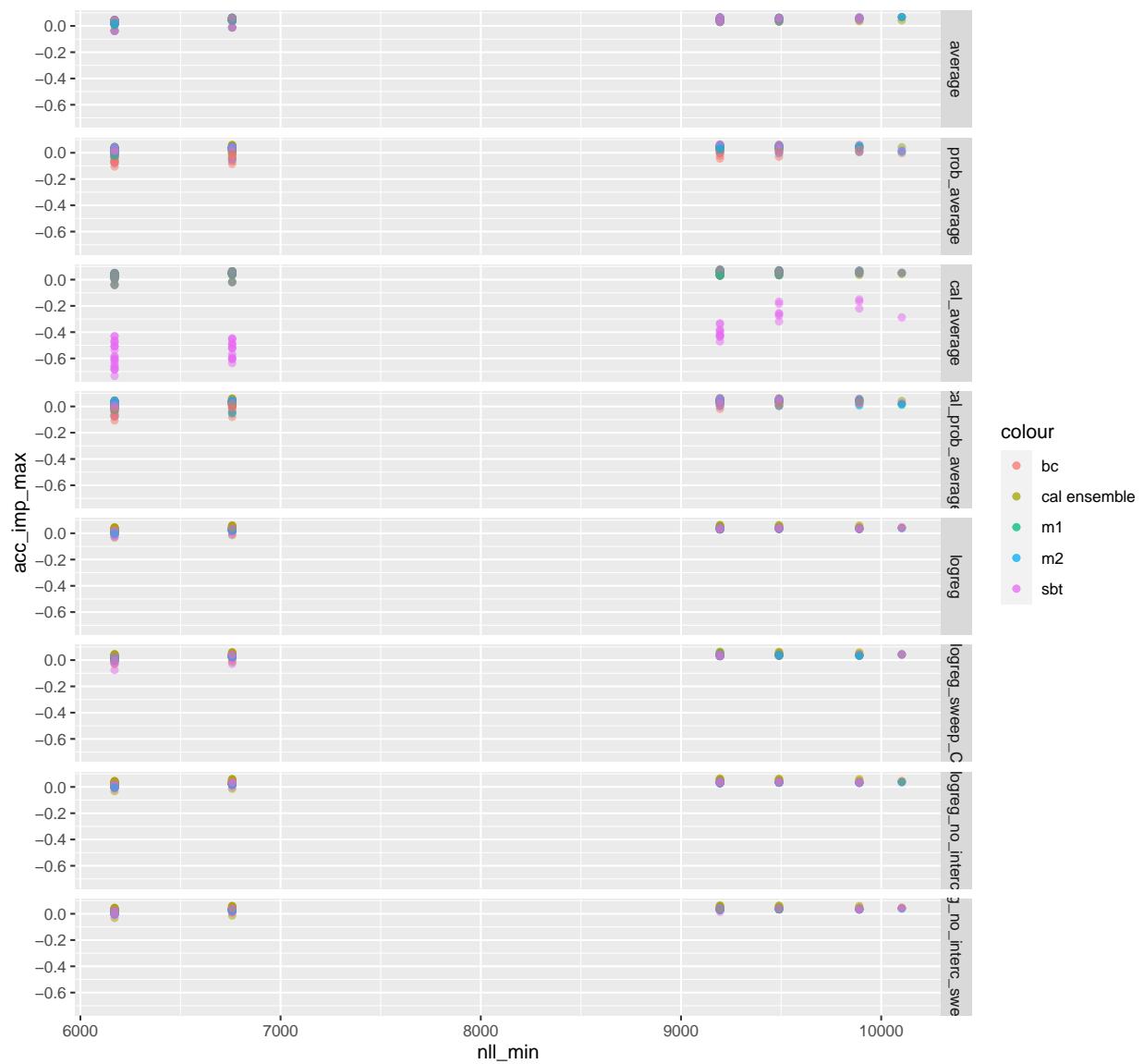
Accuracy improvement of ensemble over  
the best of networks vs nll\_var.  
Ensemble size 3



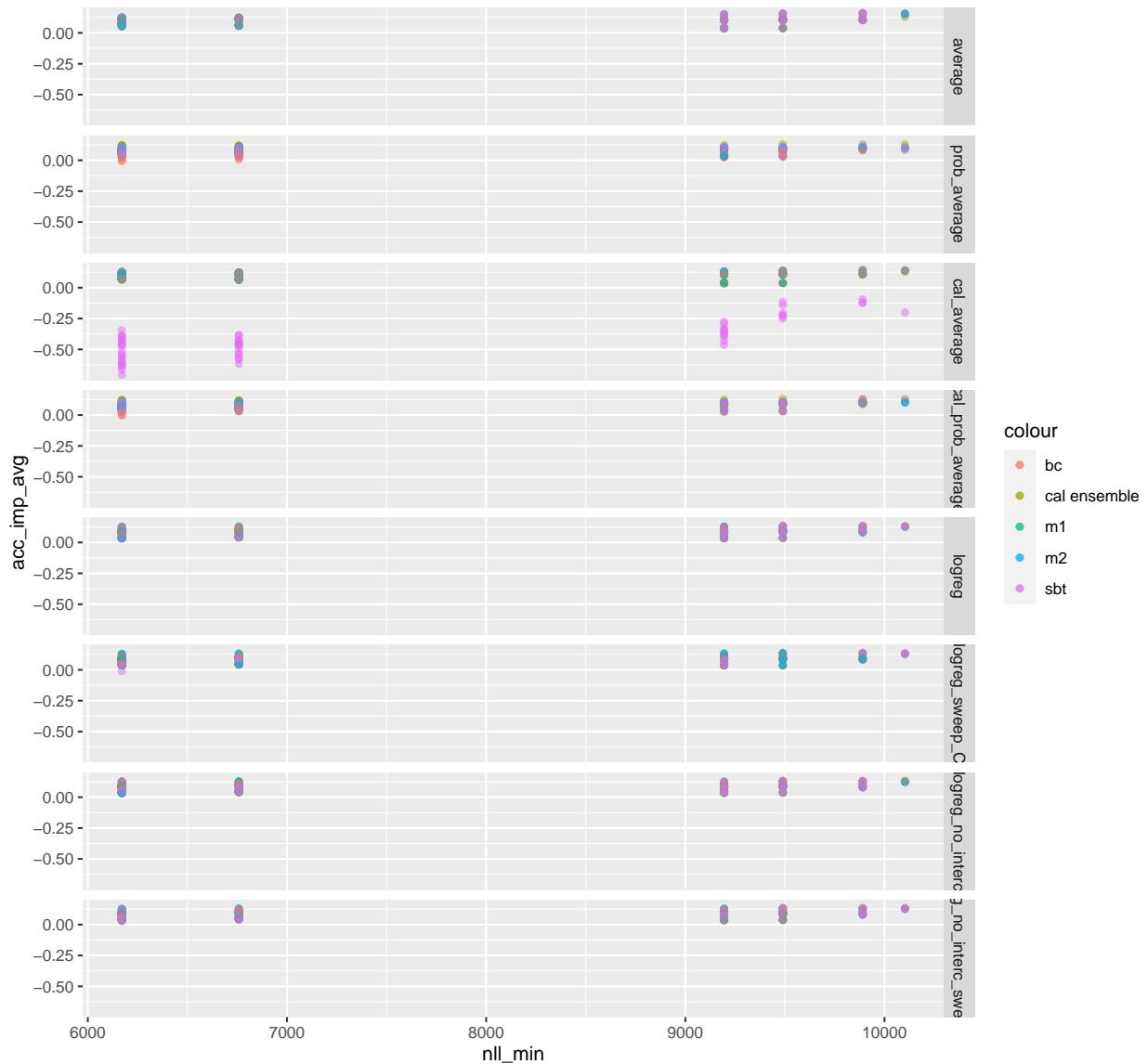
Accuracy improvement of ensemble over  
the average of networks vs nll\_var.  
Ensemble size 3



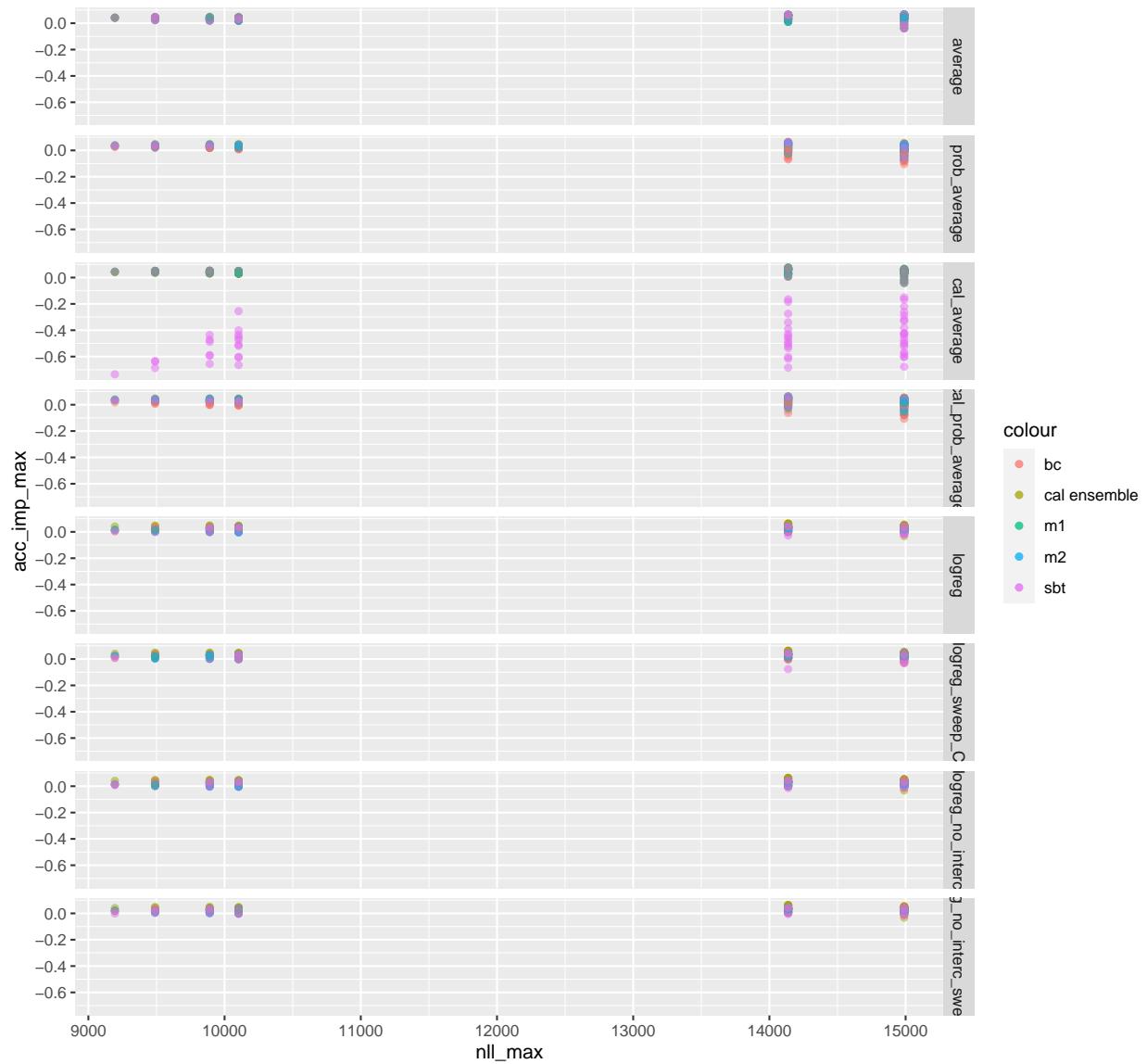
Accuracy improvement of ensemble over  
the best of networks vs nll\_min.  
Ensemble size 3



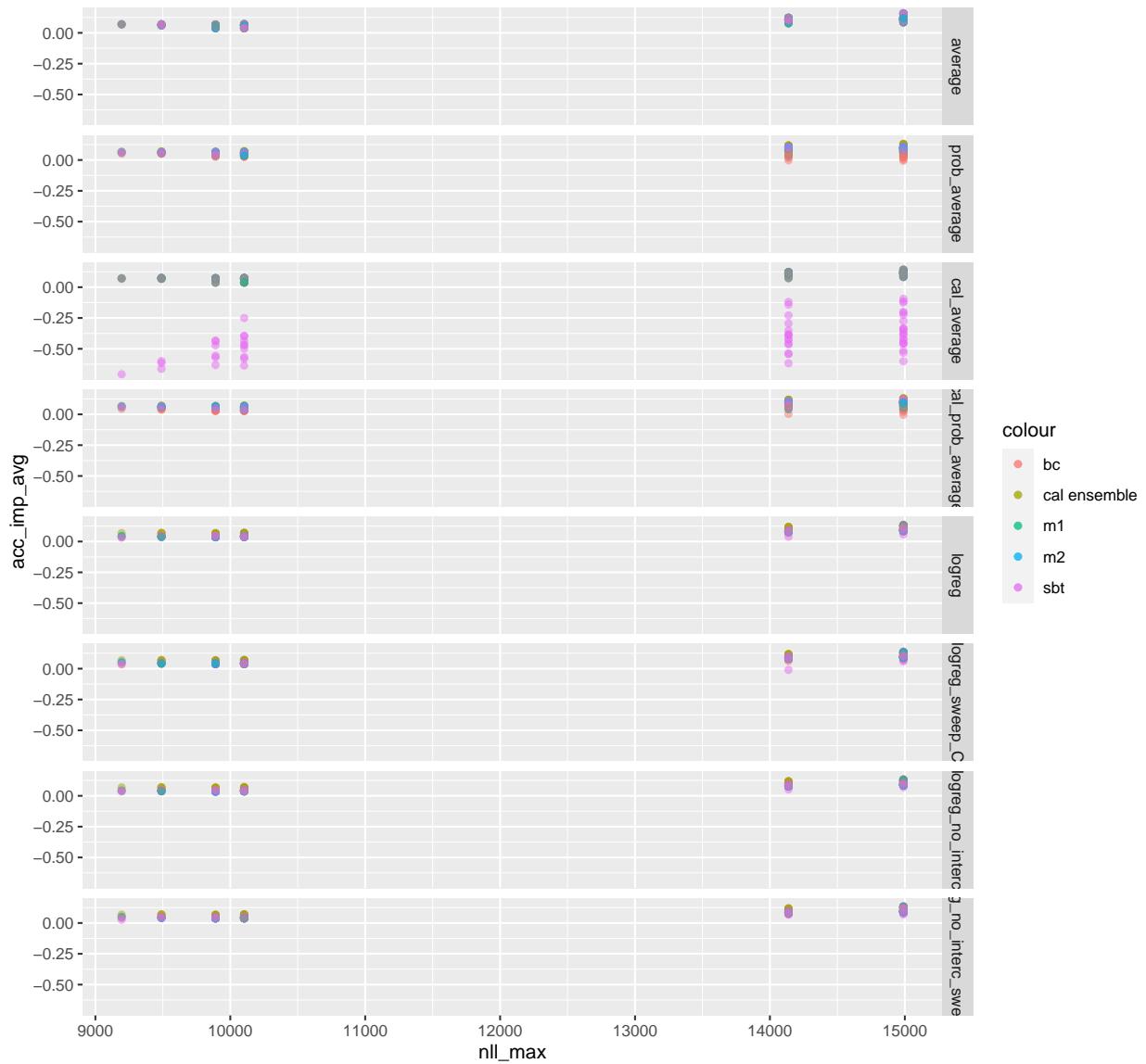
Accuracy improvement of ensemble over  
the average of networks vs nll\_min.  
Ensemble size 3



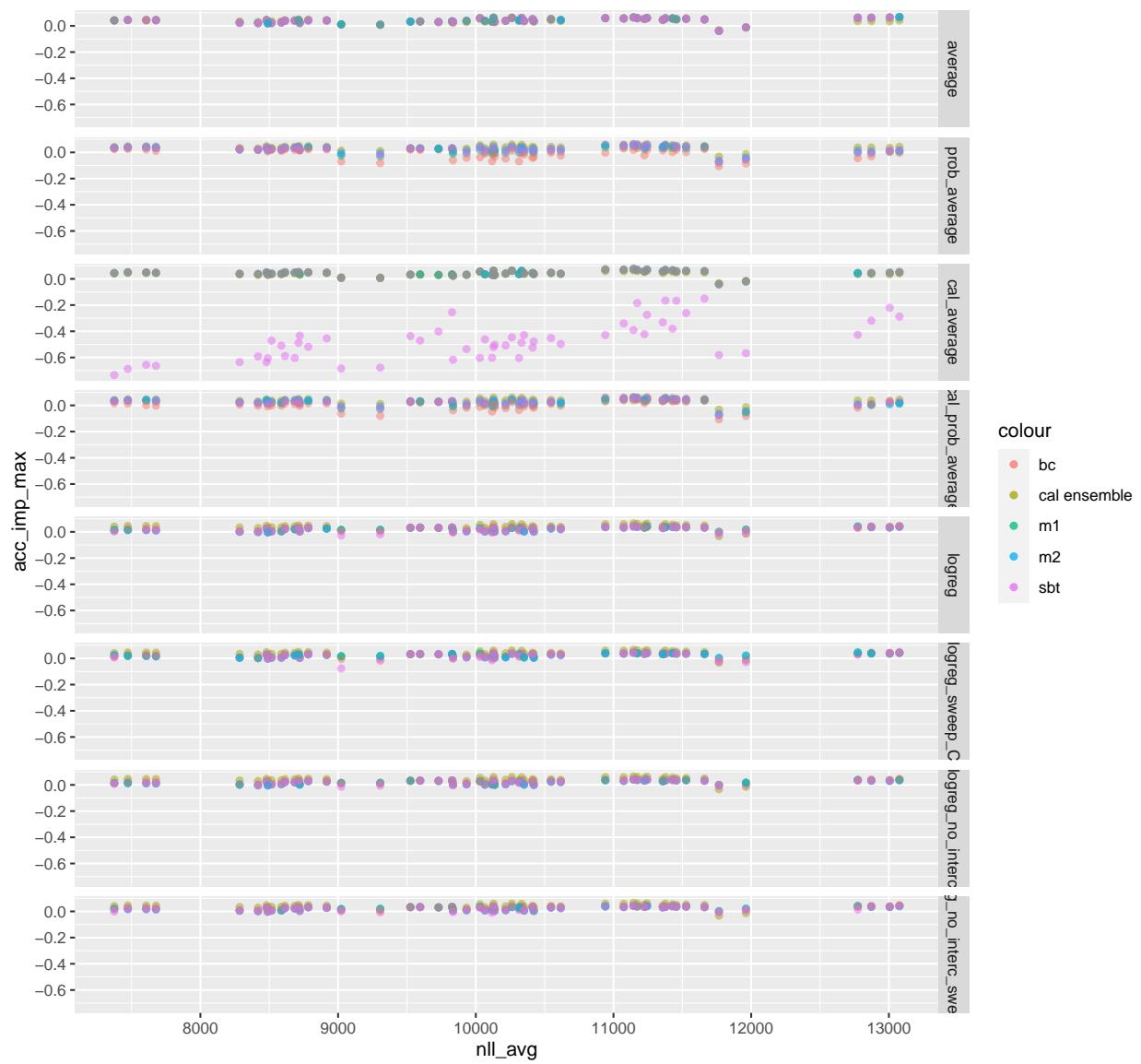
Accuracy improvement of ensemble over  
the best of networks vs nll\_max.  
Ensemble size 3



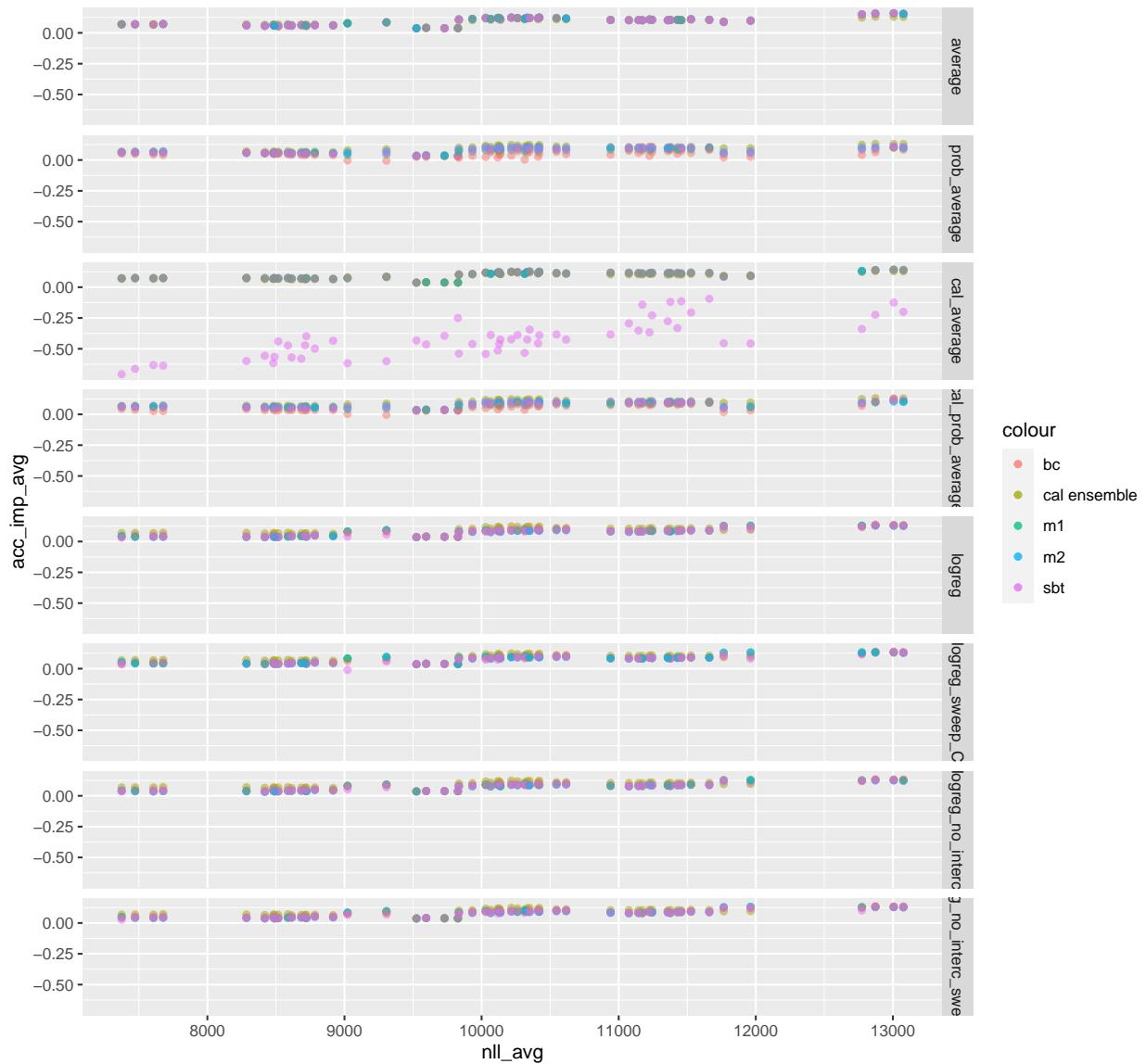
Accuracy improvement of ensemble over  
the average of networks vs nll\_max.  
Ensemble size 3



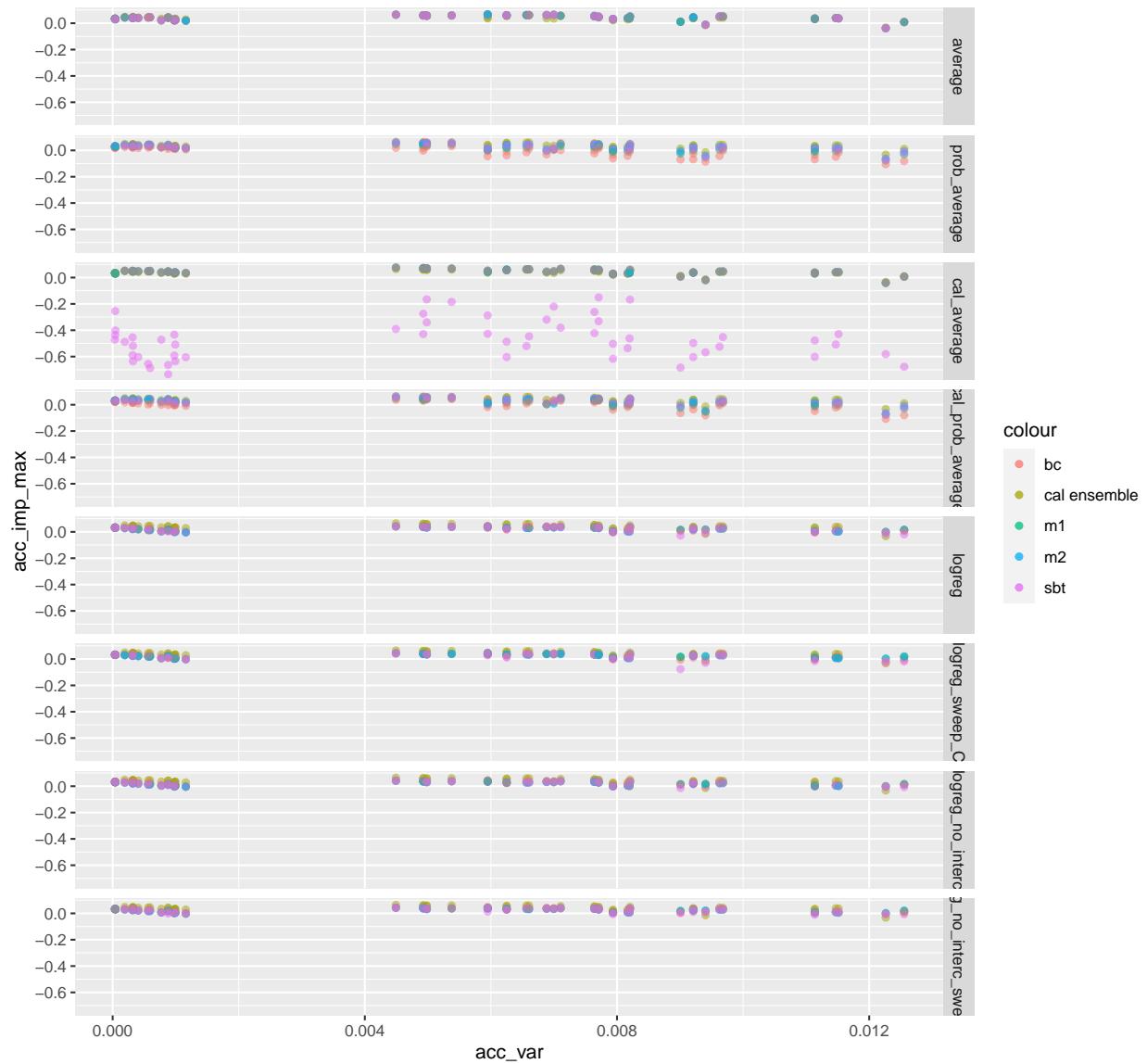
Accuracy improvement of ensemble over  
the best of networks vs nll\_avg.  
Ensemble size 3



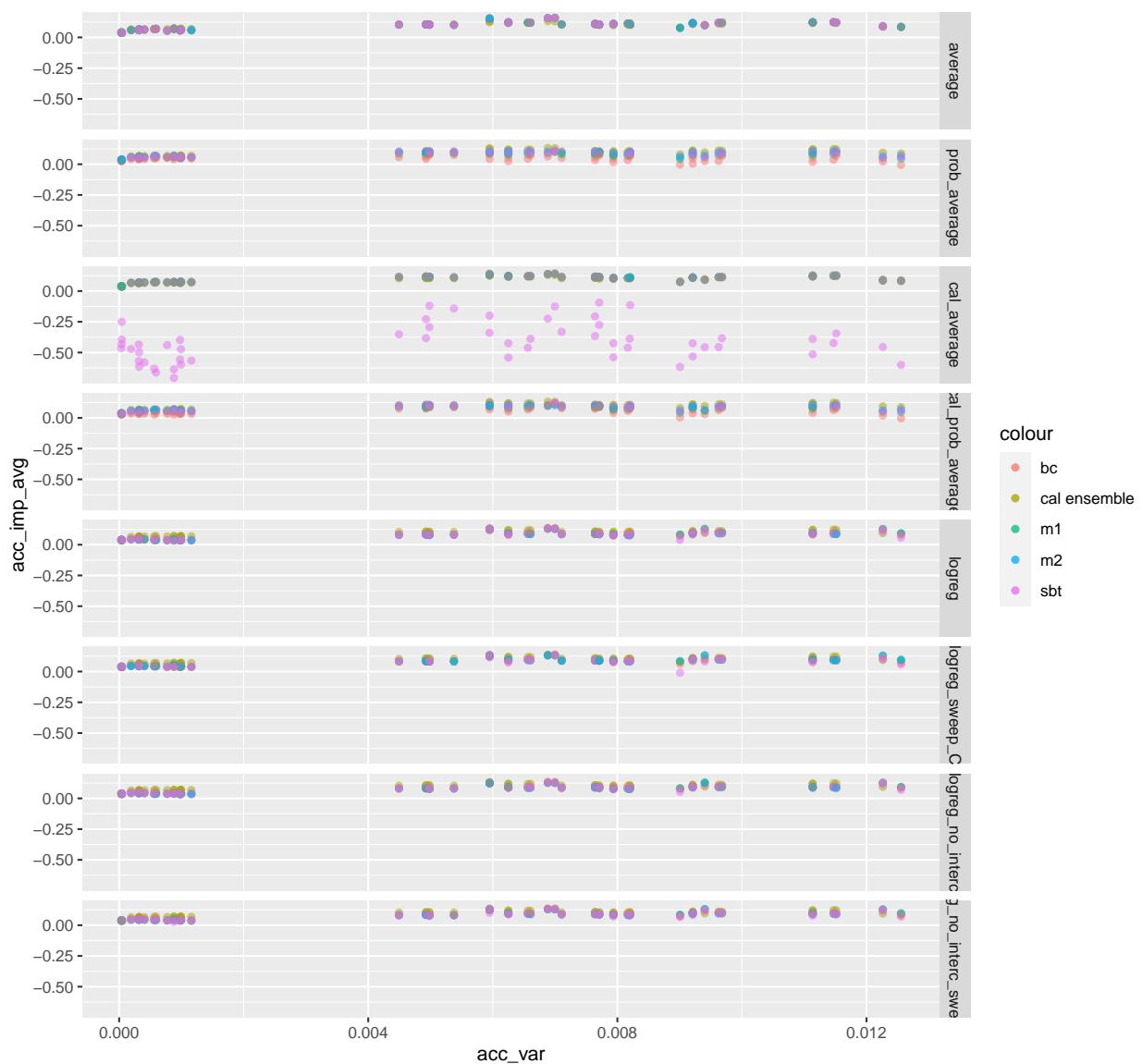
Accuracy improvement of ensemble over  
the average of networks vs nll\_avg.  
Ensemble size 3



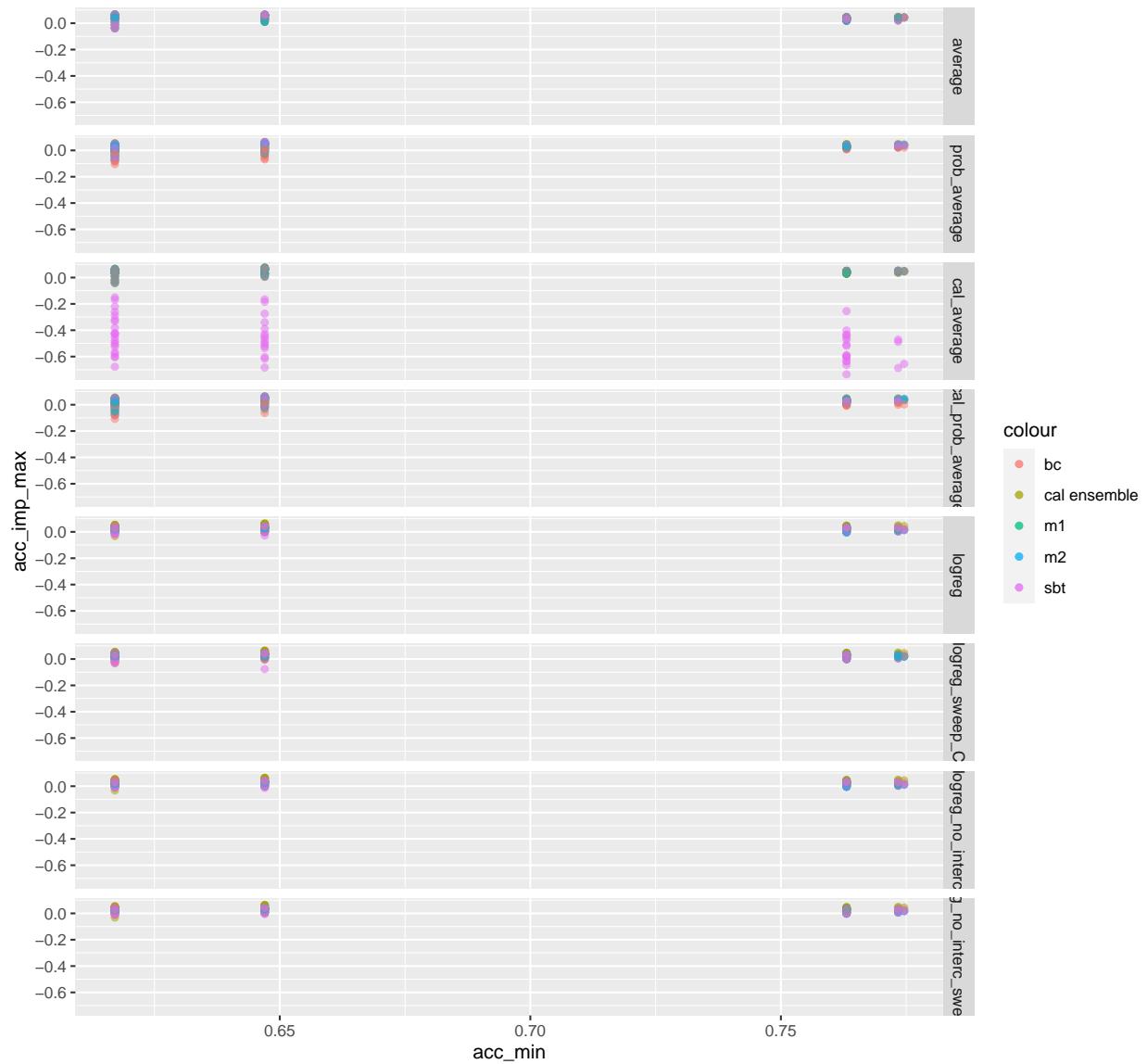
Accuracy improvement of ensemble over  
the best of networks vs acc\_var.  
Ensemble size 3



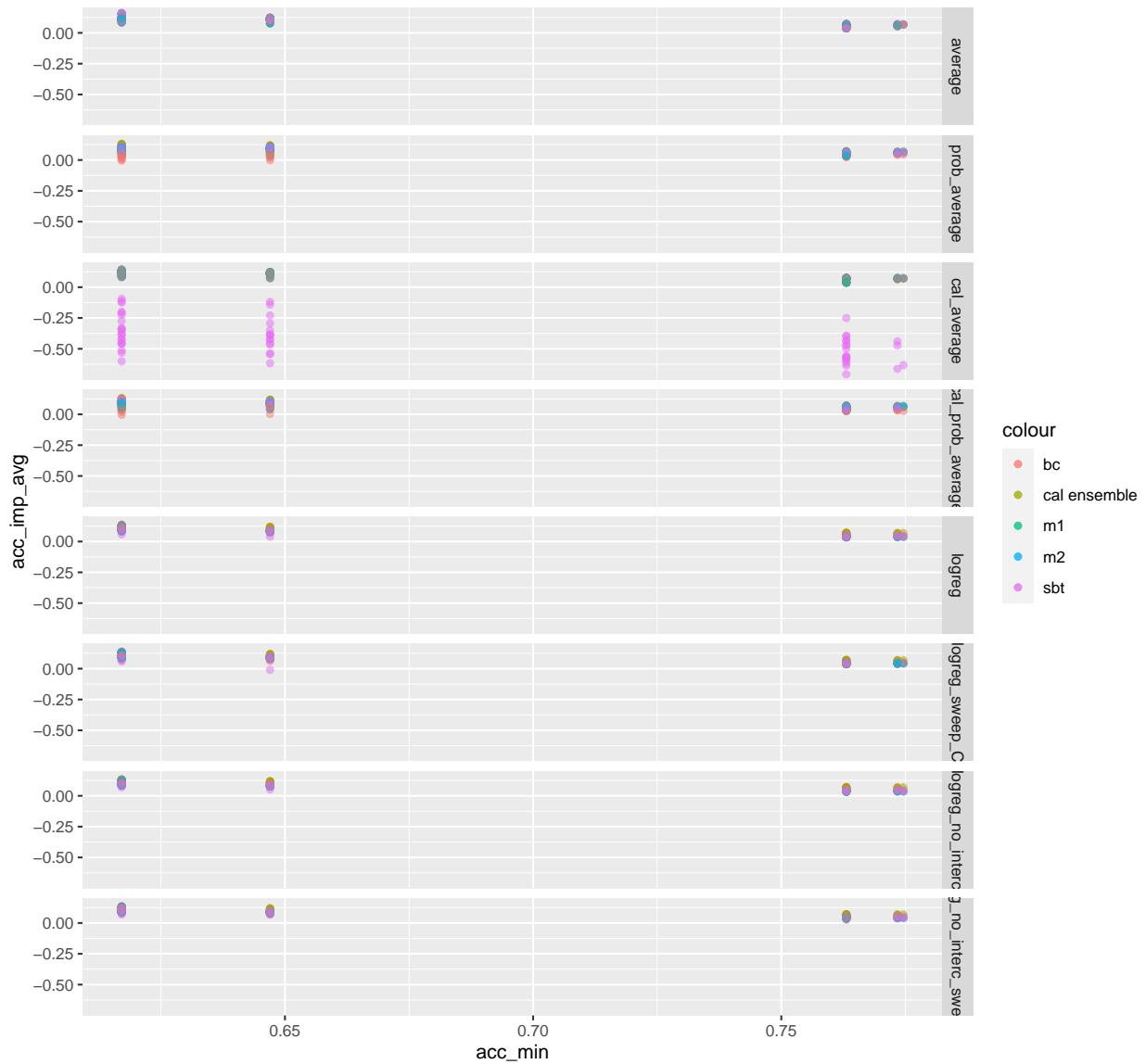
Accuracy improvement of ensemble over  
the average of networks vs acc\_var.  
Ensemble size 3



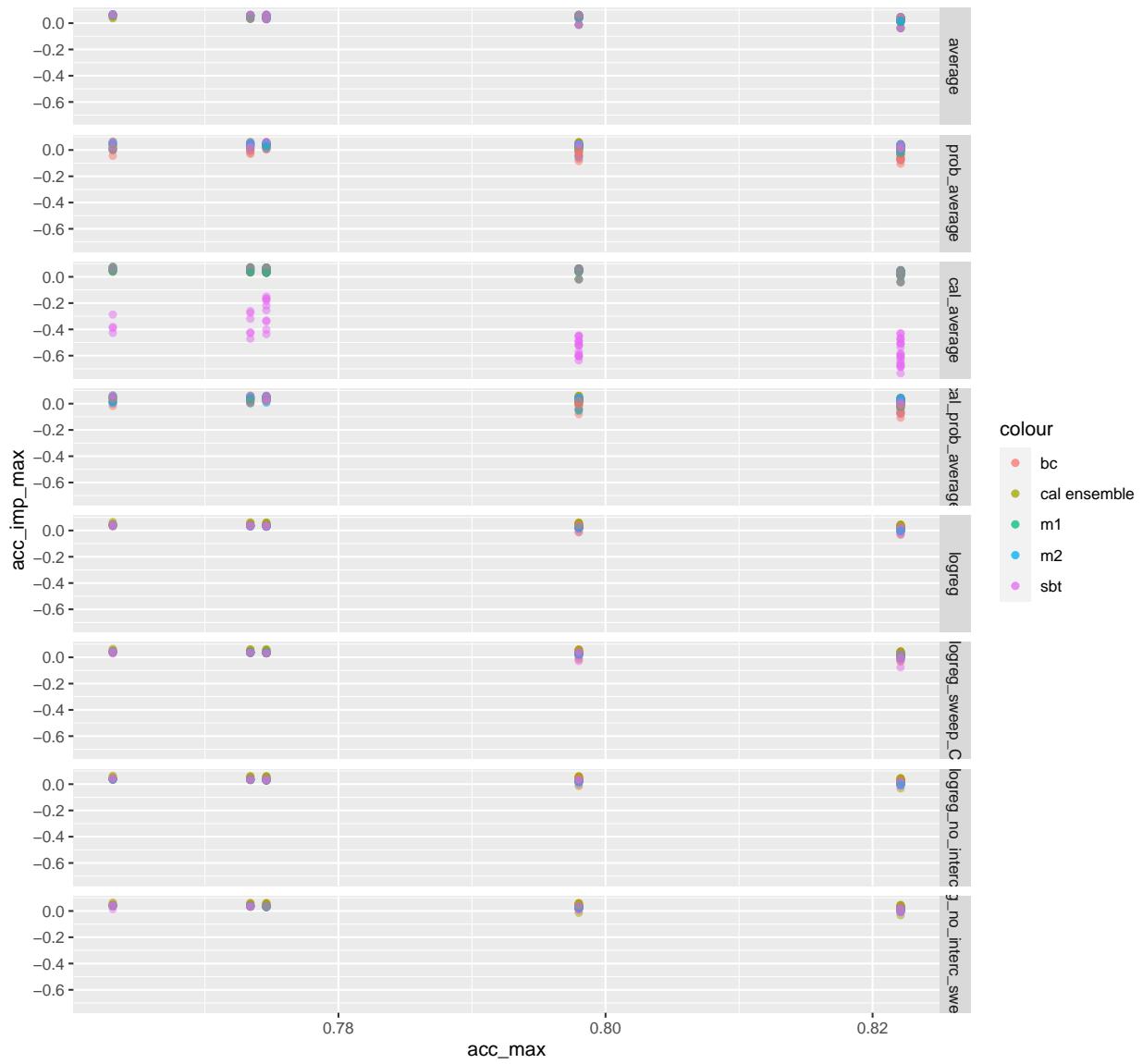
Accuracy improvement of ensemble over  
the best of networks vs acc\_min.  
Ensemble size 3



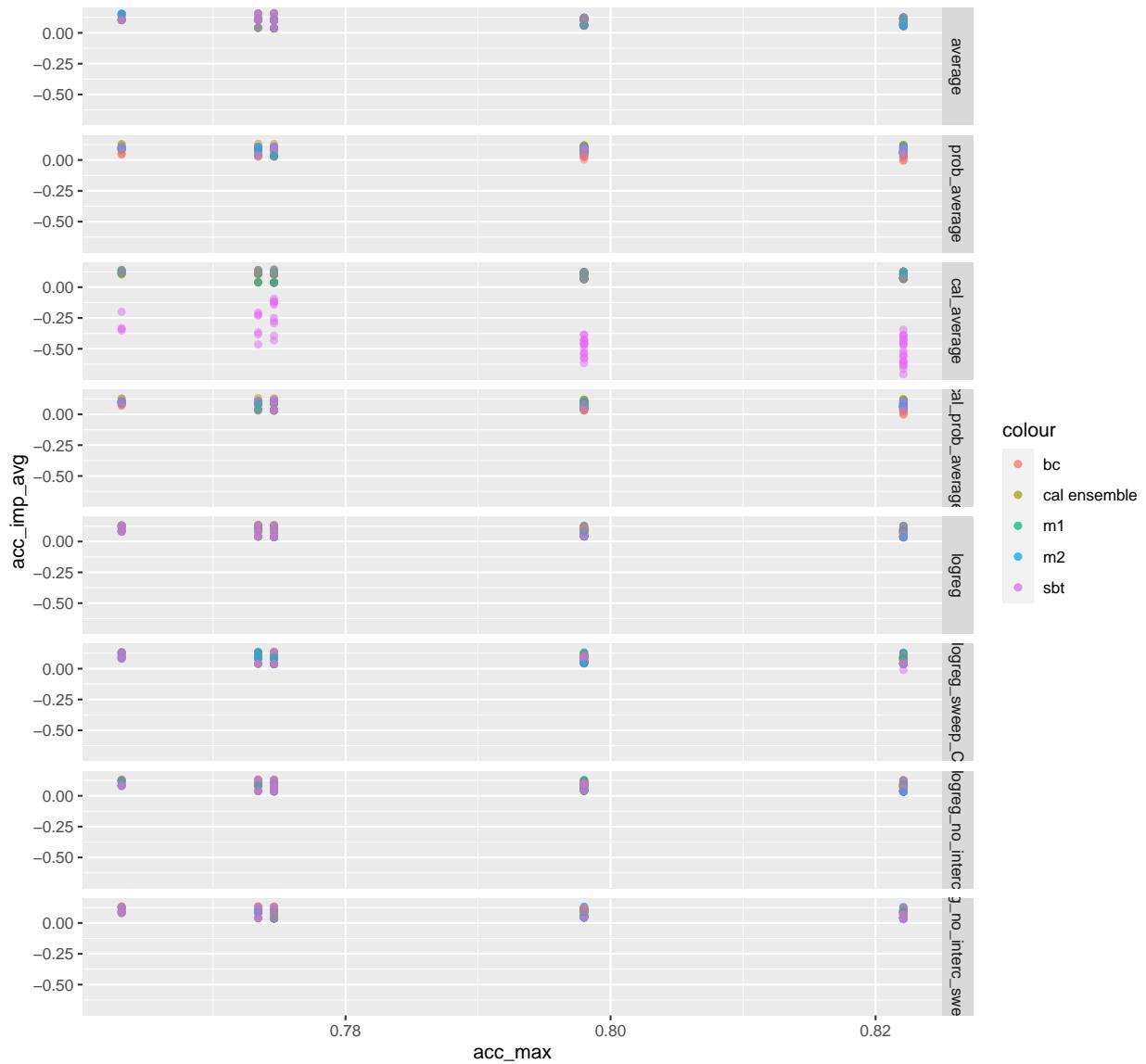
Accuracy improvement of ensemble over  
the average of networks vs acc\_min.  
Ensemble size 3



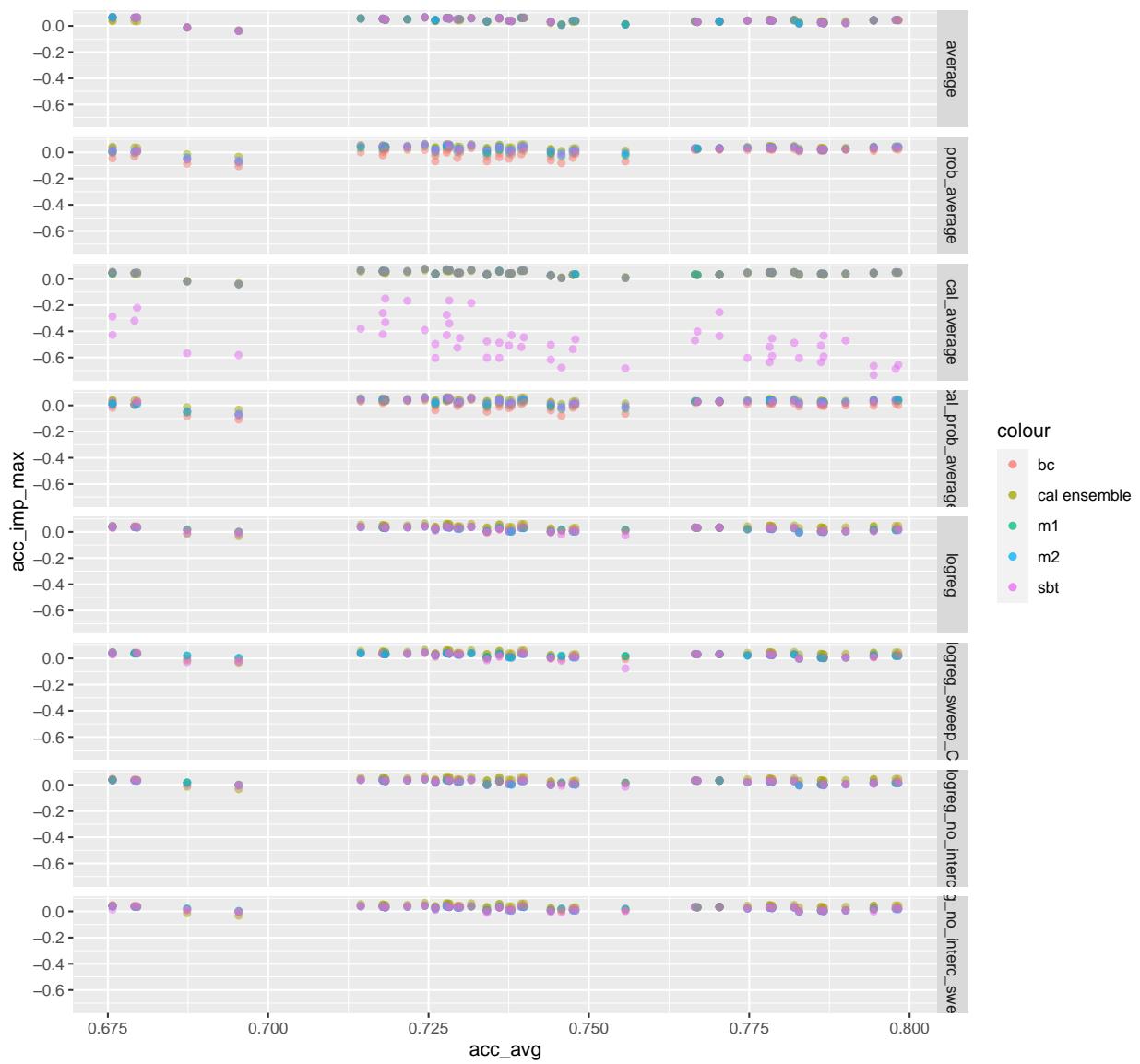
Accuracy improvement of ensemble over  
the best of networks vs acc\_max.  
Ensemble size 3



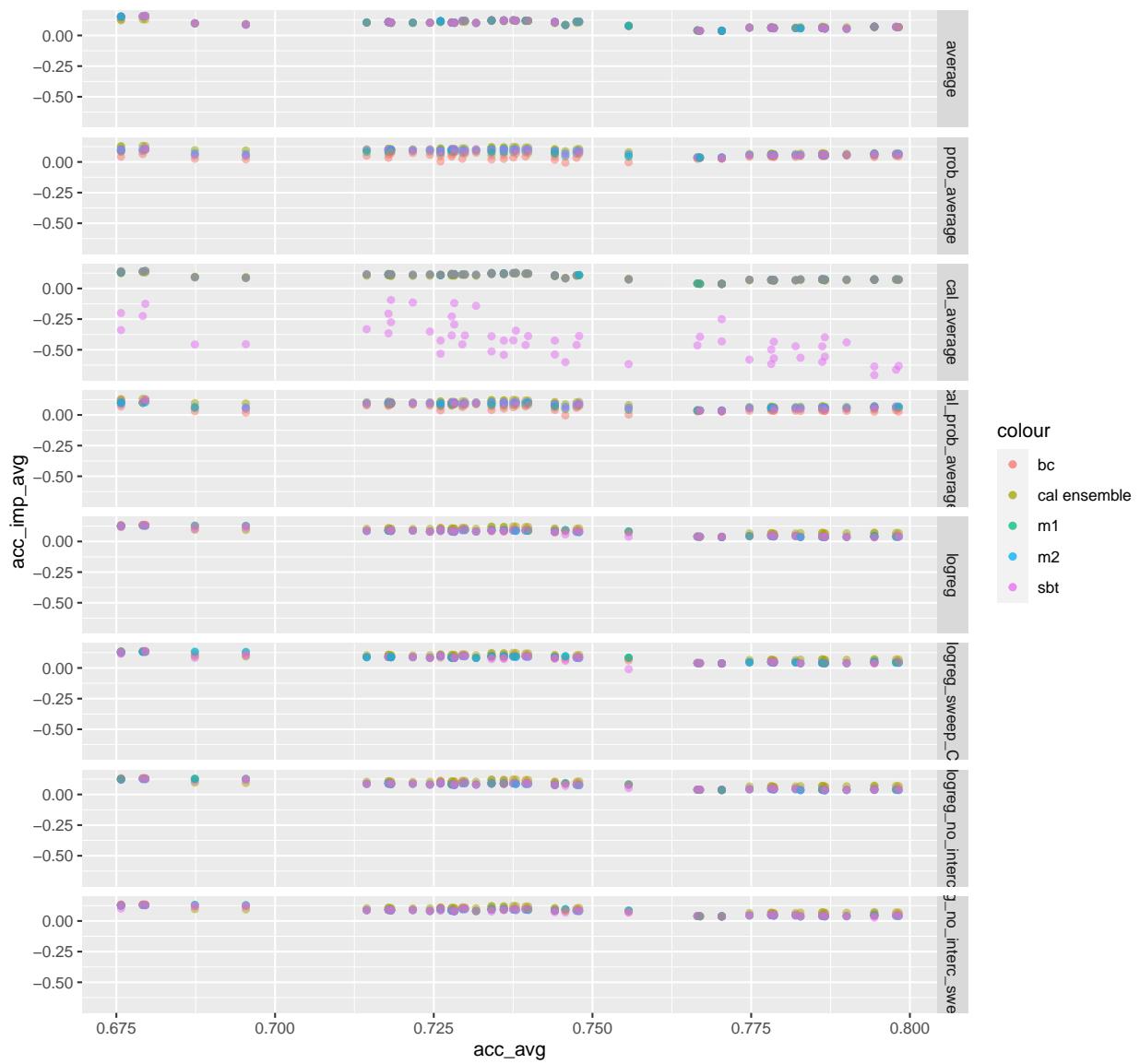
Accuracy improvement of ensemble over  
the average of networks vs acc\_max.  
Ensemble size 3



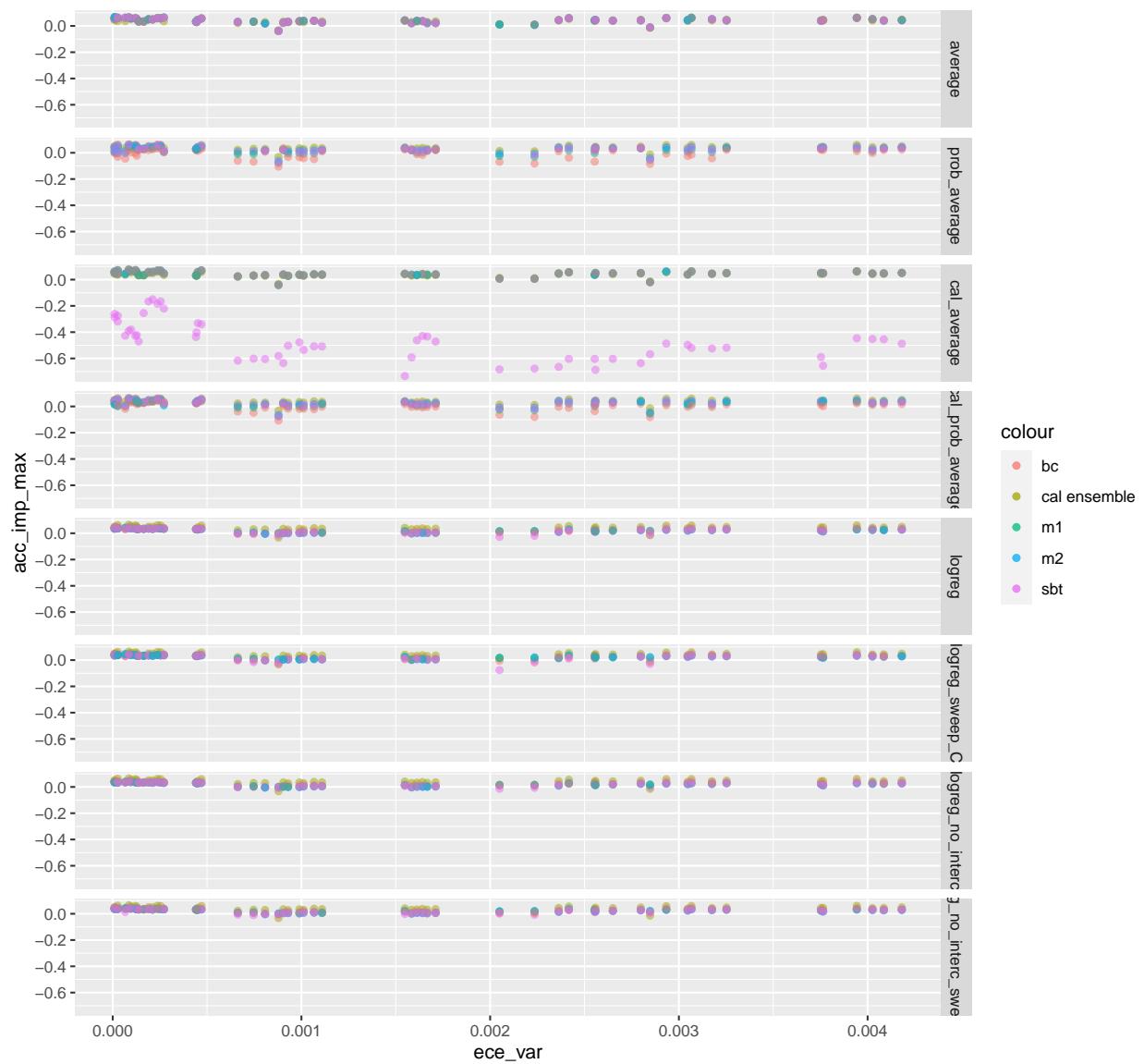
Accuracy improvement of ensemble over  
the best of networks vs acc\_avg.  
Ensemble size 3



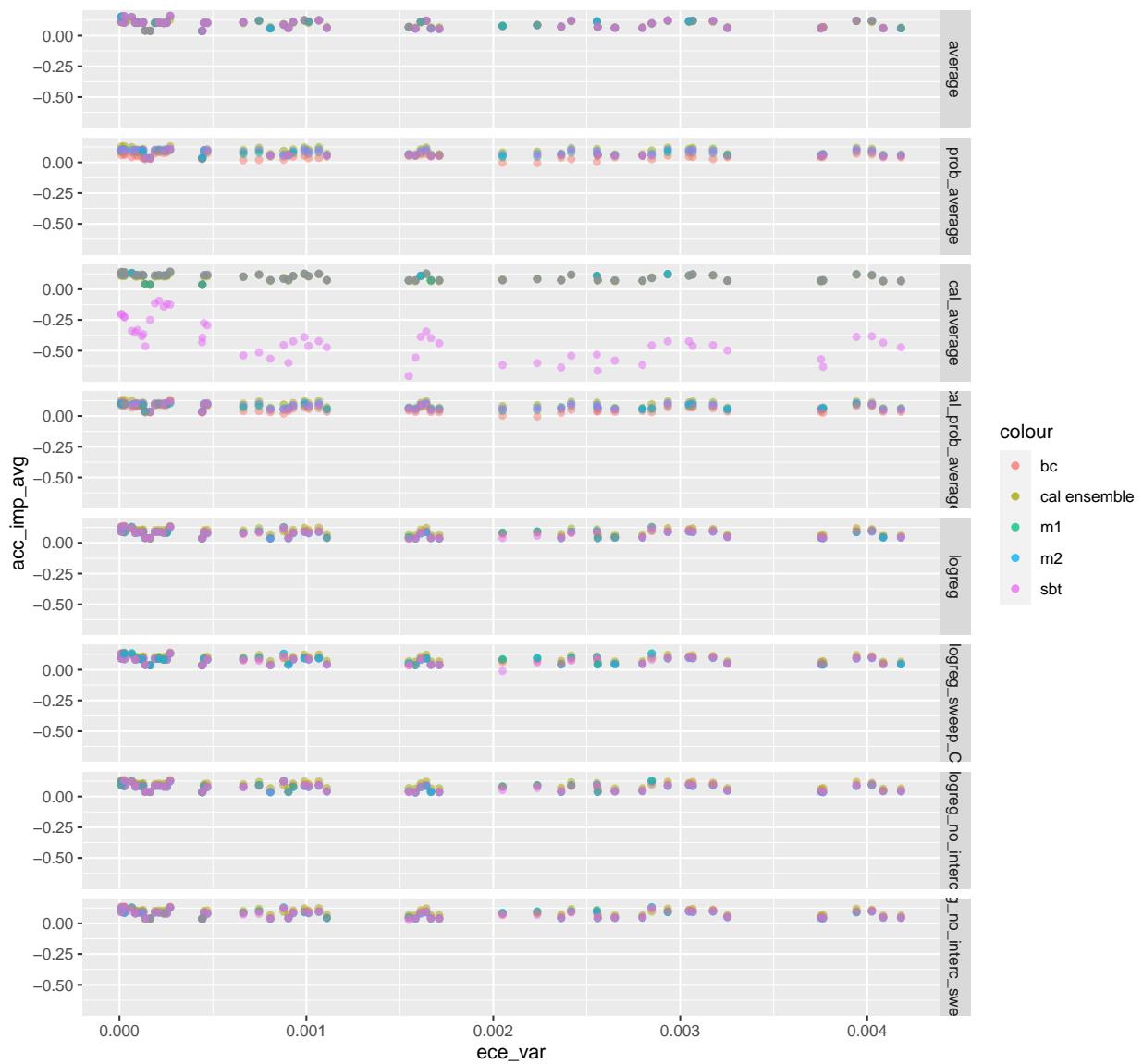
Accuracy improvement of ensemble over  
the average of networks vs acc\_avg.  
Ensemble size 3



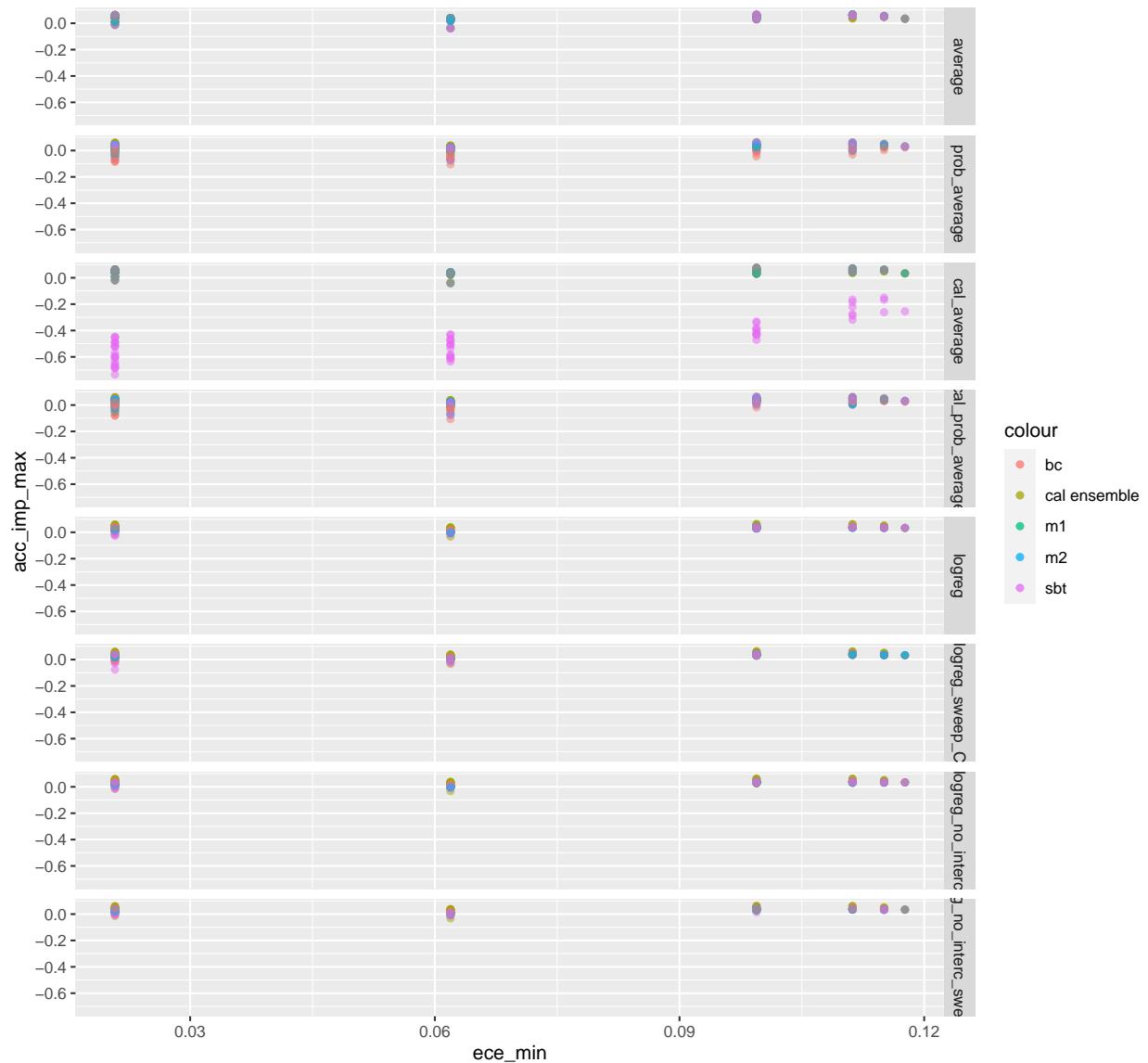
Accuracy improvement of ensemble over  
the best of networks vs ece\_var.  
Ensemble size 3



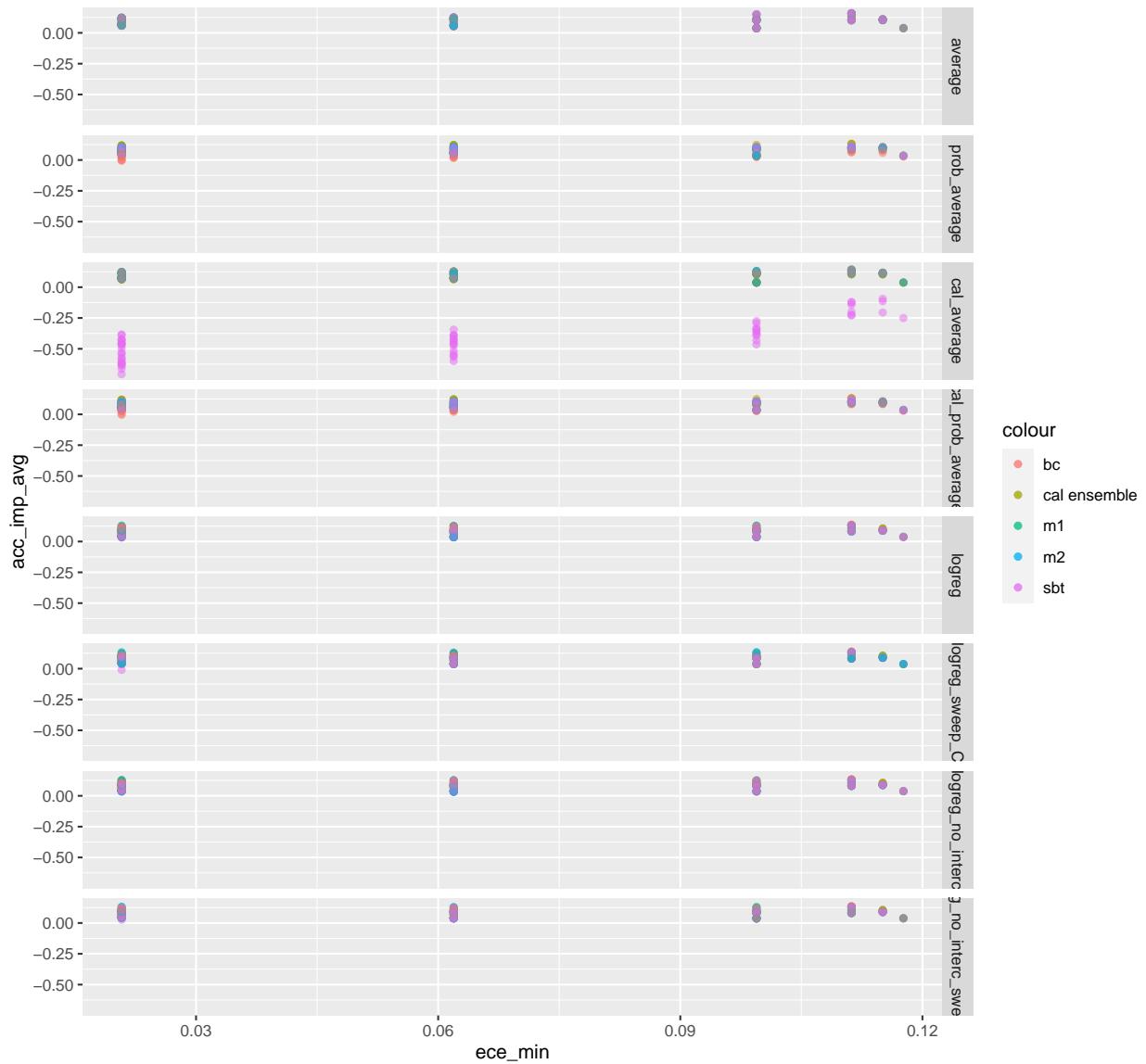
Accuracy improvement of ensemble over  
the average of networks vs ece\_var.  
Ensemble size 3



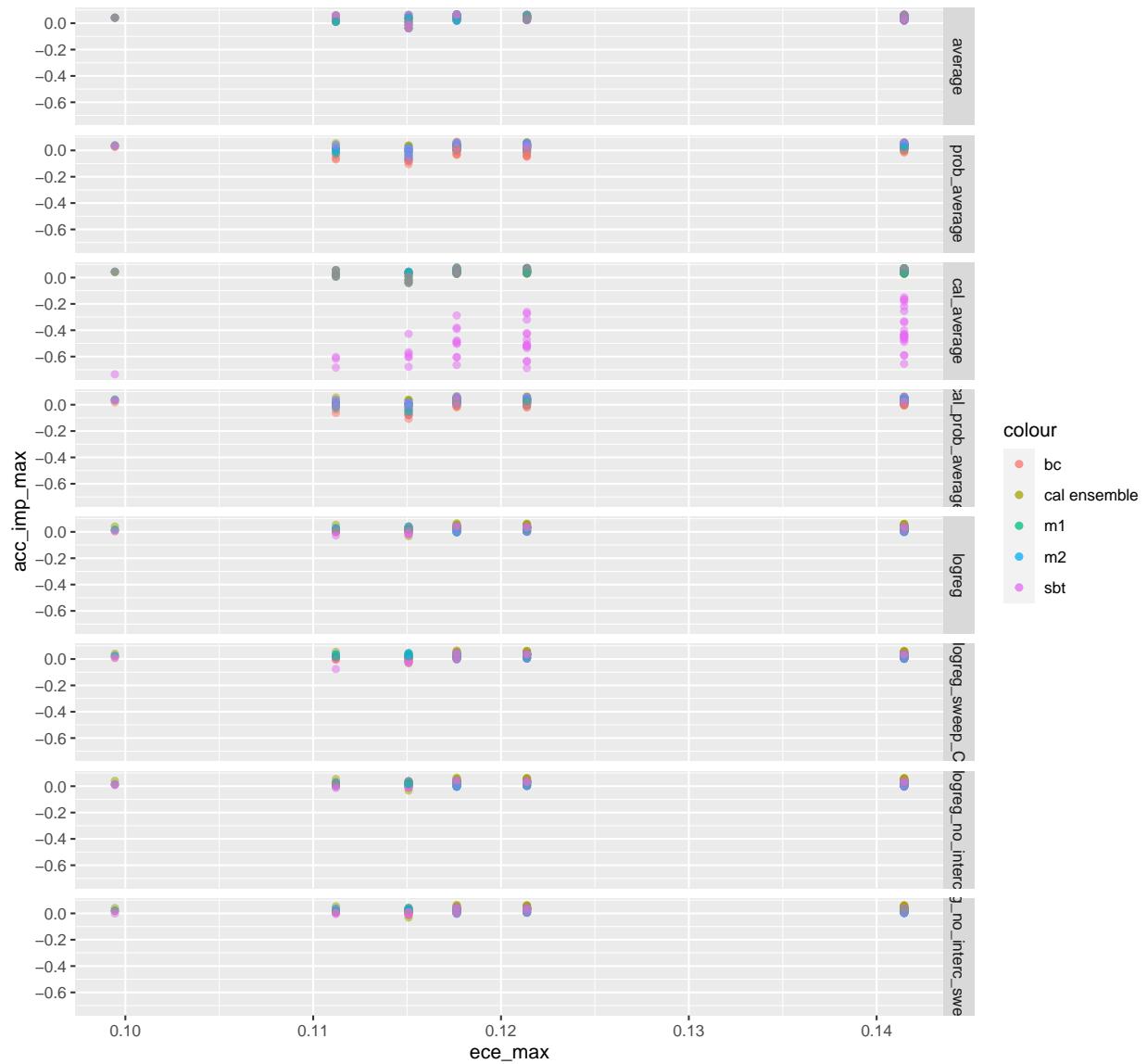
Accuracy improvement of ensemble over  
the best of networks vs ece\_min.  
Ensemble size 3



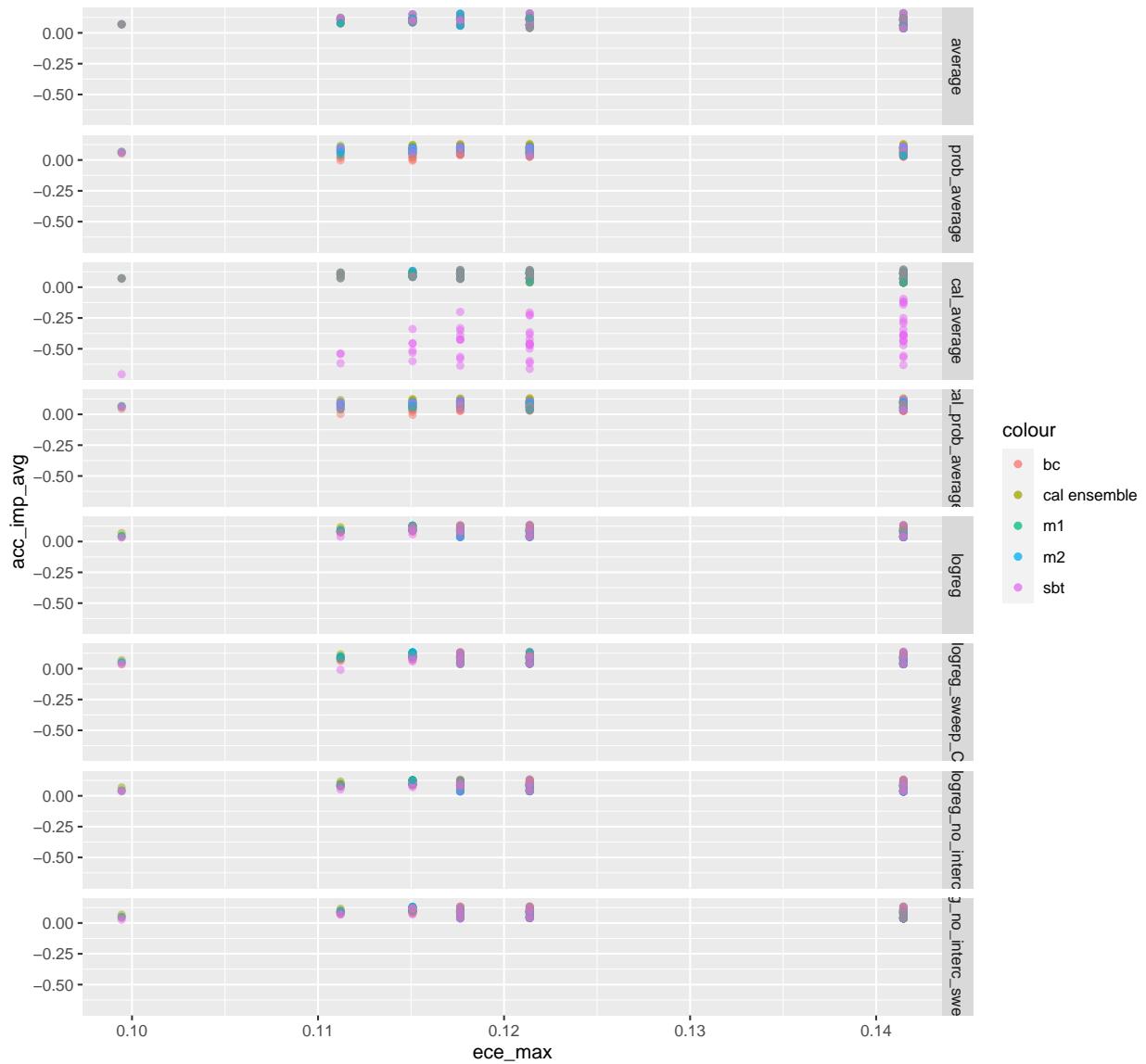
Accuracy improvement of ensemble over  
the average of networks vs ece\_min.  
Ensemble size 3



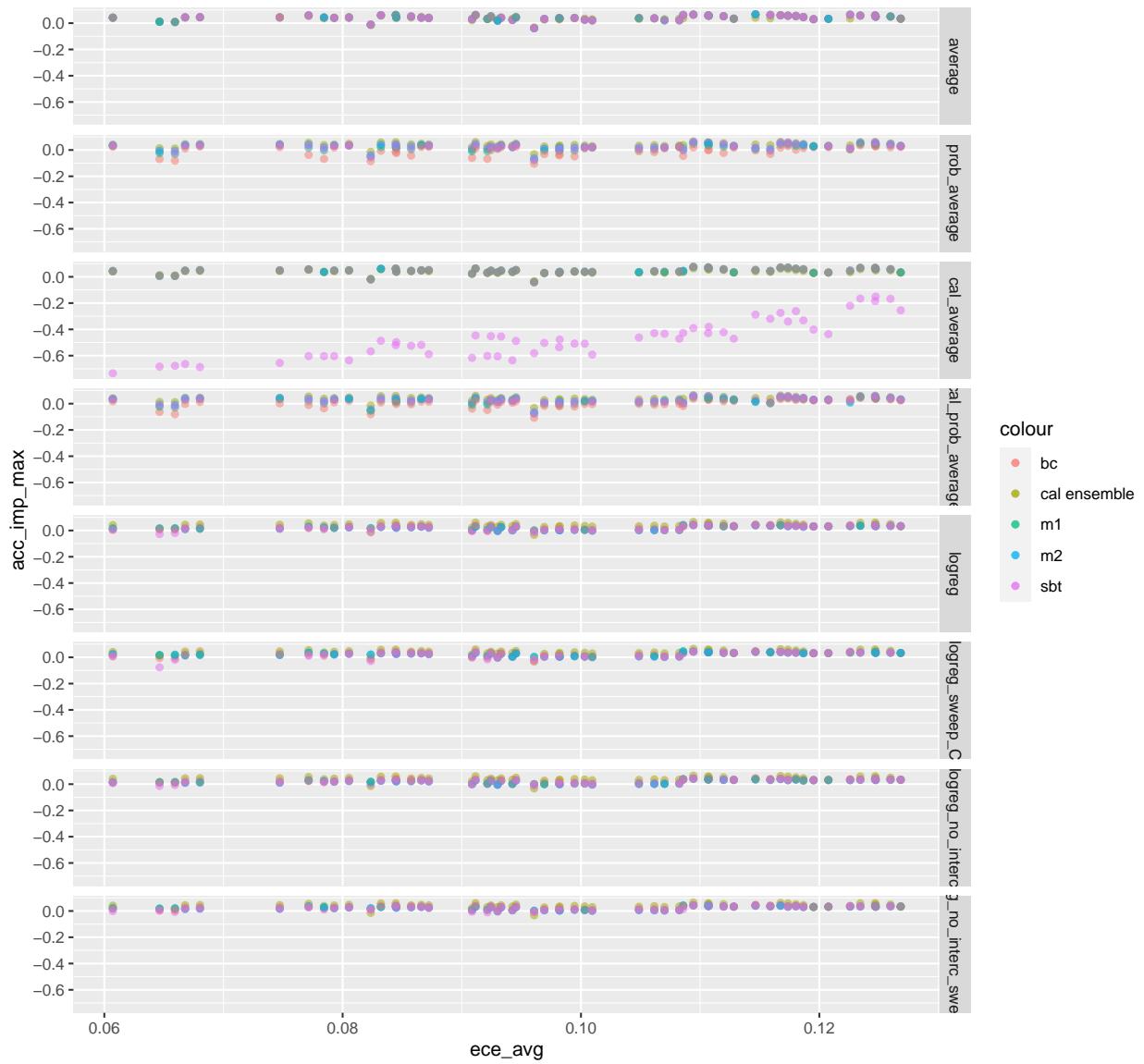
Accuracy improvement of ensemble over  
the best of networks vs ece\_max.  
Ensemble size 3



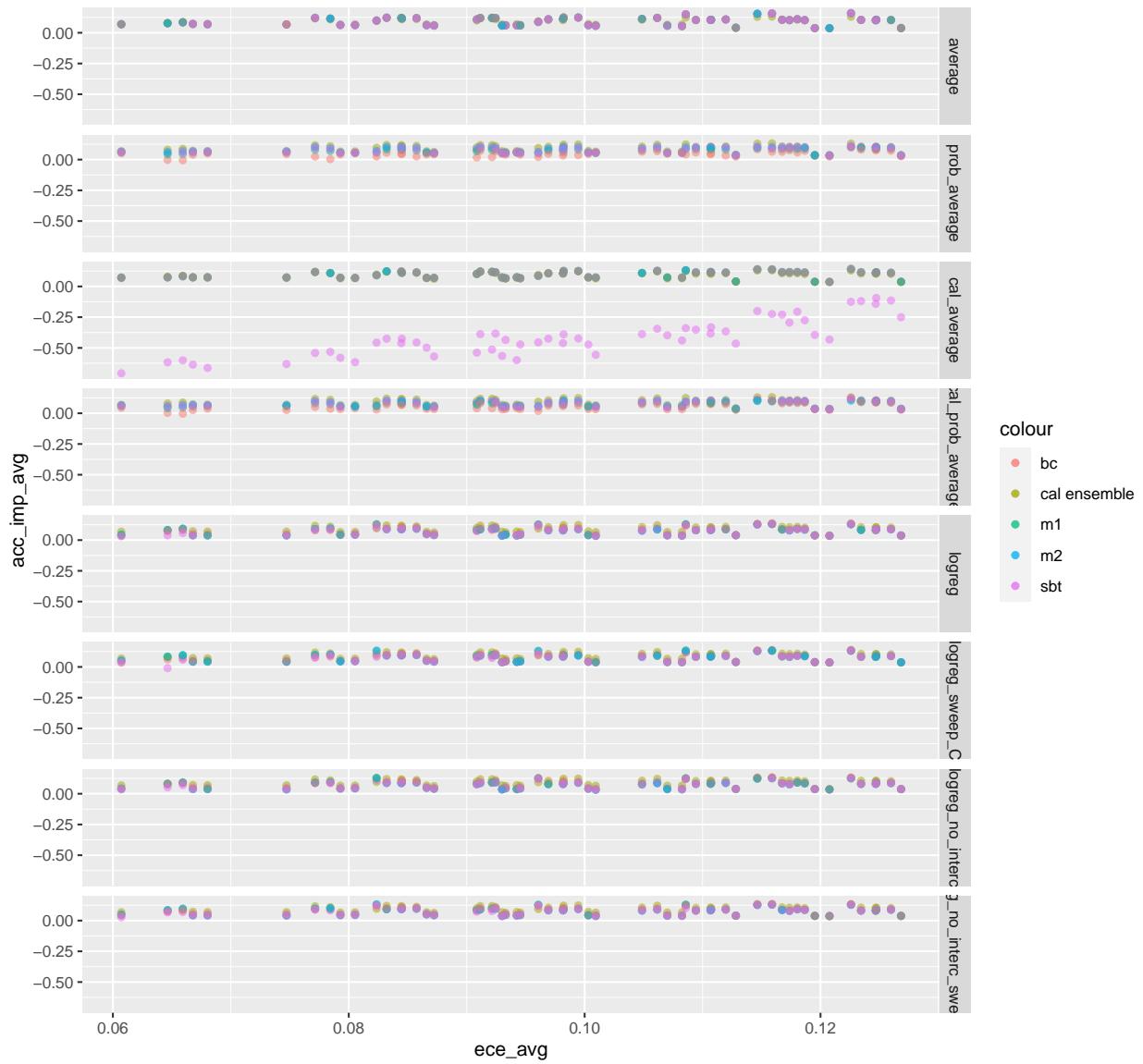
Accuracy improvement of ensemble over  
the average of networks vs ece\_max.  
Ensemble size 3



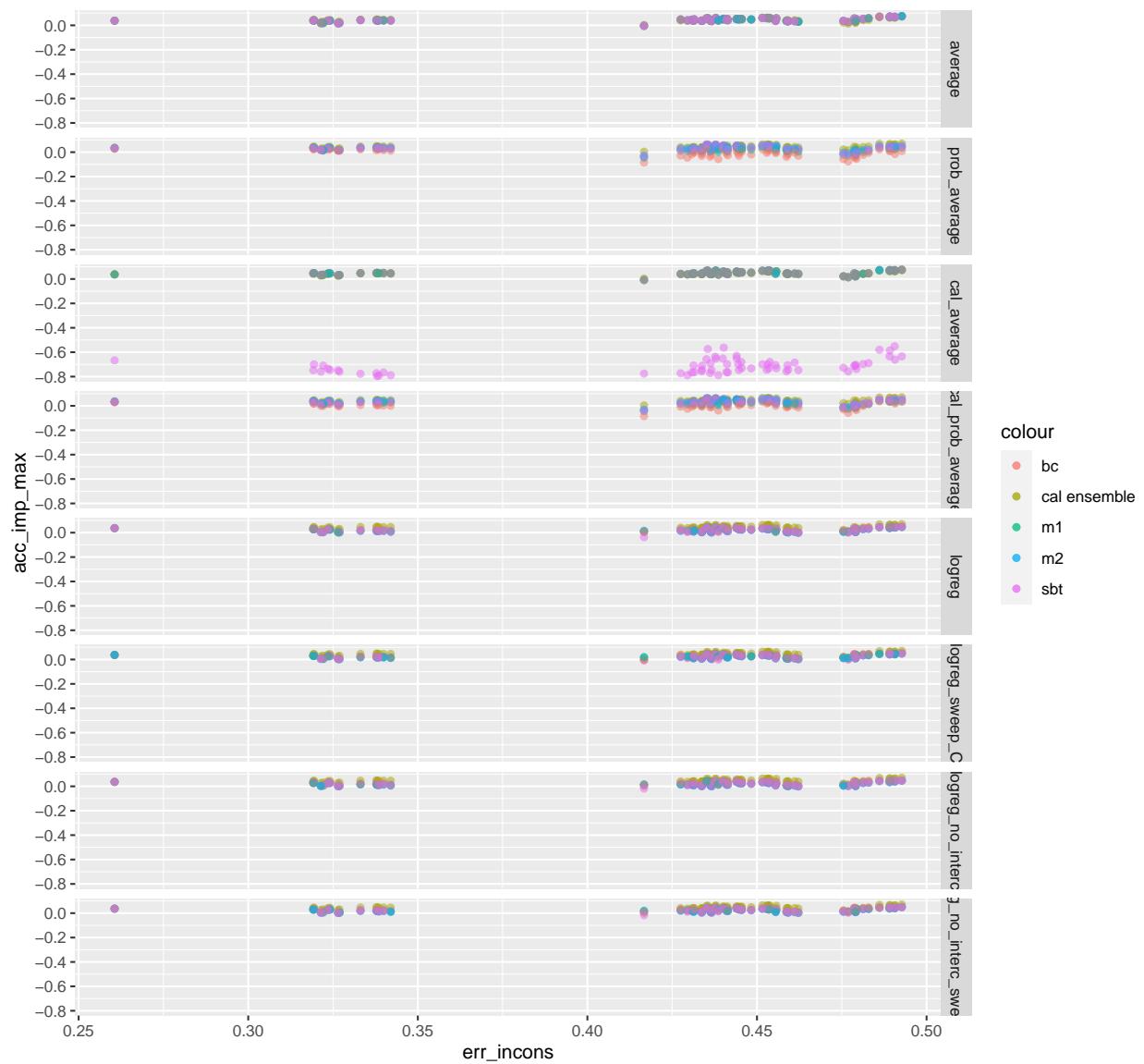
Accuracy improvement of ensemble over  
the best of networks vs ece\_avg.  
Ensemble size 3



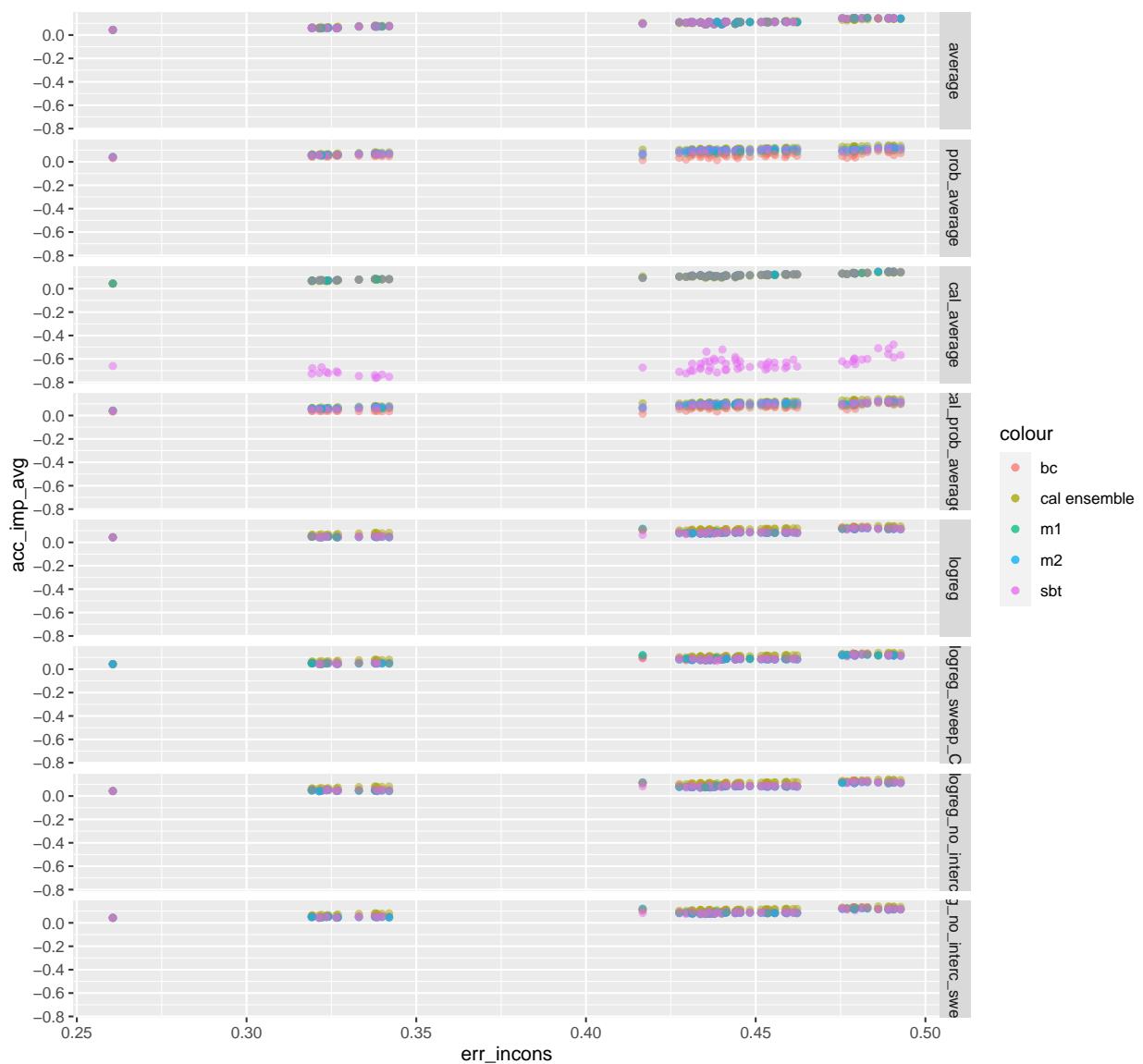
Accuracy improvement of ensemble over  
the average of networks vs ece\_avg.  
Ensemble size 3



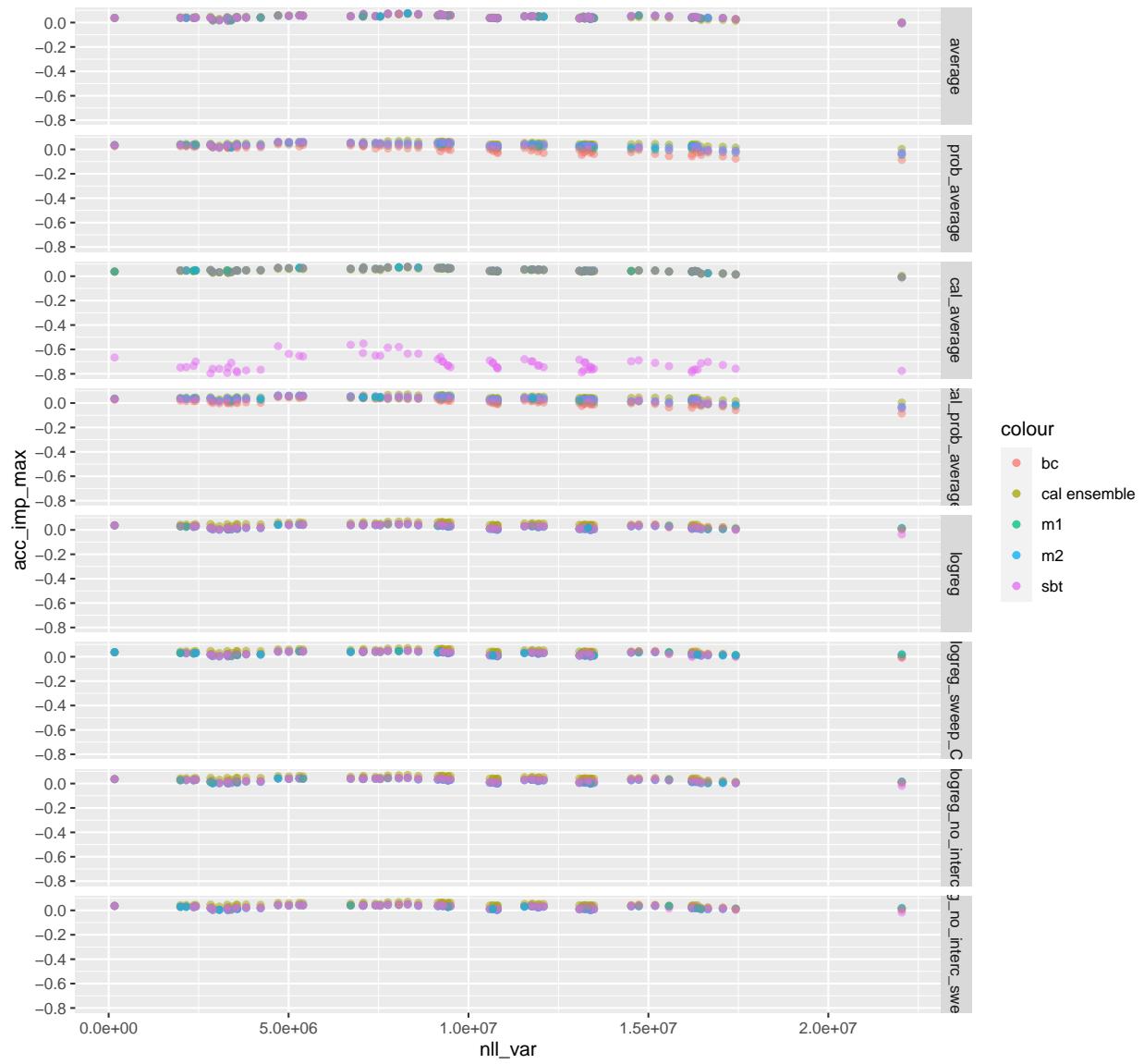
Accuracy improvement of ensemble over  
the best of networks vs err\_incons.  
Ensemble size 4



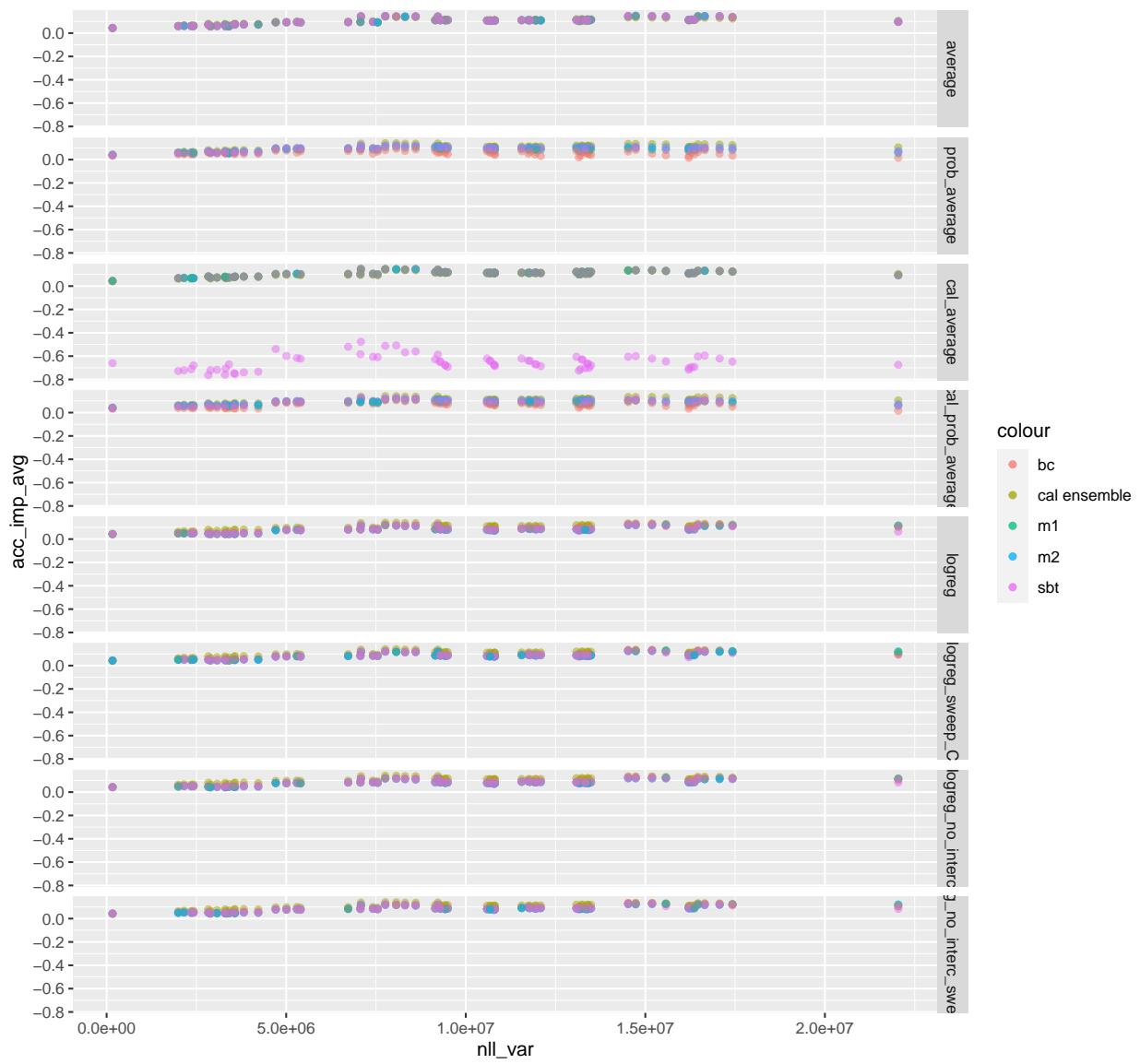
Accuracy improvement of ensemble over  
the average of networks vs err\_incons.  
Ensemble size 4



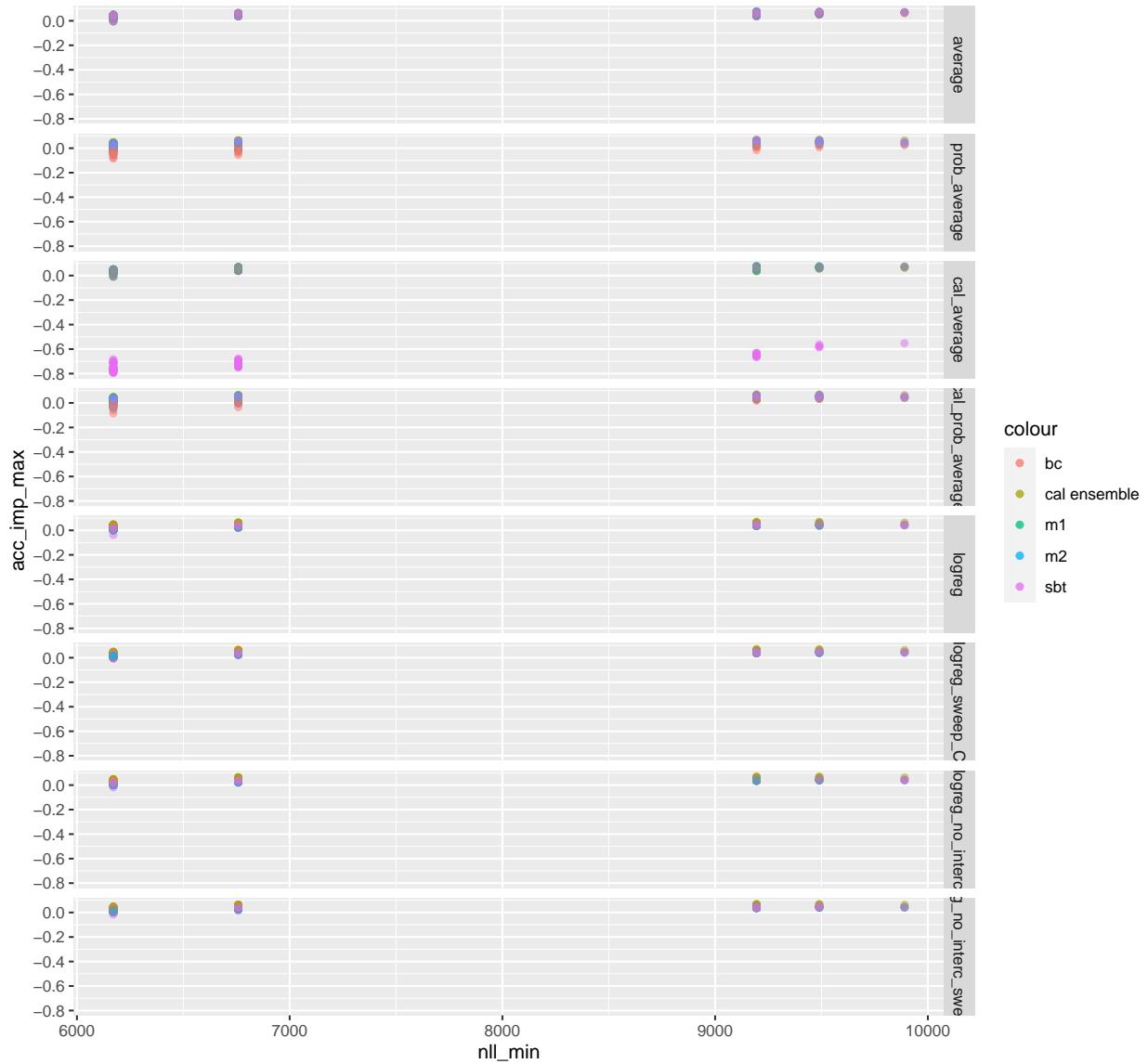
Accuracy improvement of ensemble over  
the best of networks vs nll\_var.  
Ensemble size 4



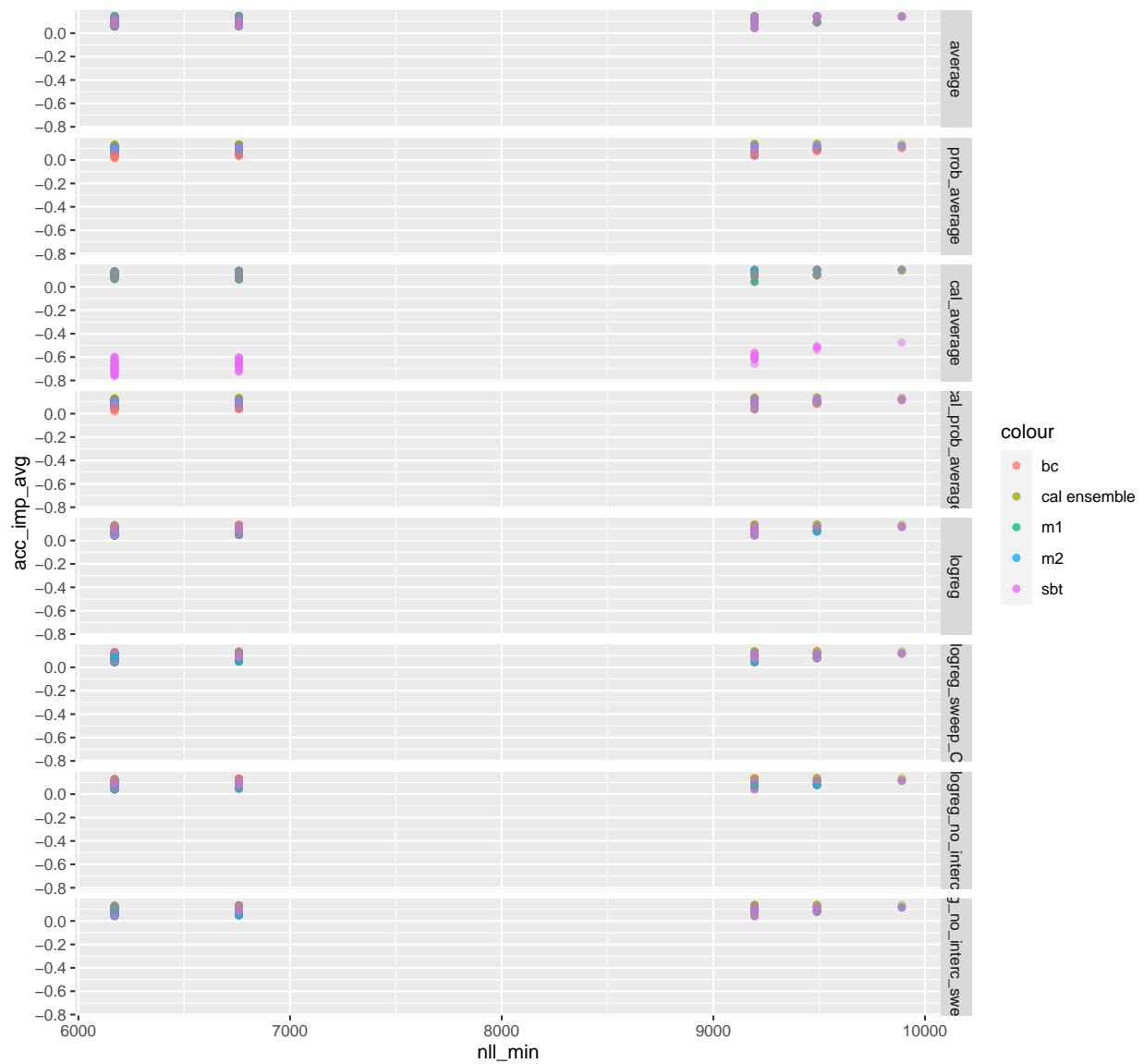
Accuracy improvement of ensemble over  
the average of networks vs nll\_var.  
Ensemble size 4



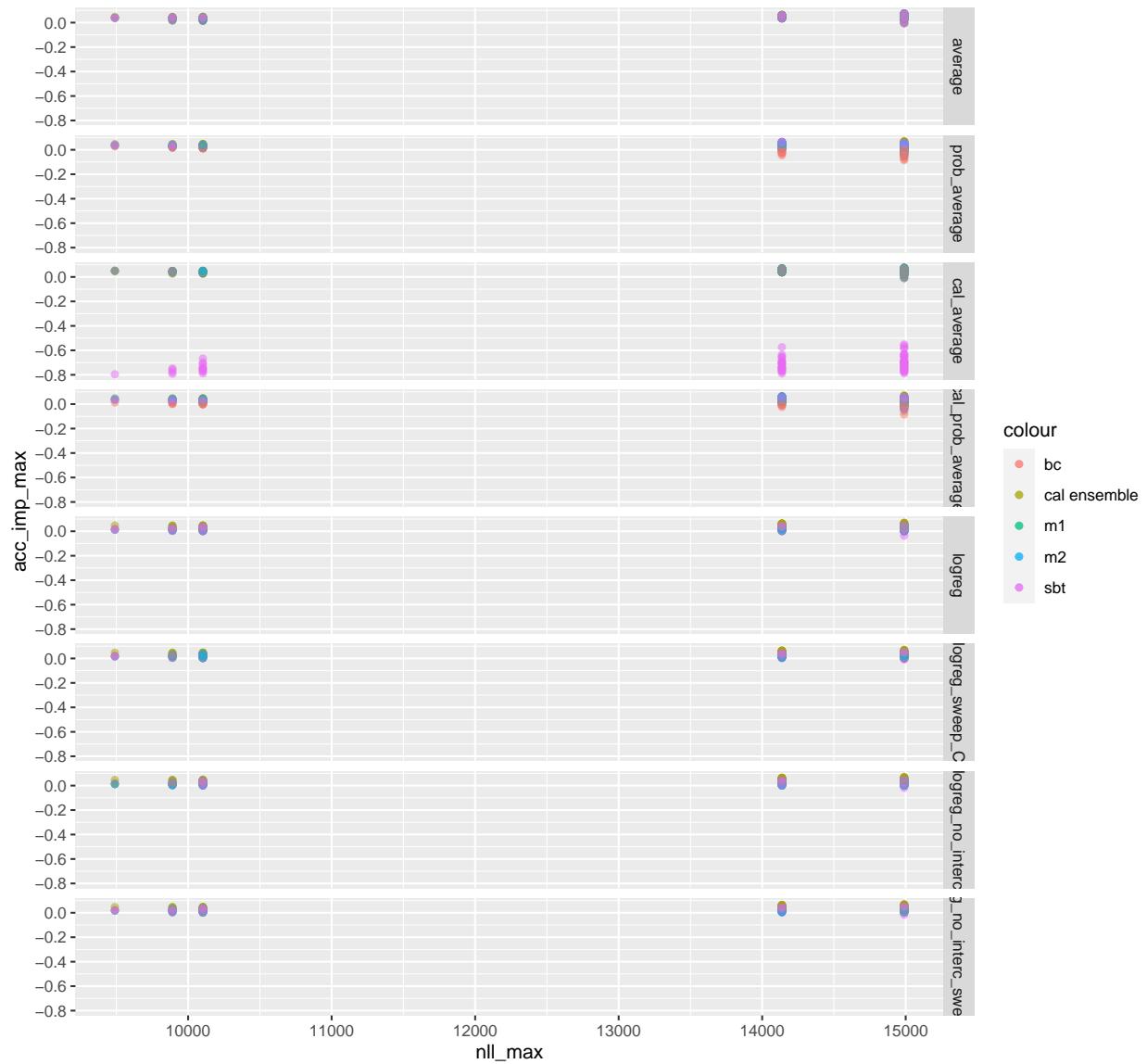
Accuracy improvement of ensemble over  
the best of networks vs nll\_min.  
Ensemble size 4



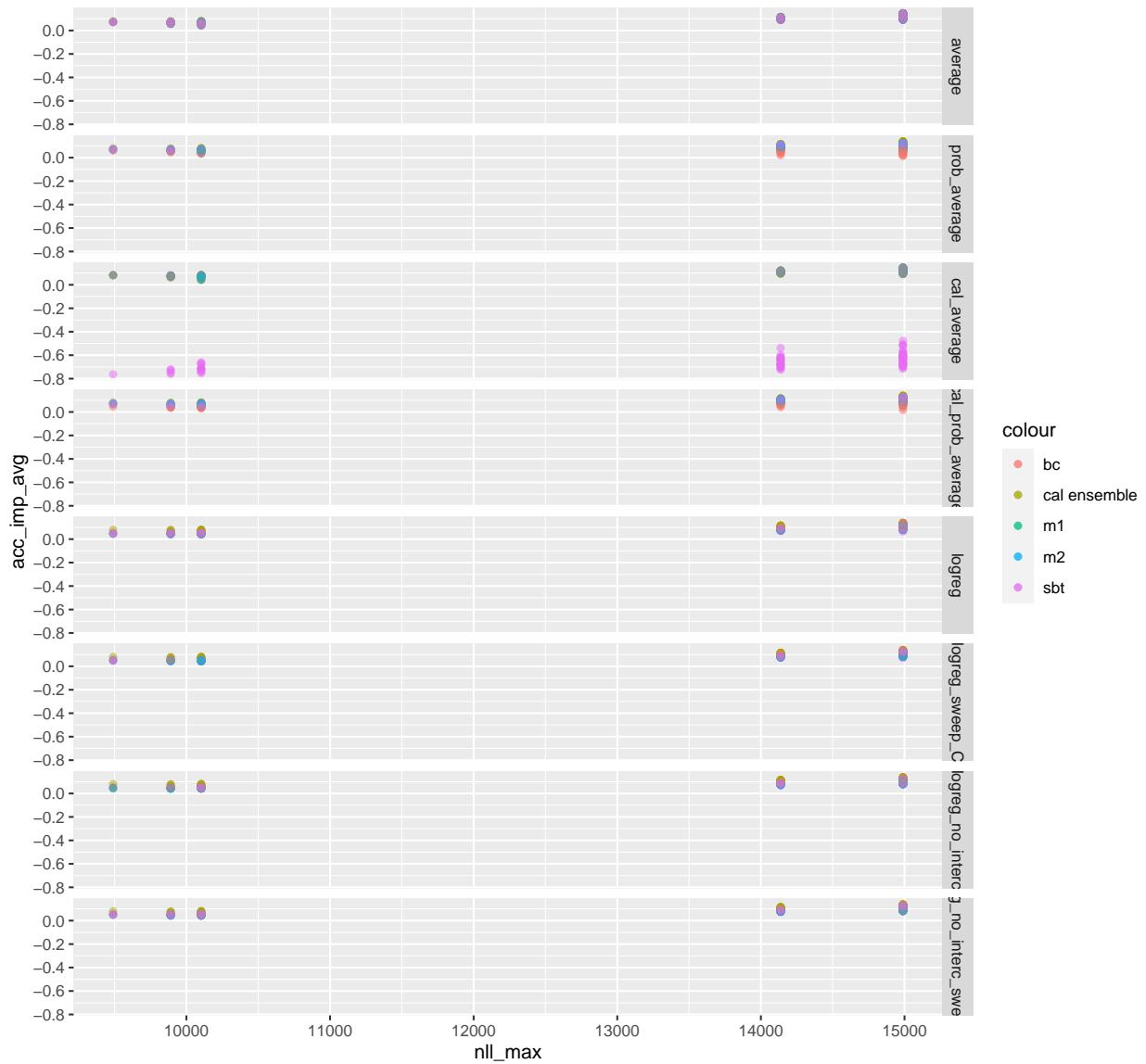
Accuracy improvement of ensemble over  
the average of networks vs nll\_min.  
Ensemble size 4



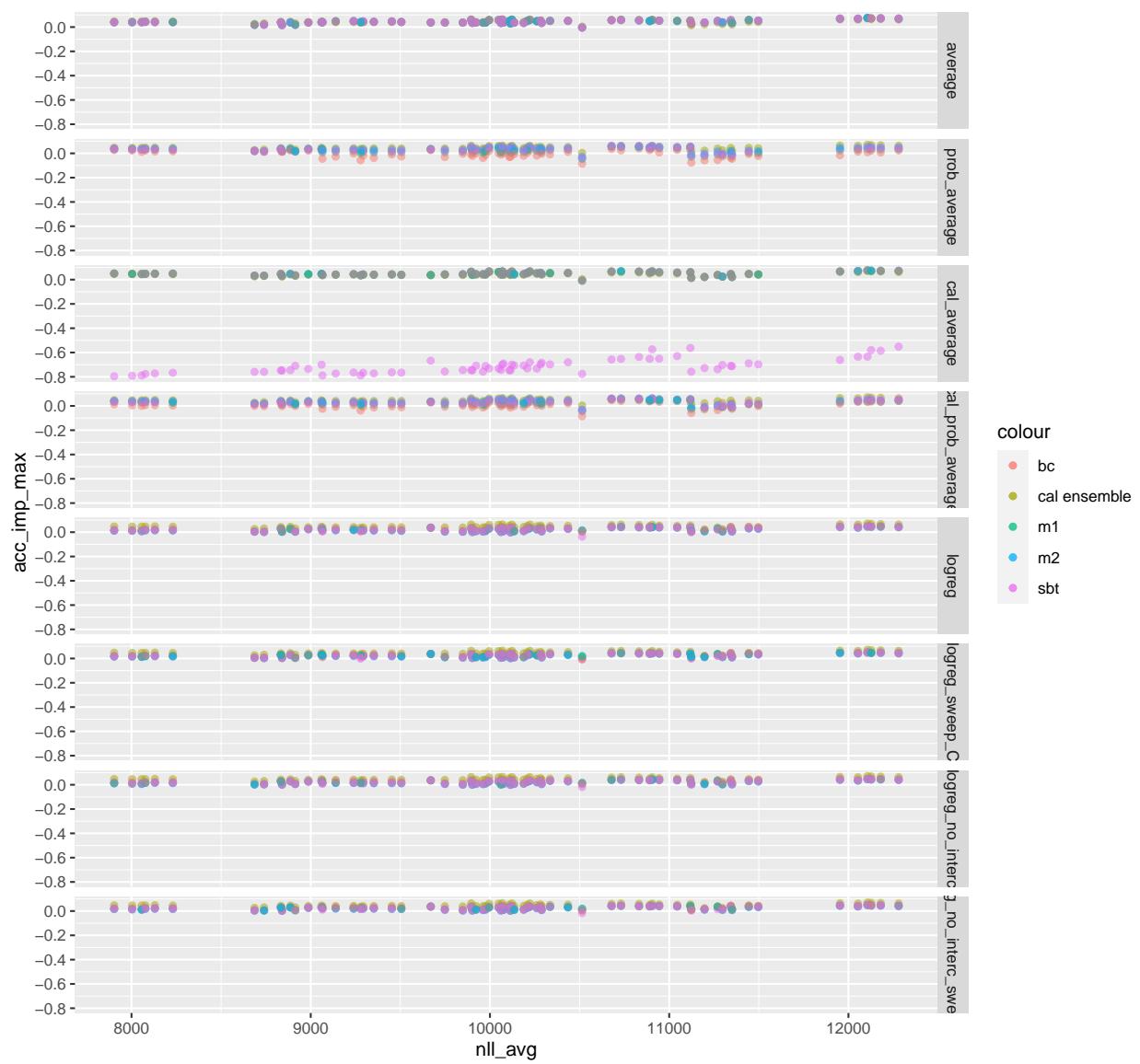
Accuracy improvement of ensemble over  
the best of networks vs nll\_max.  
Ensemble size 4



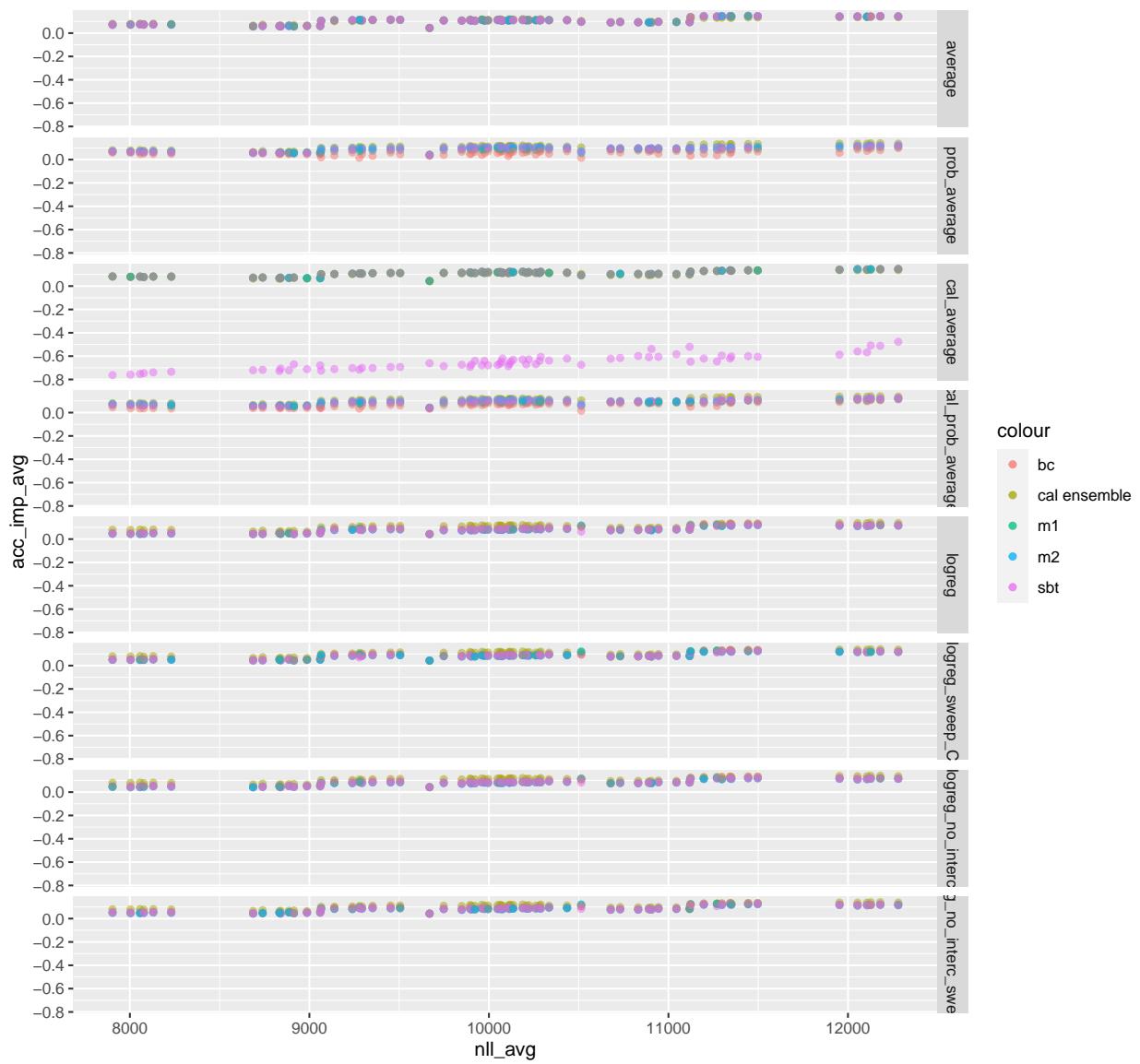
Accuracy improvement of ensemble over  
the average of networks vs nll\_max.  
Ensemble size 4



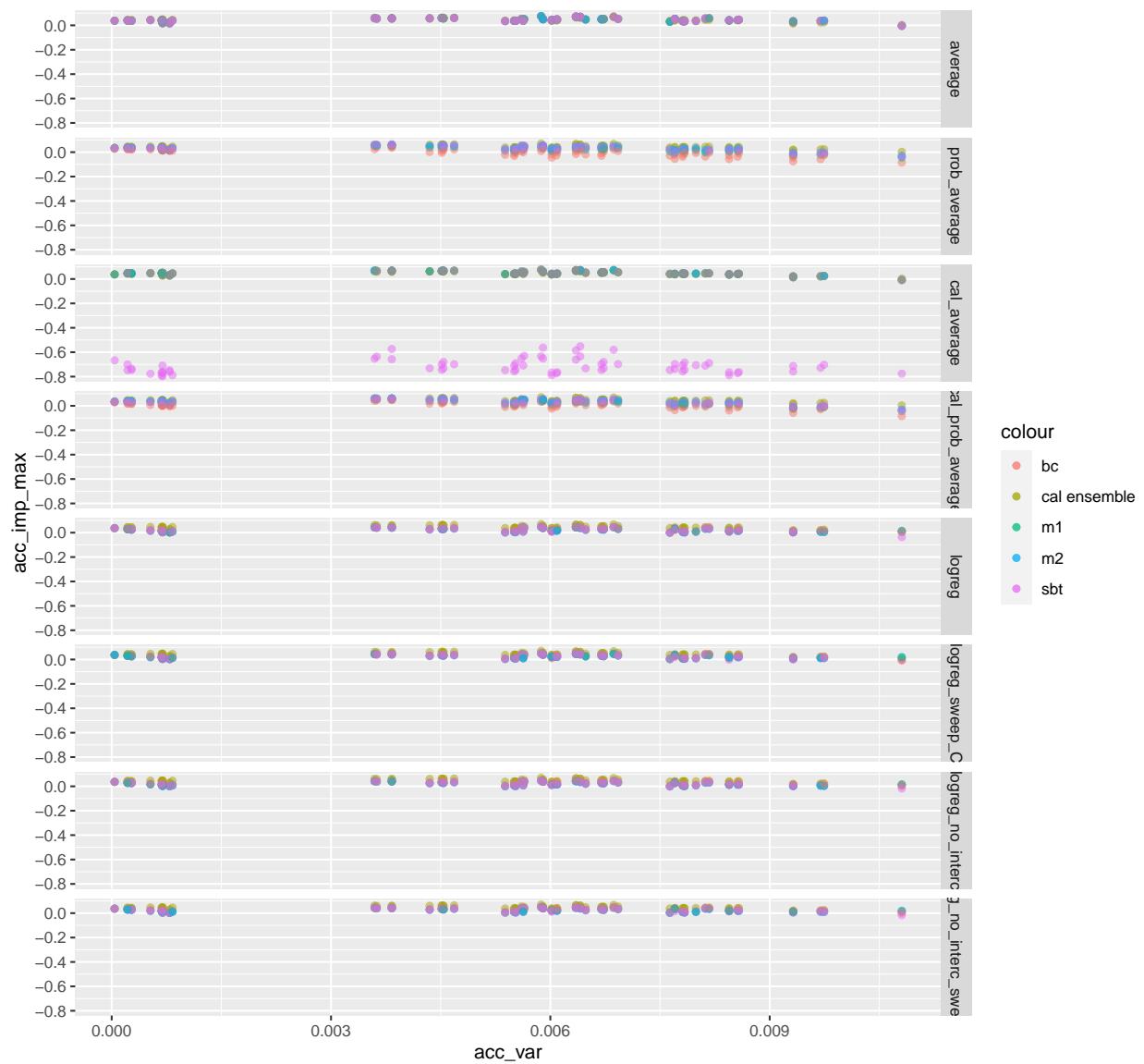
Accuracy improvement of ensemble over  
the best of networks vs nll\_avg.  
Ensemble size 4



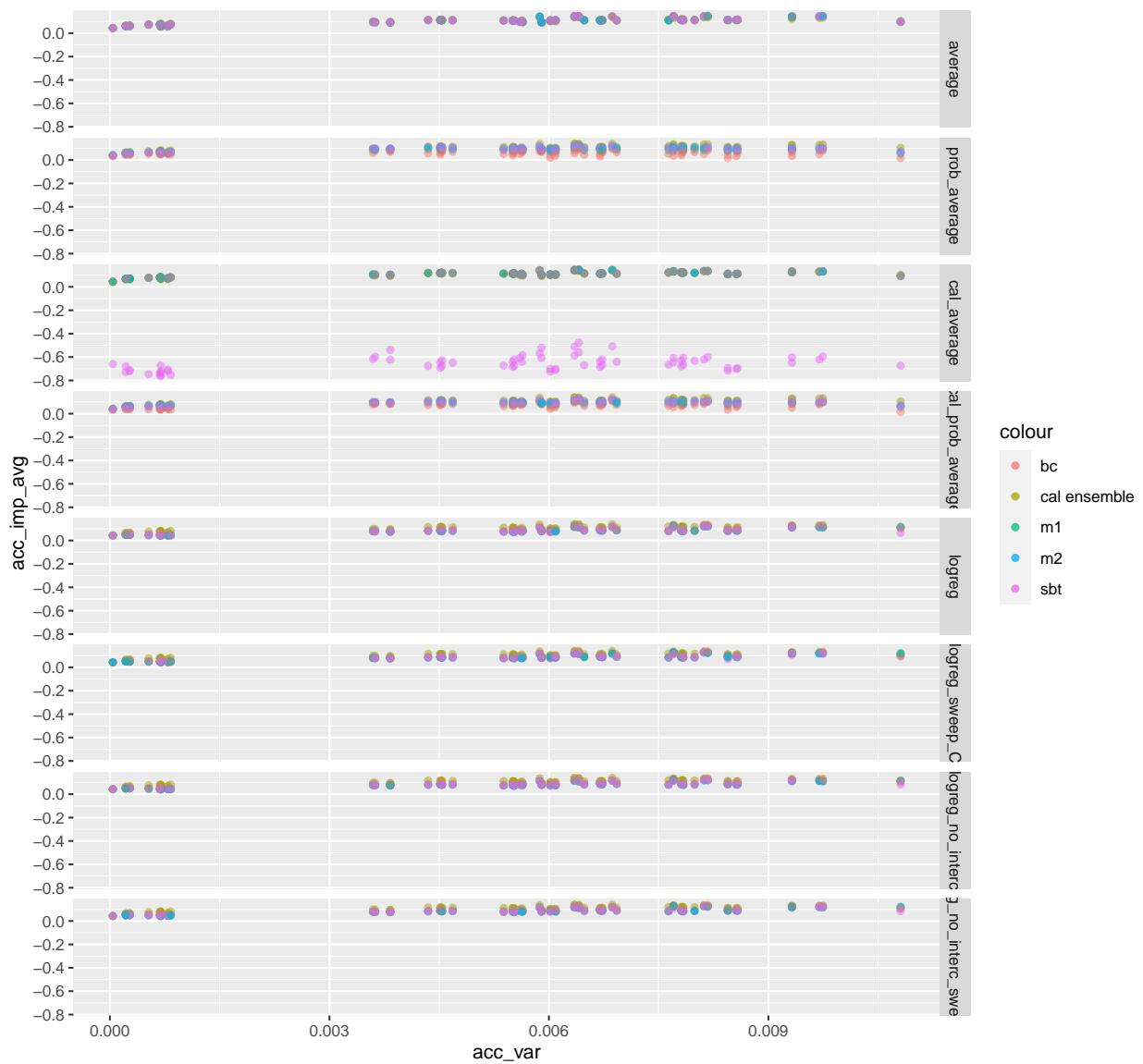
Accuracy improvement of ensemble over  
the average of networks vs nll\_avg.  
Ensemble size 4



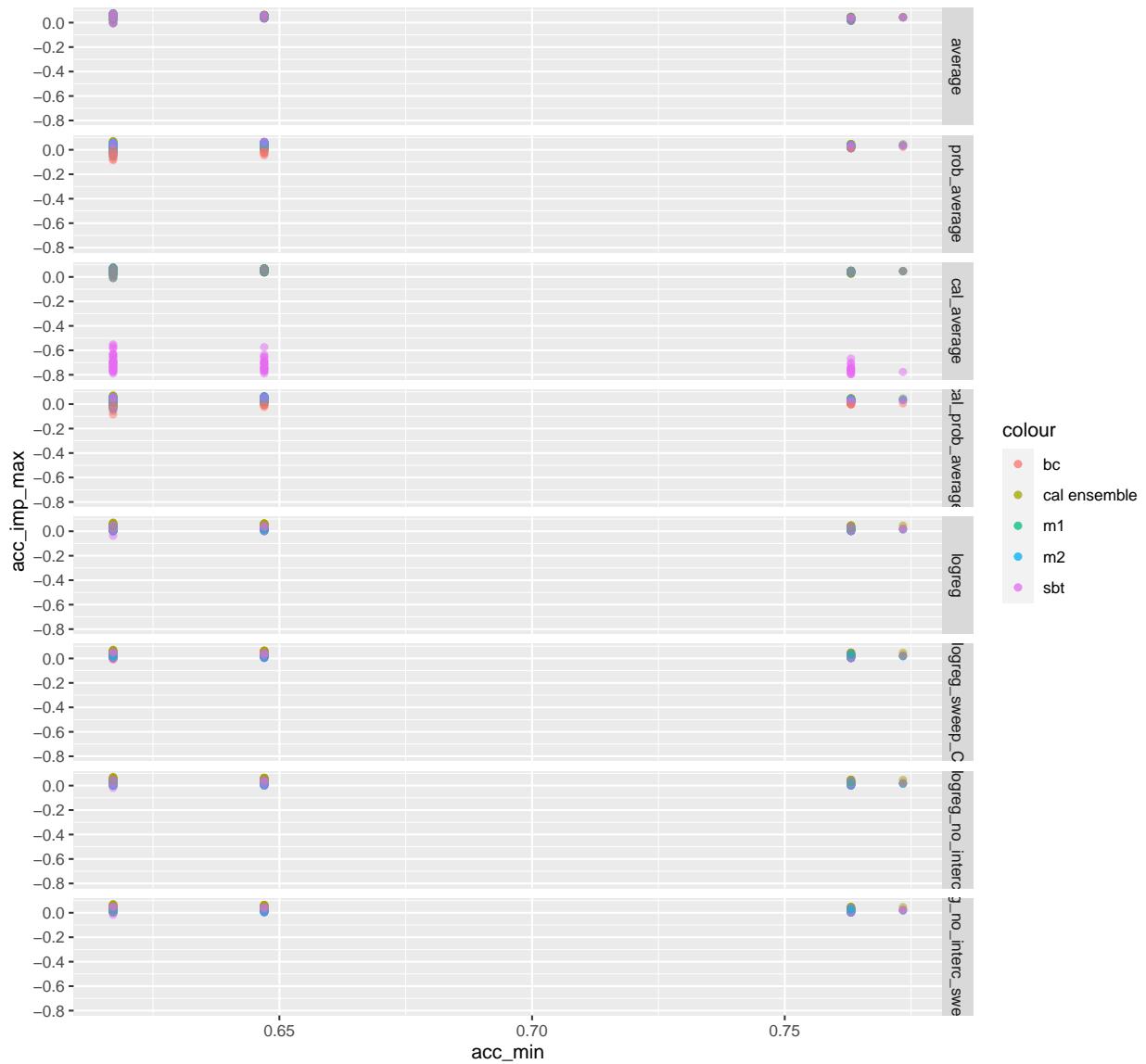
Accuracy improvement of ensemble over  
the best of networks vs acc\_var.  
Ensemble size 4



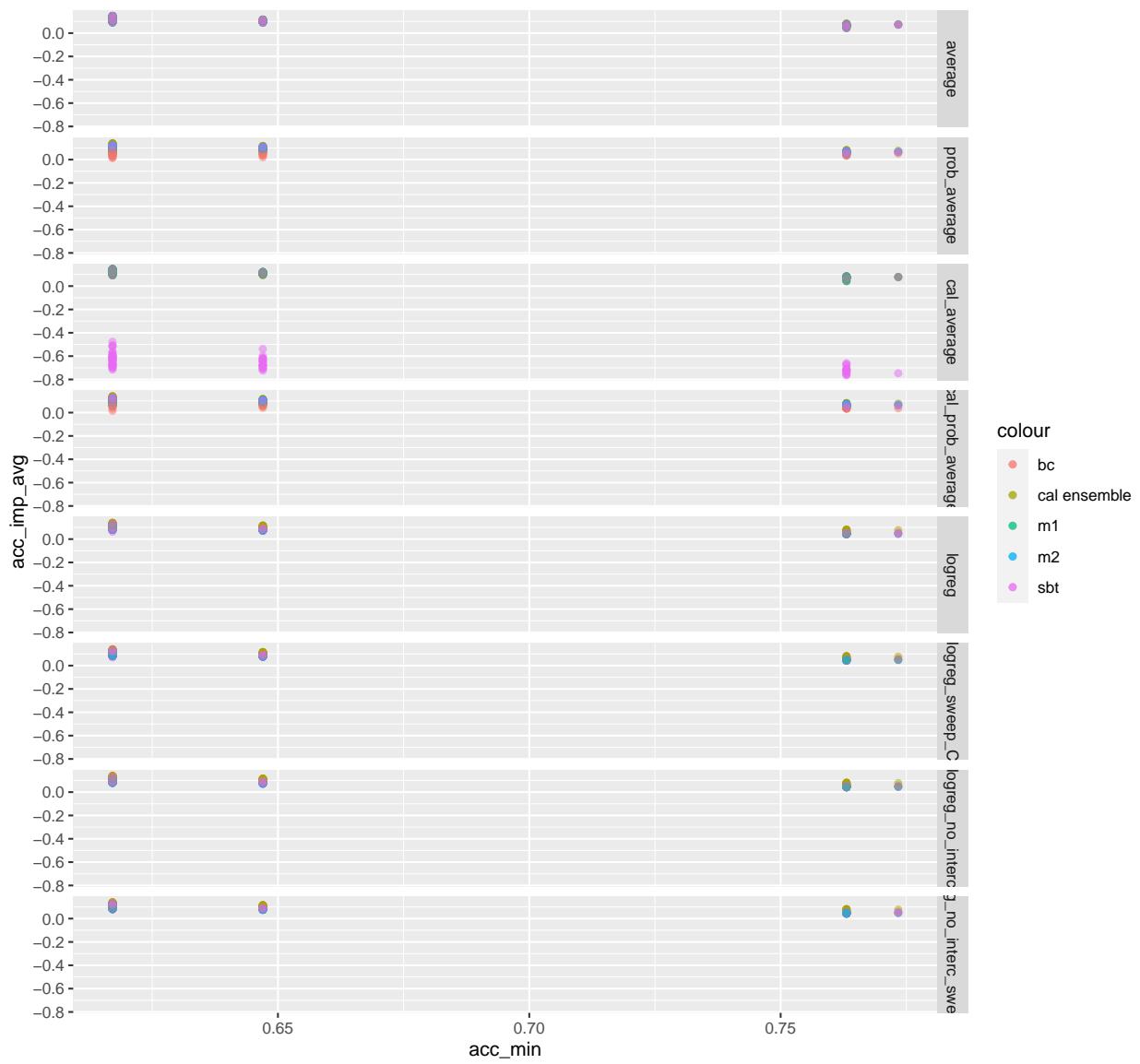
Accuracy improvement of ensemble over  
the average of networks vs acc\_var.  
Ensemble size 4



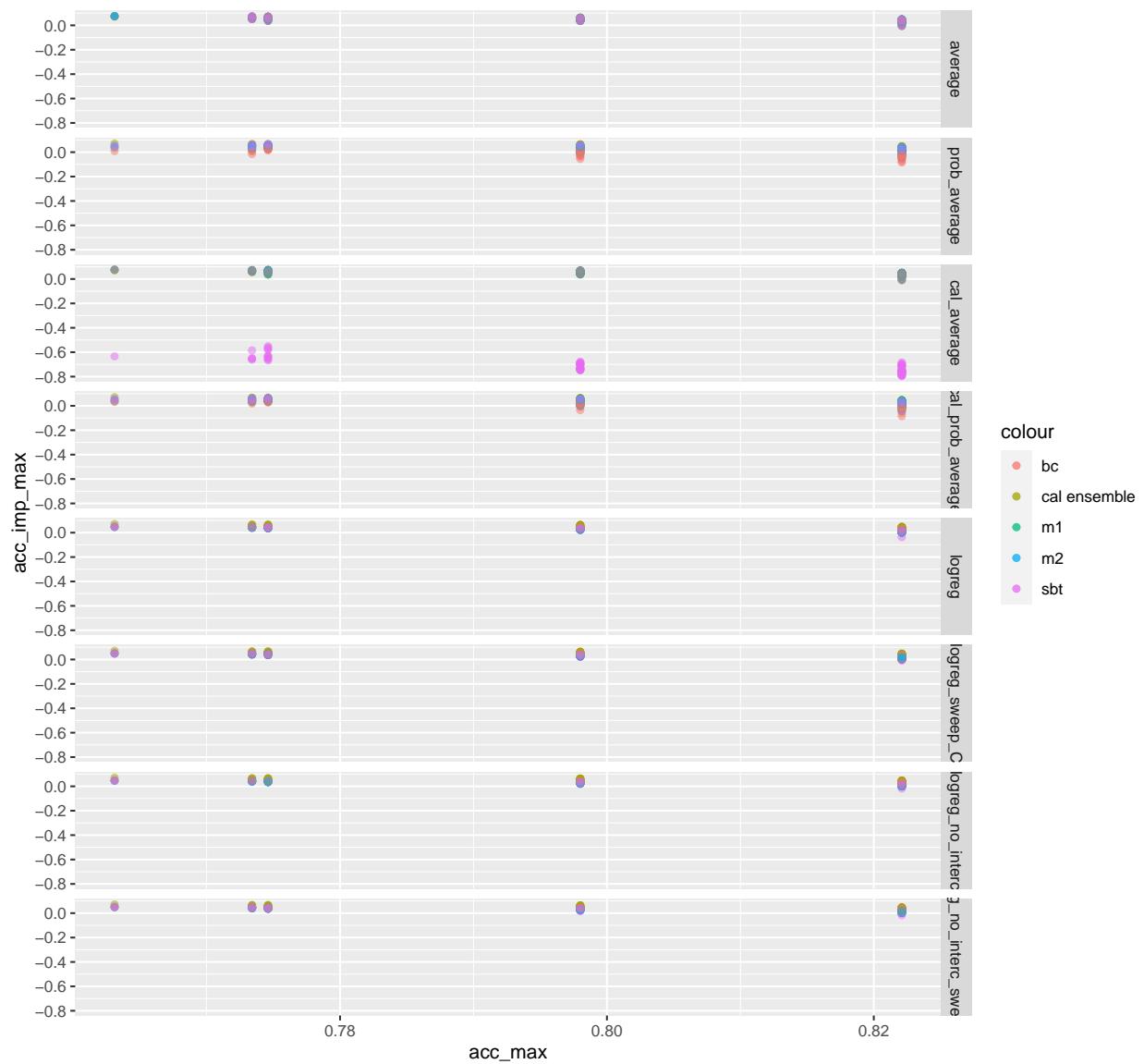
Accuracy improvement of ensemble over  
the best of networks vs acc\_min.  
Ensemble size 4



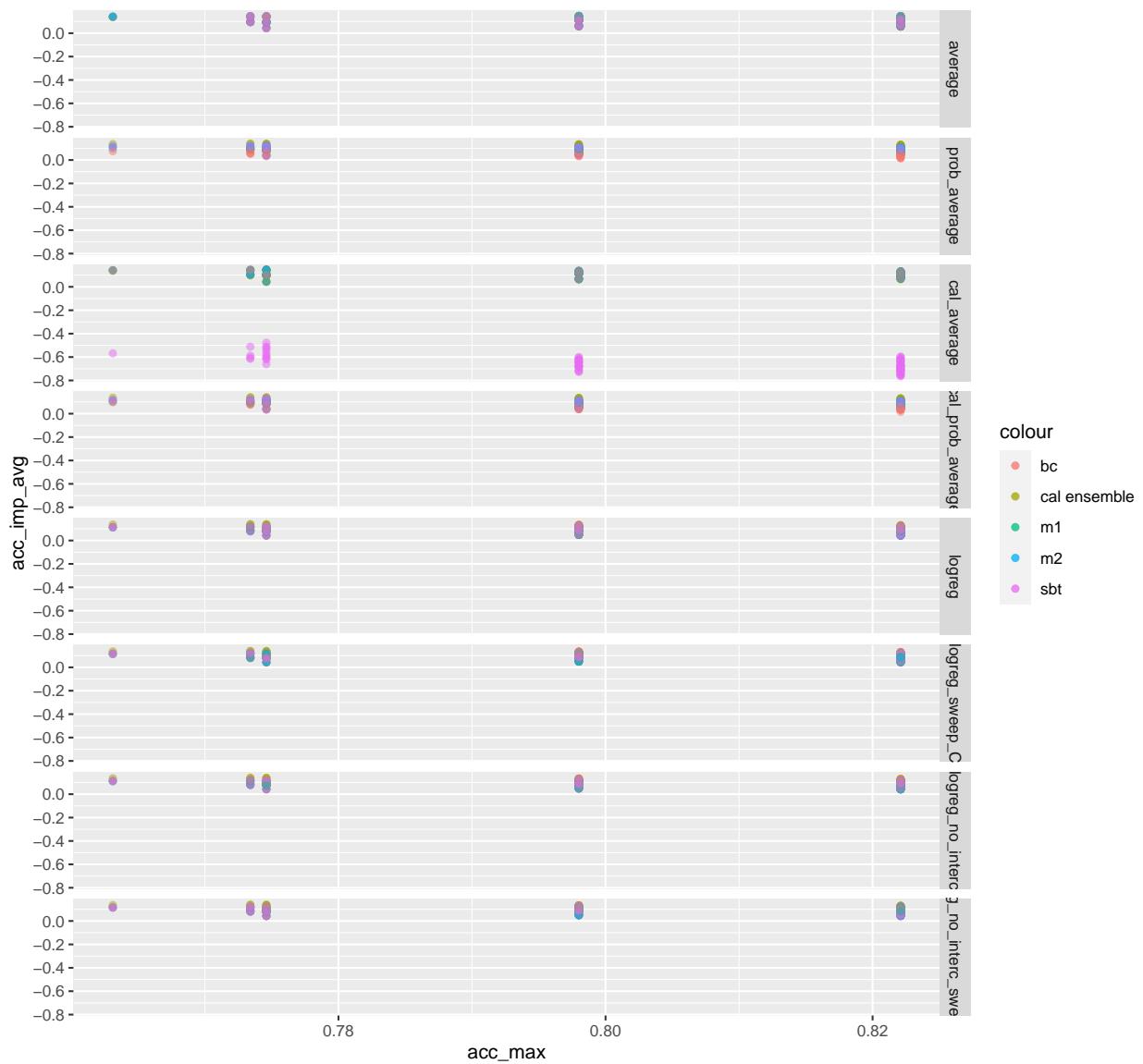
Accuracy improvement of ensemble over  
the average of networks vs acc\_min.  
Ensemble size 4



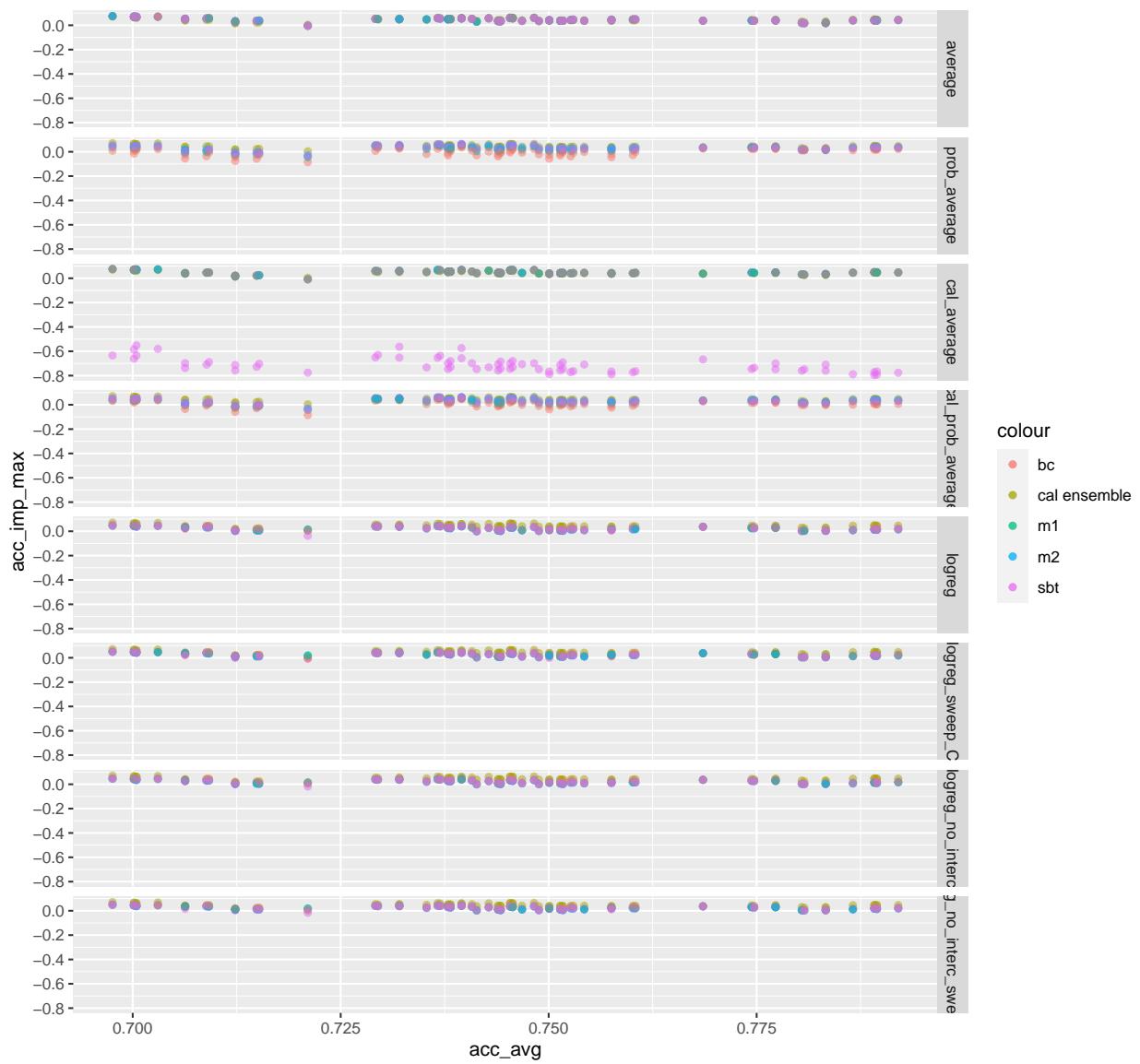
Accuracy improvement of ensemble over  
the best of networks vs acc\_max.  
Ensemble size 4



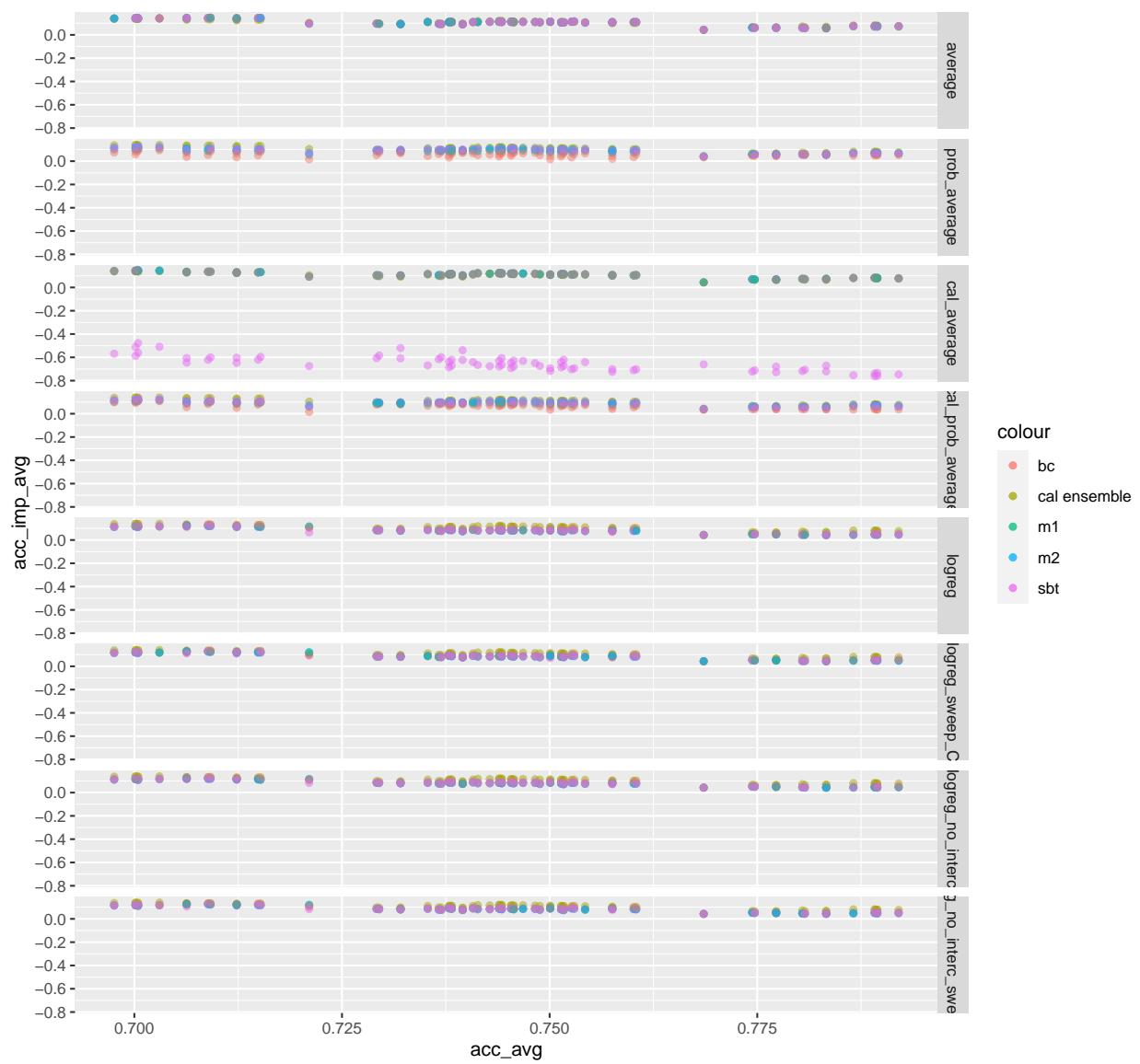
Accuracy improvement of ensemble over  
the average of networks vs acc\_max.  
Ensemble size 4



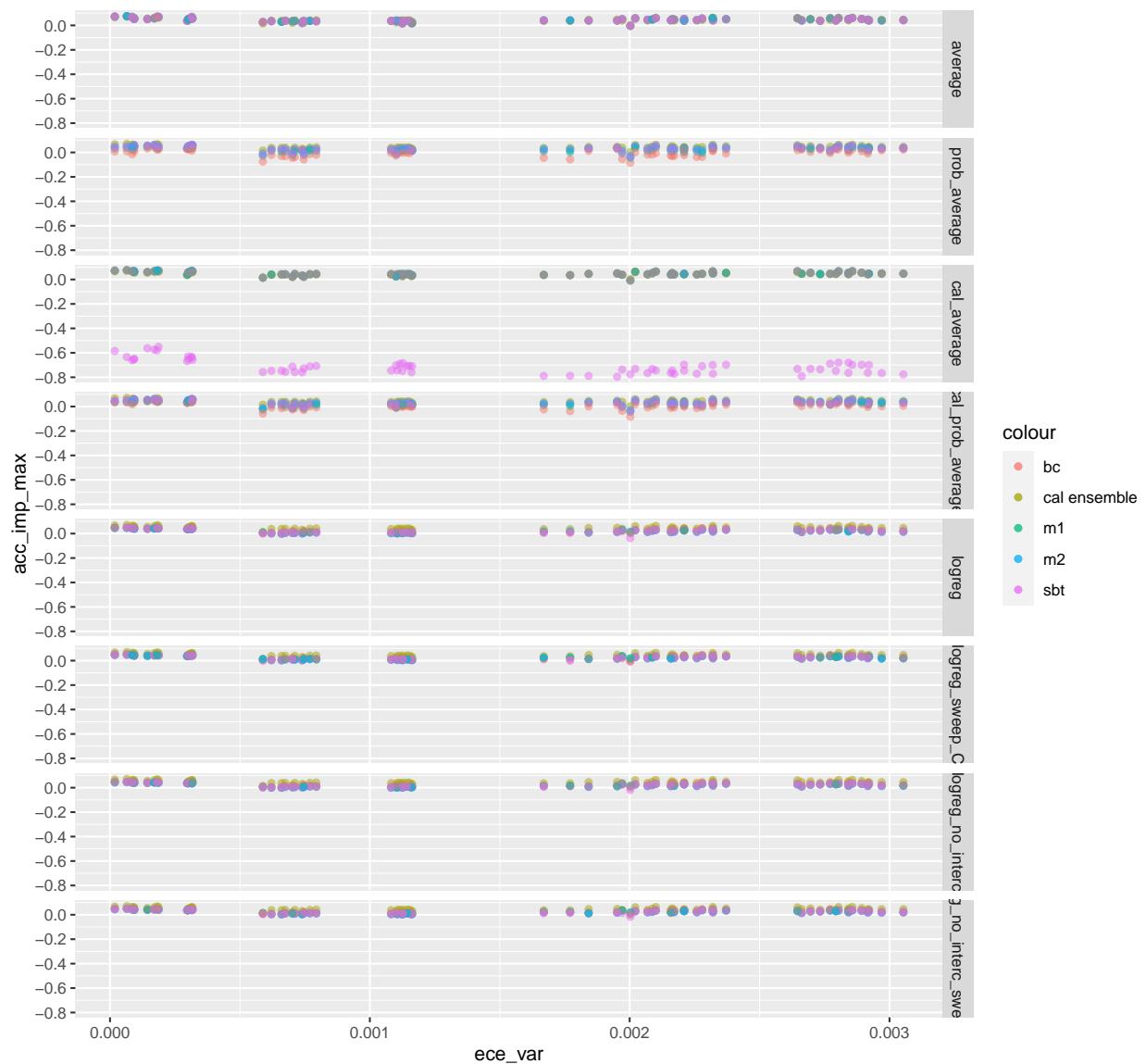
Accuracy improvement of ensemble over  
the best of networks vs acc\_avg.  
Ensemble size 4



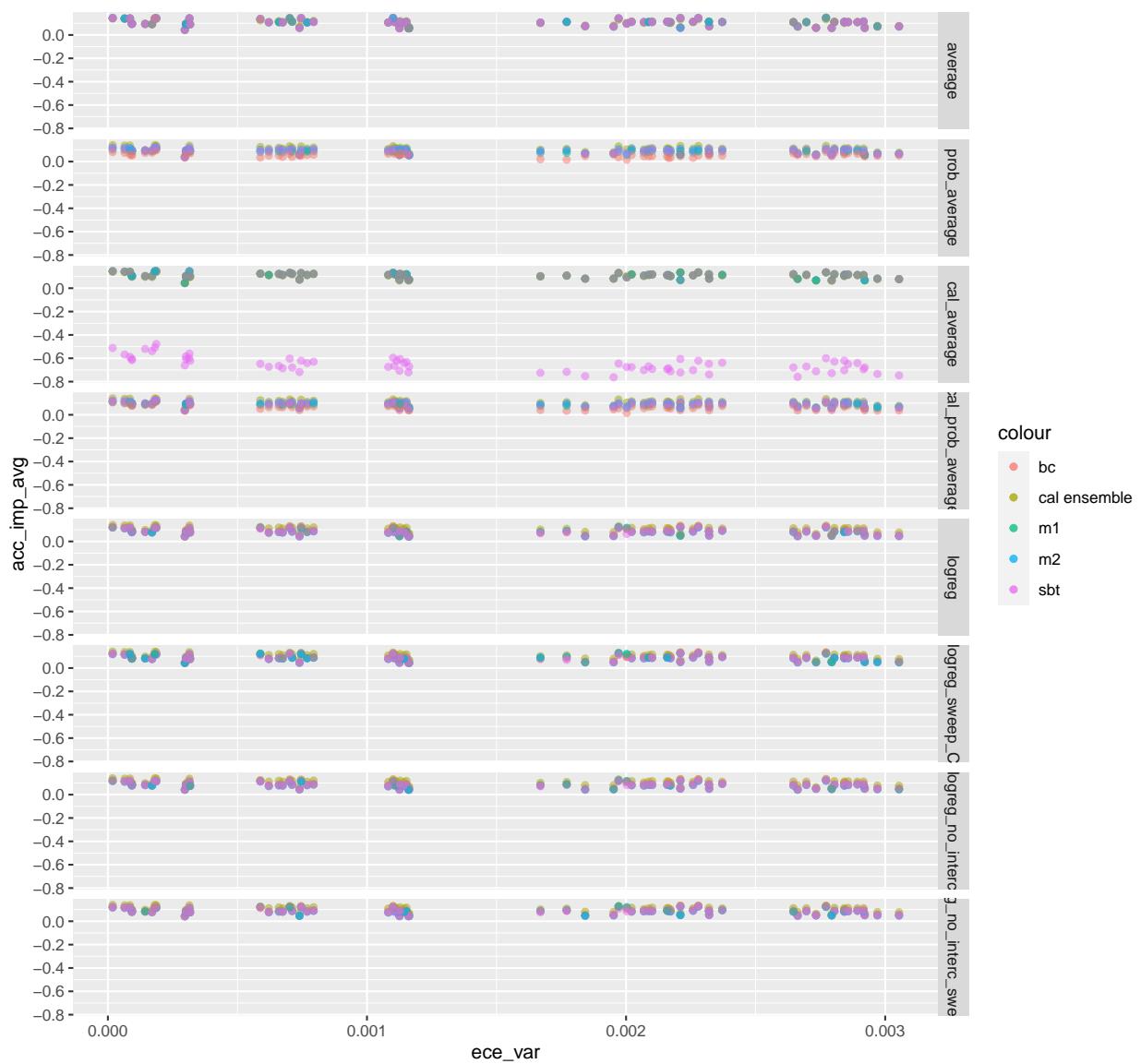
Accuracy improvement of ensemble over  
the average of networks vs acc\_avg.  
Ensemble size 4



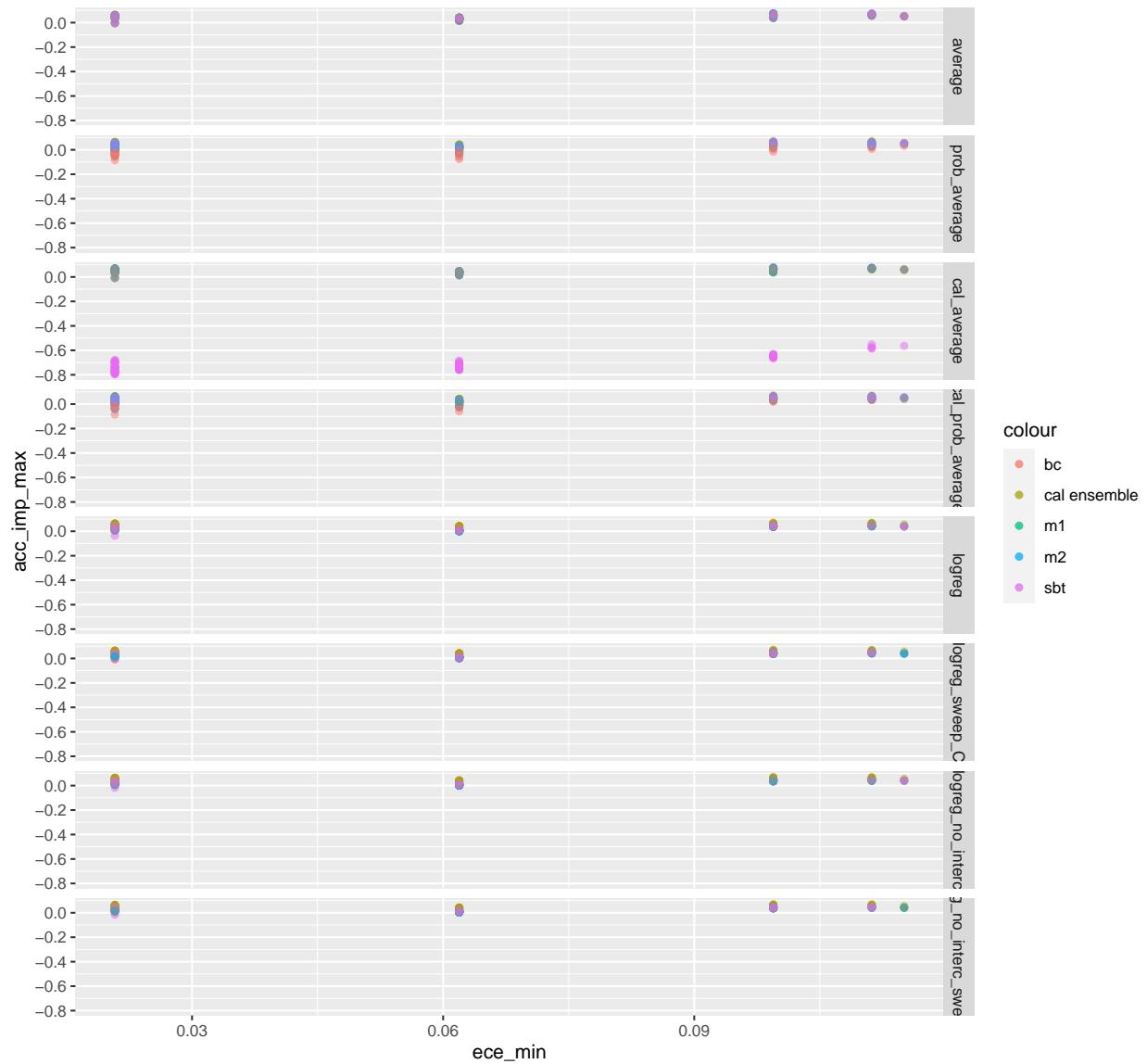
Accuracy improvement of ensemble over  
the best of networks vs ece\_var.  
Ensemble size 4



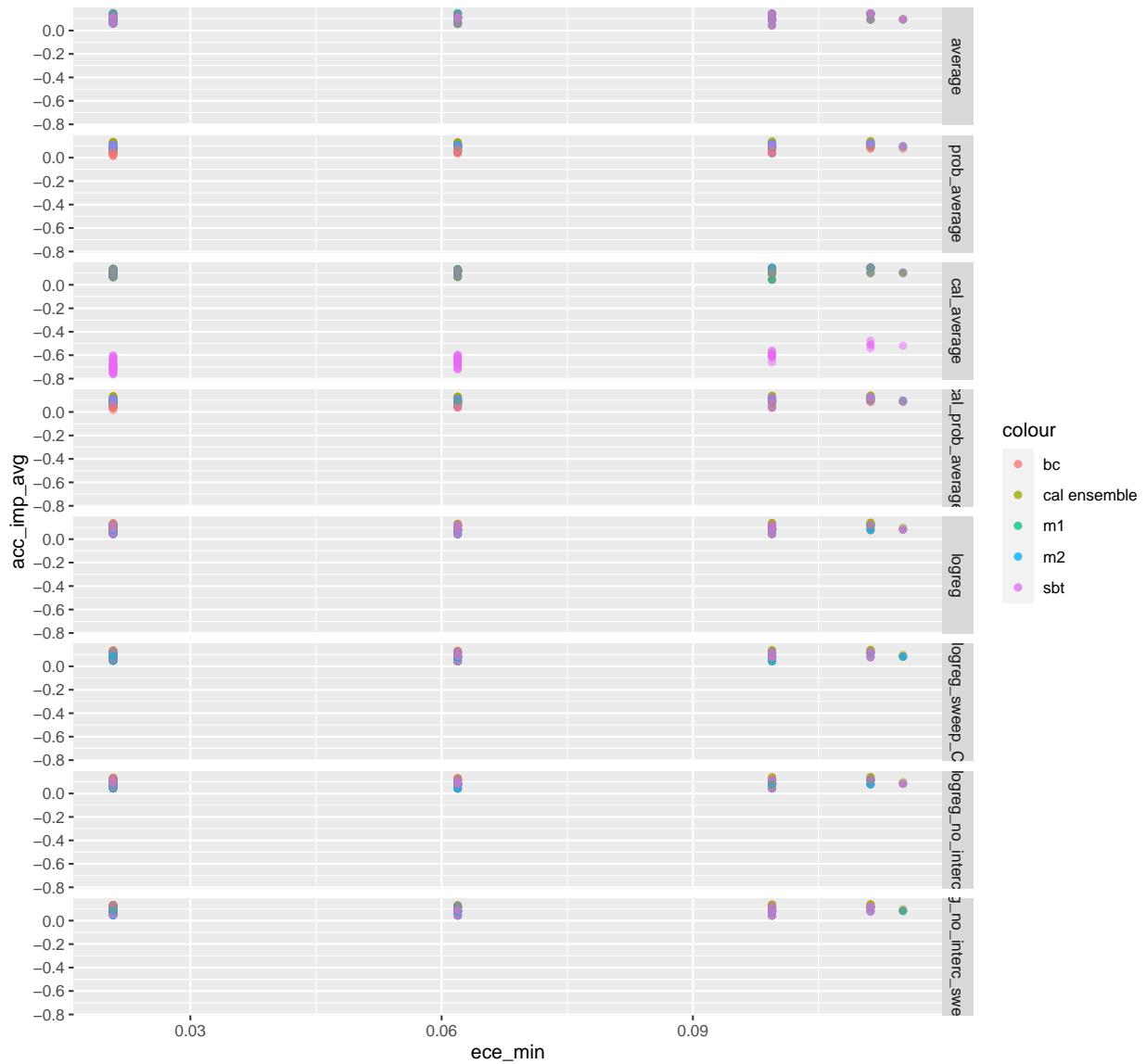
Accuracy improvement of ensemble over  
the average of networks vs ece\_var.  
Ensemble size 4



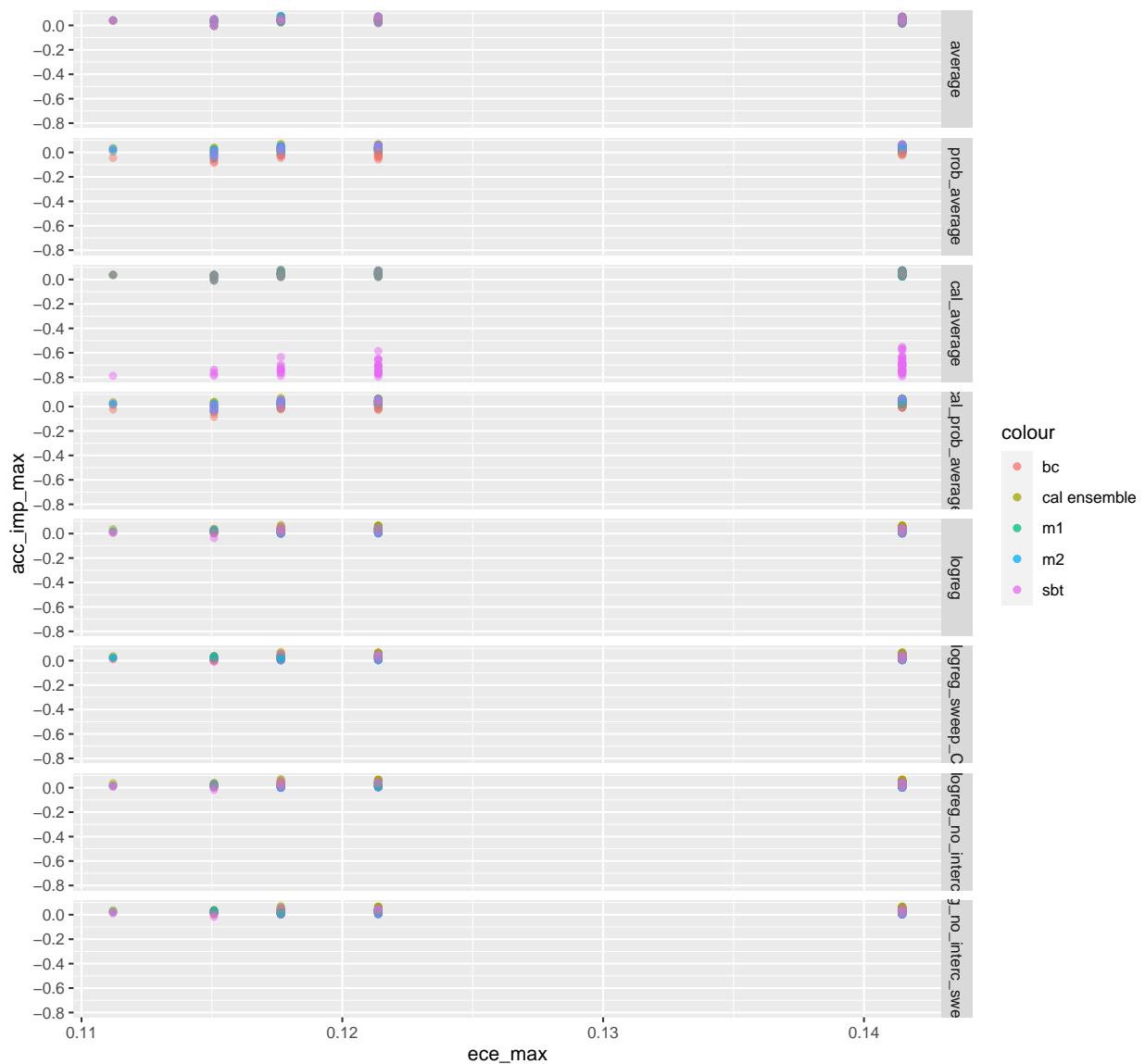
Accuracy improvement of ensemble over  
the best of networks vs ece\_min.  
Ensemble size 4



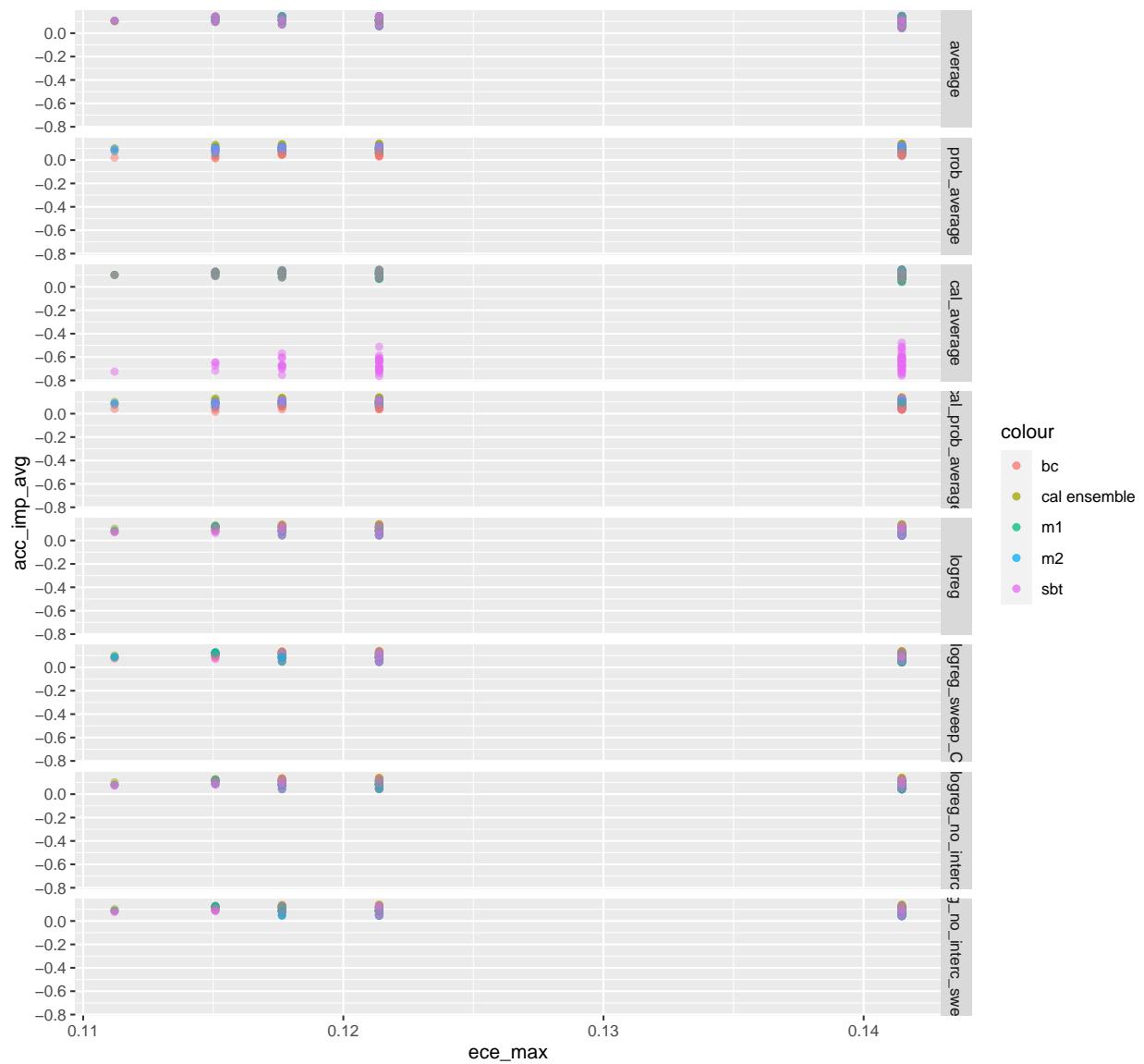
Accuracy improvement of ensemble over  
the average of networks vs ece\_min.  
Ensemble size 4



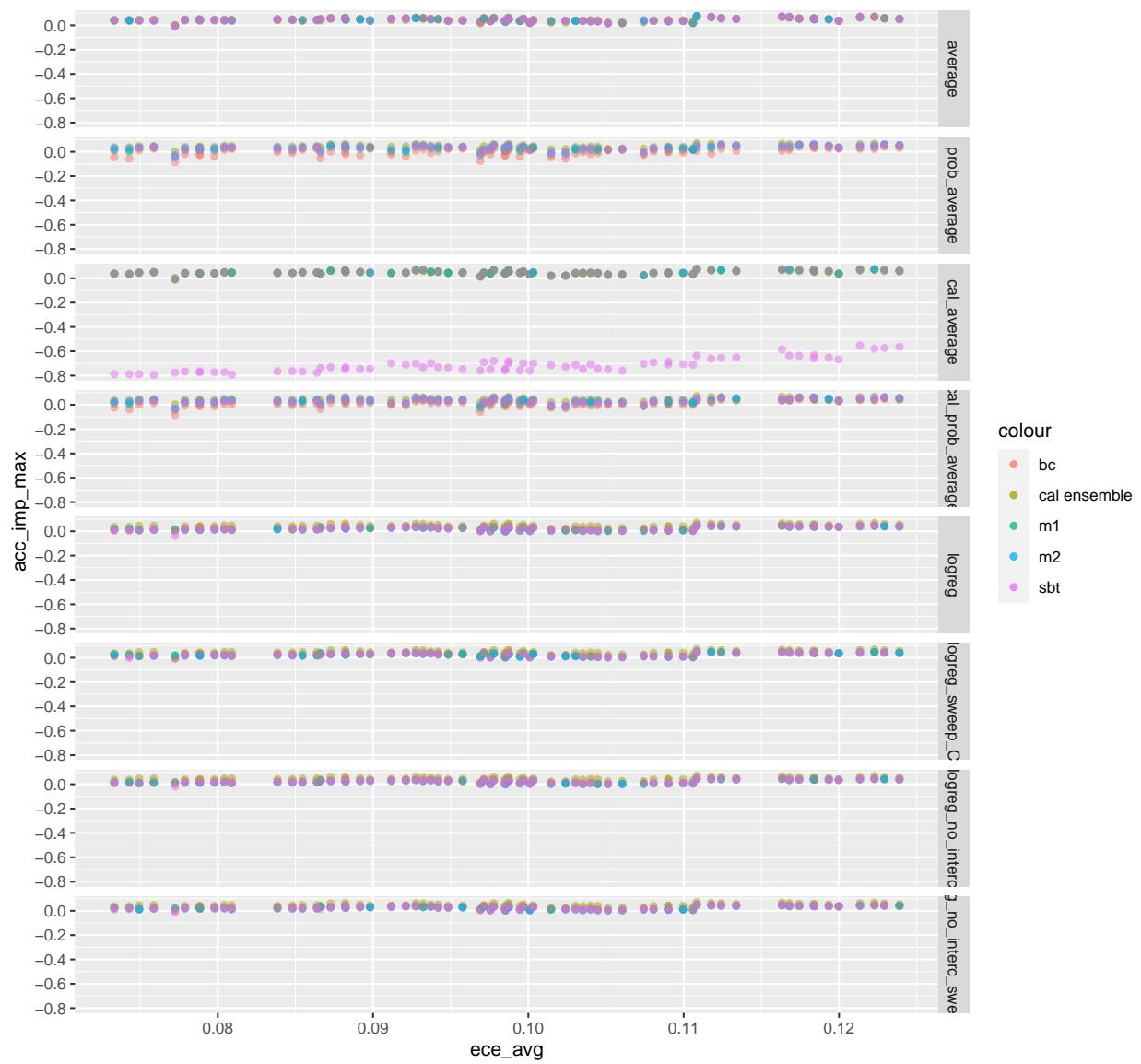
Accuracy improvement of ensemble over  
the best of networks vs ece\_max.  
Ensemble size 4



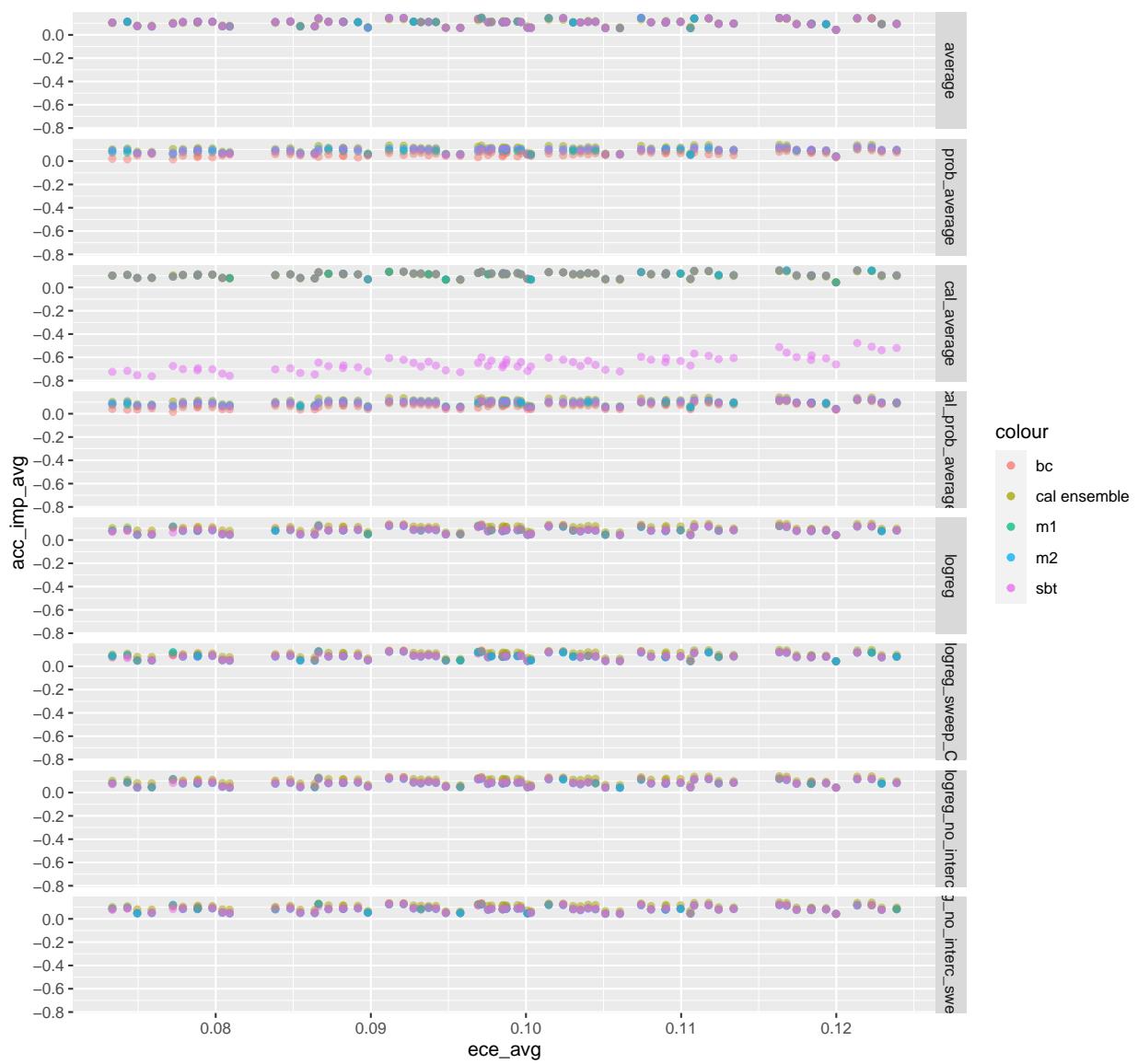
Accuracy improvement of ensemble over  
the average of networks vs ece\_max.  
Ensemble size 4



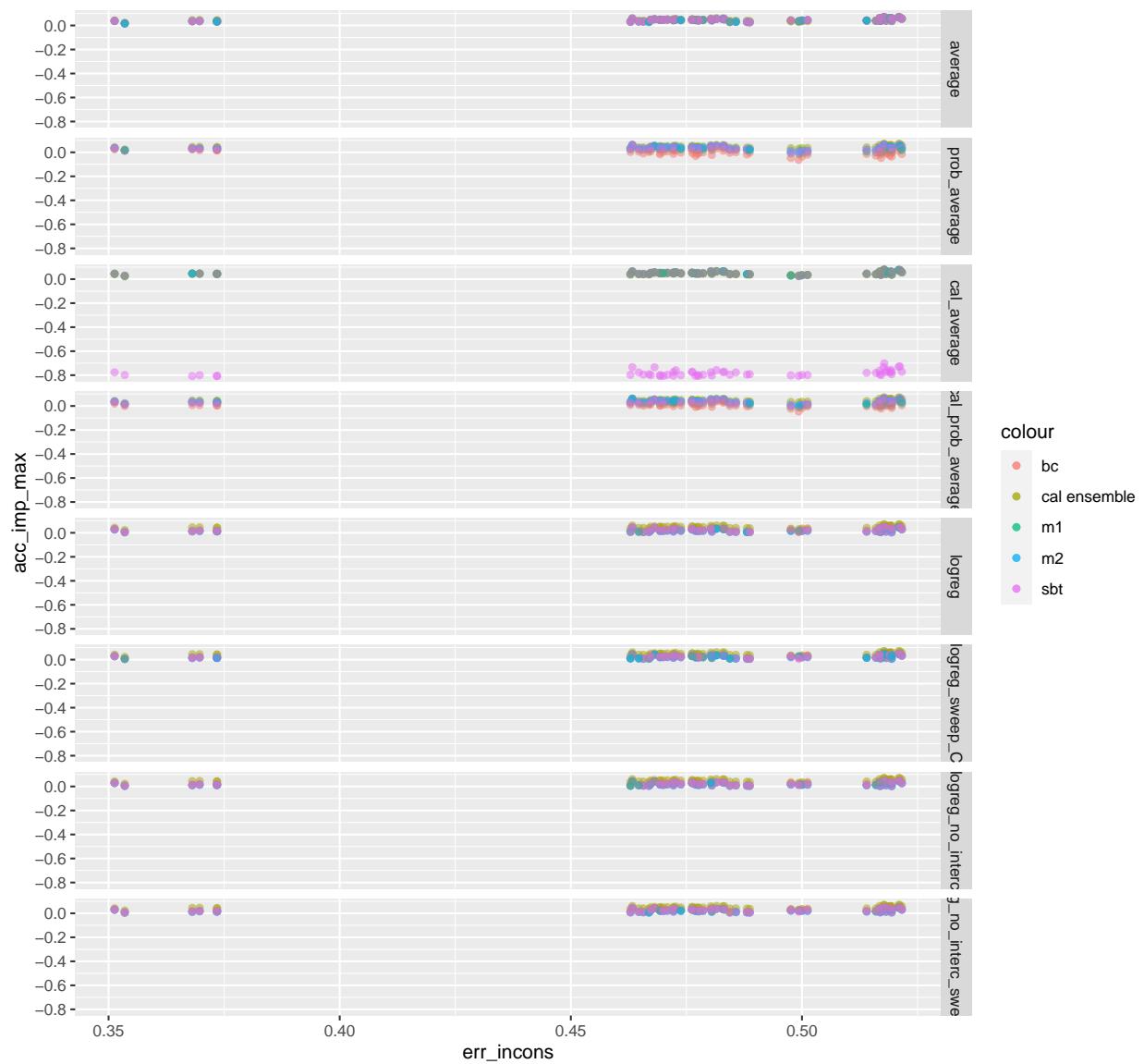
Accuracy improvement of ensemble over  
the best of networks vs ece\_avg.  
Ensemble size 4



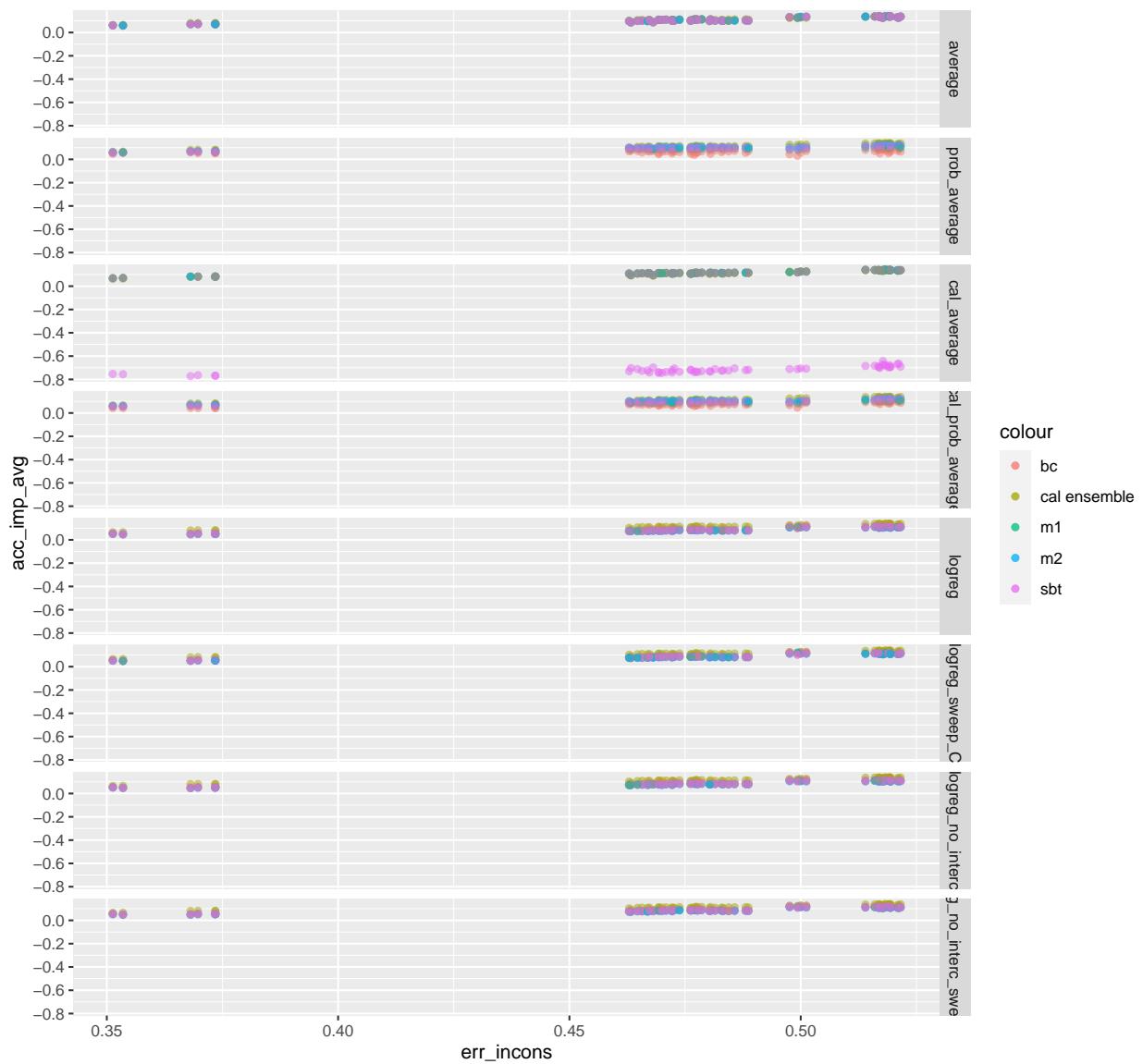
Accuracy improvement of ensemble over  
the average of networks vs ece\_avg.  
Ensemble size 4



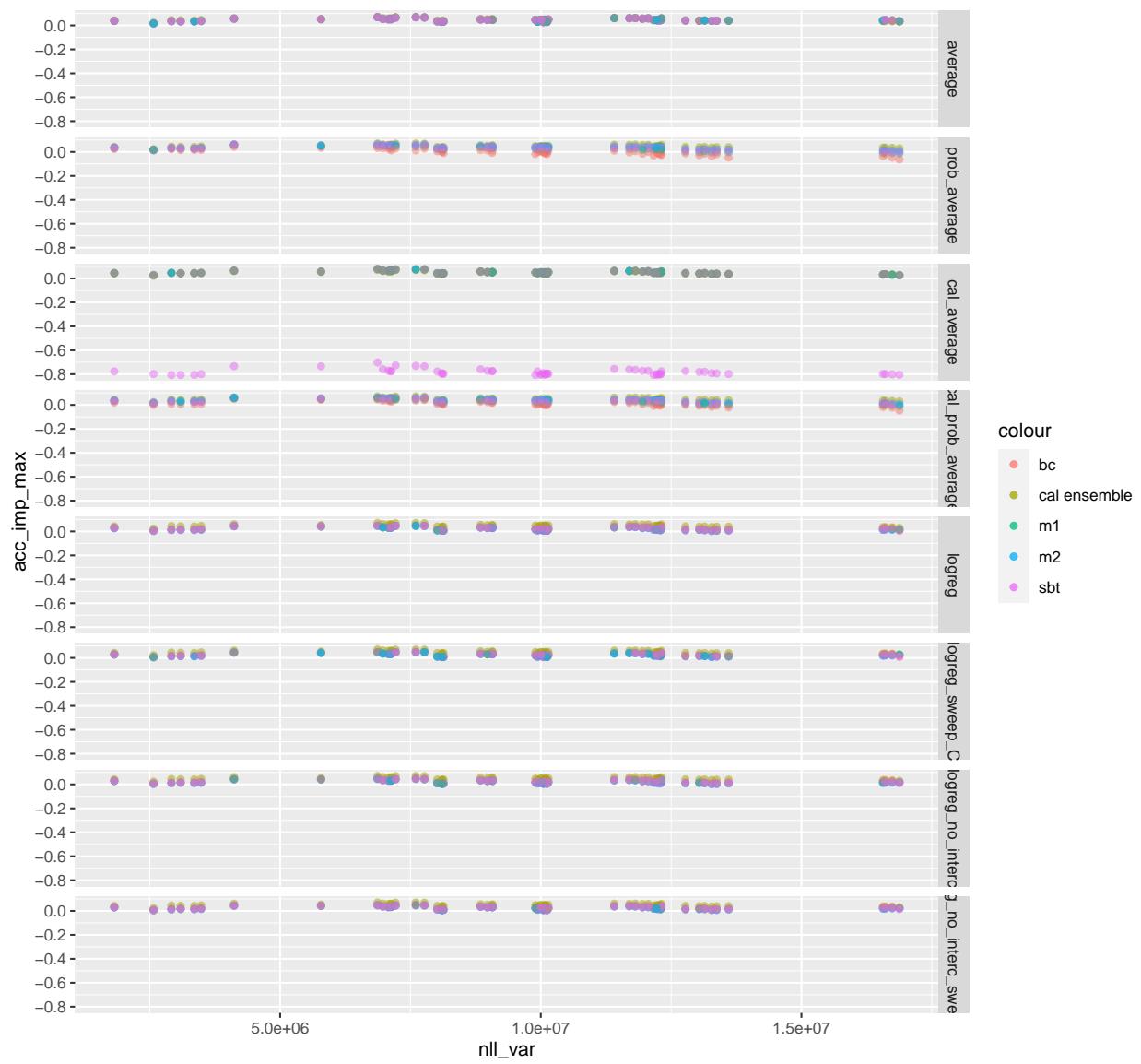
Accuracy improvement of ensemble over  
the best of networks vs err\_incons.  
Ensemble size 5



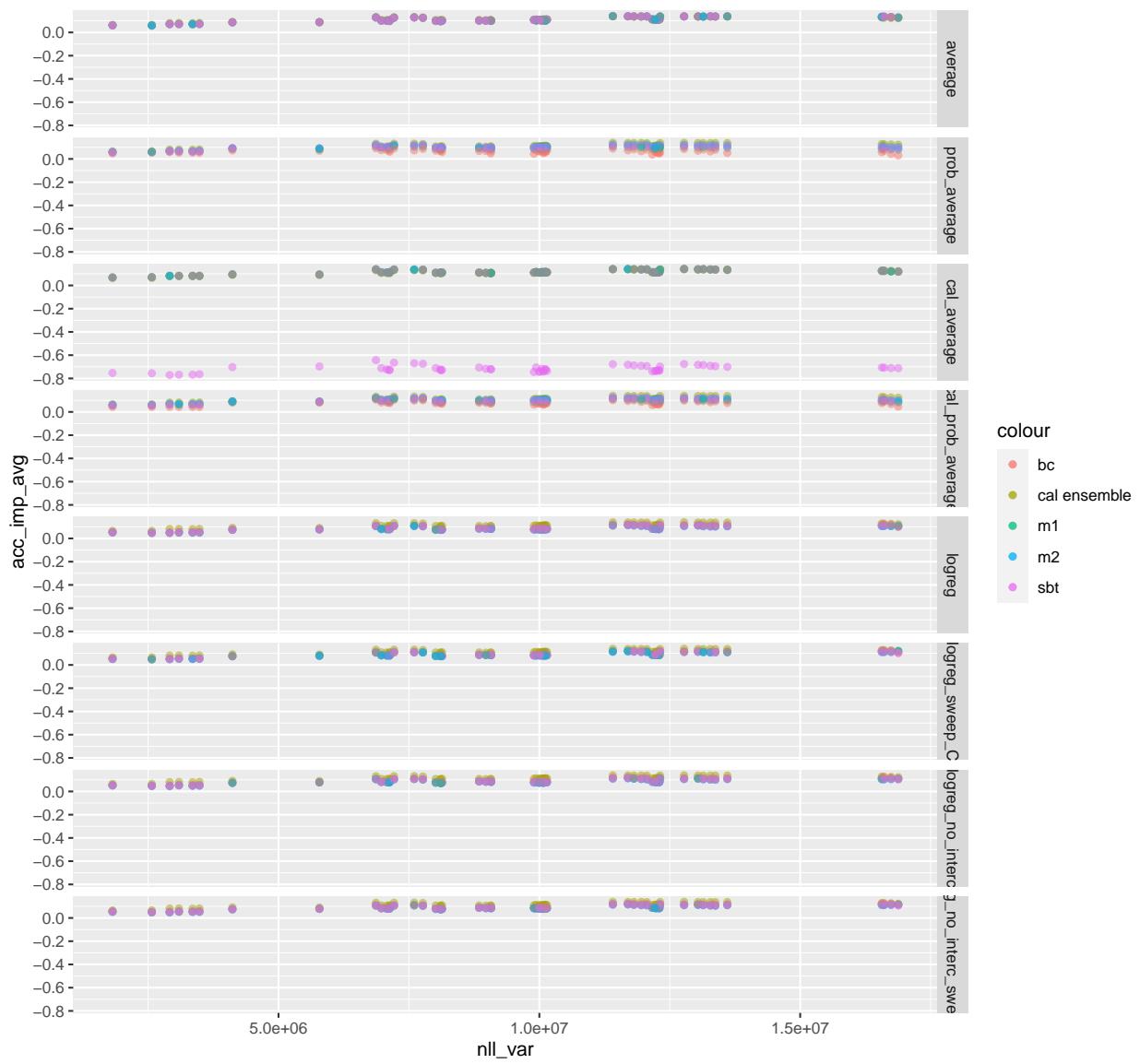
Accuracy improvement of ensemble over  
the average of networks vs err\_incons.  
Ensemble size 5



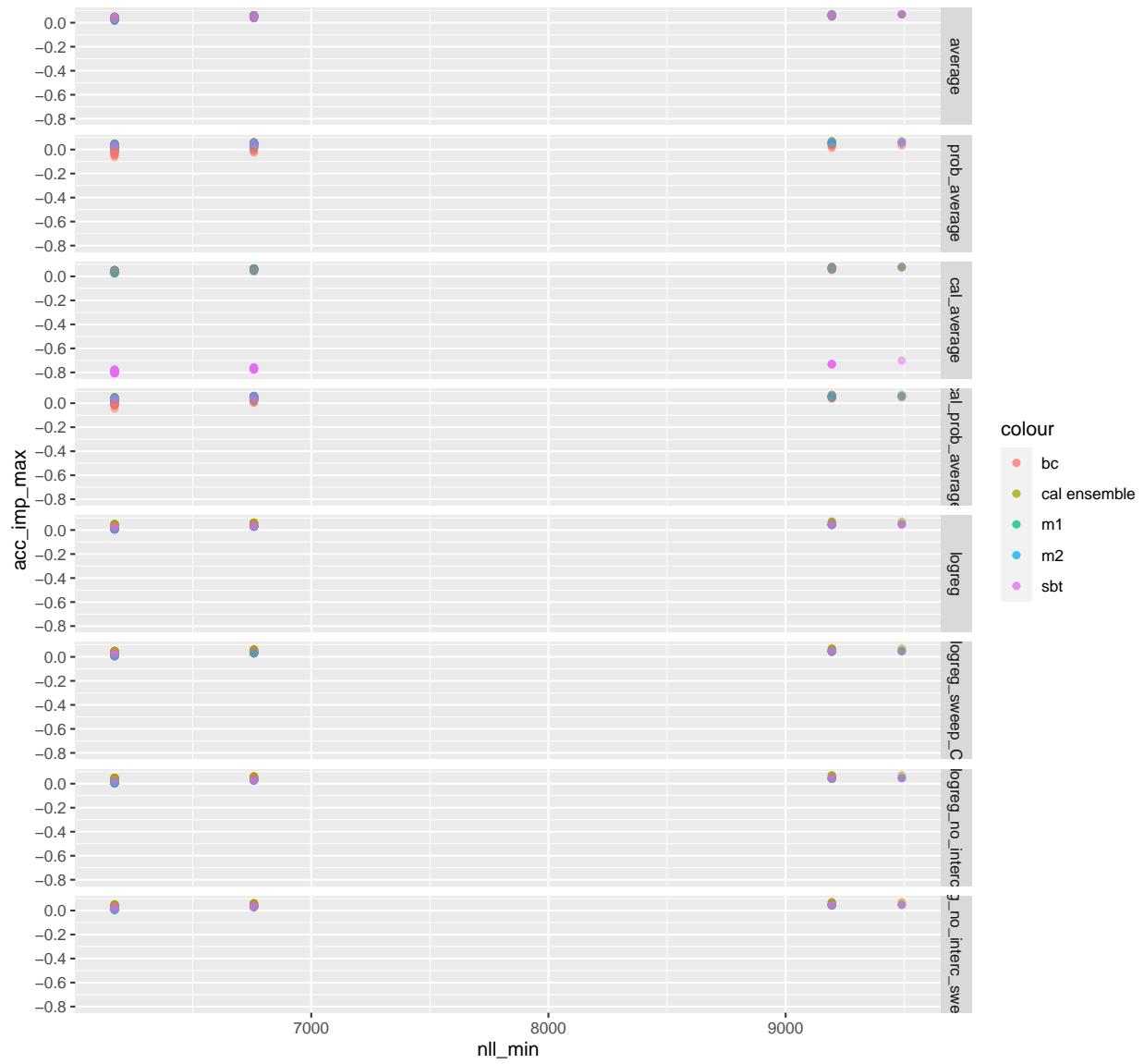
Accuracy improvement of ensemble over  
the best of networks vs nll\_var.  
Ensemble size 5



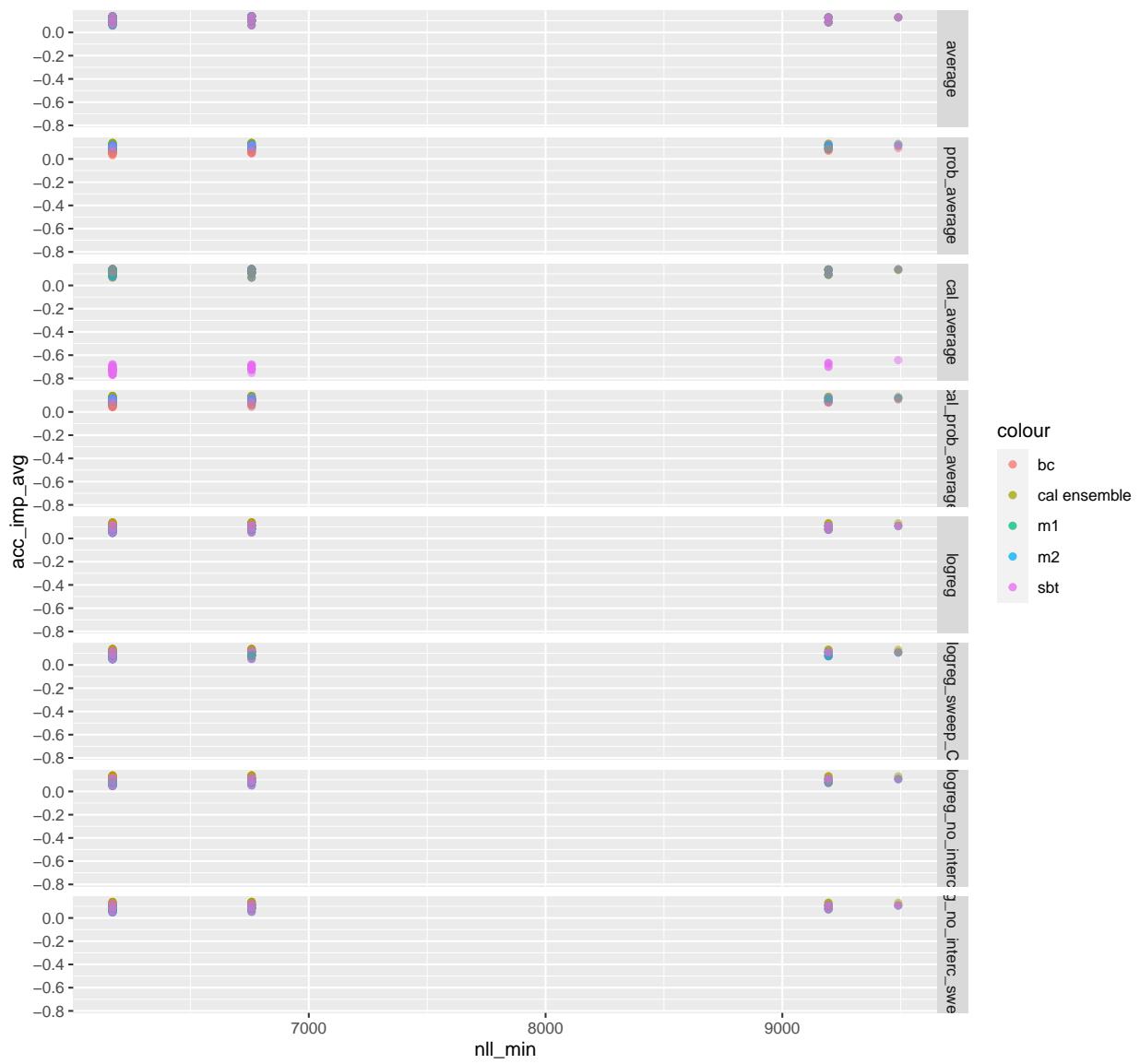
Accuracy improvement of ensemble over  
the average of networks vs nll\_var.  
Ensemble size 5



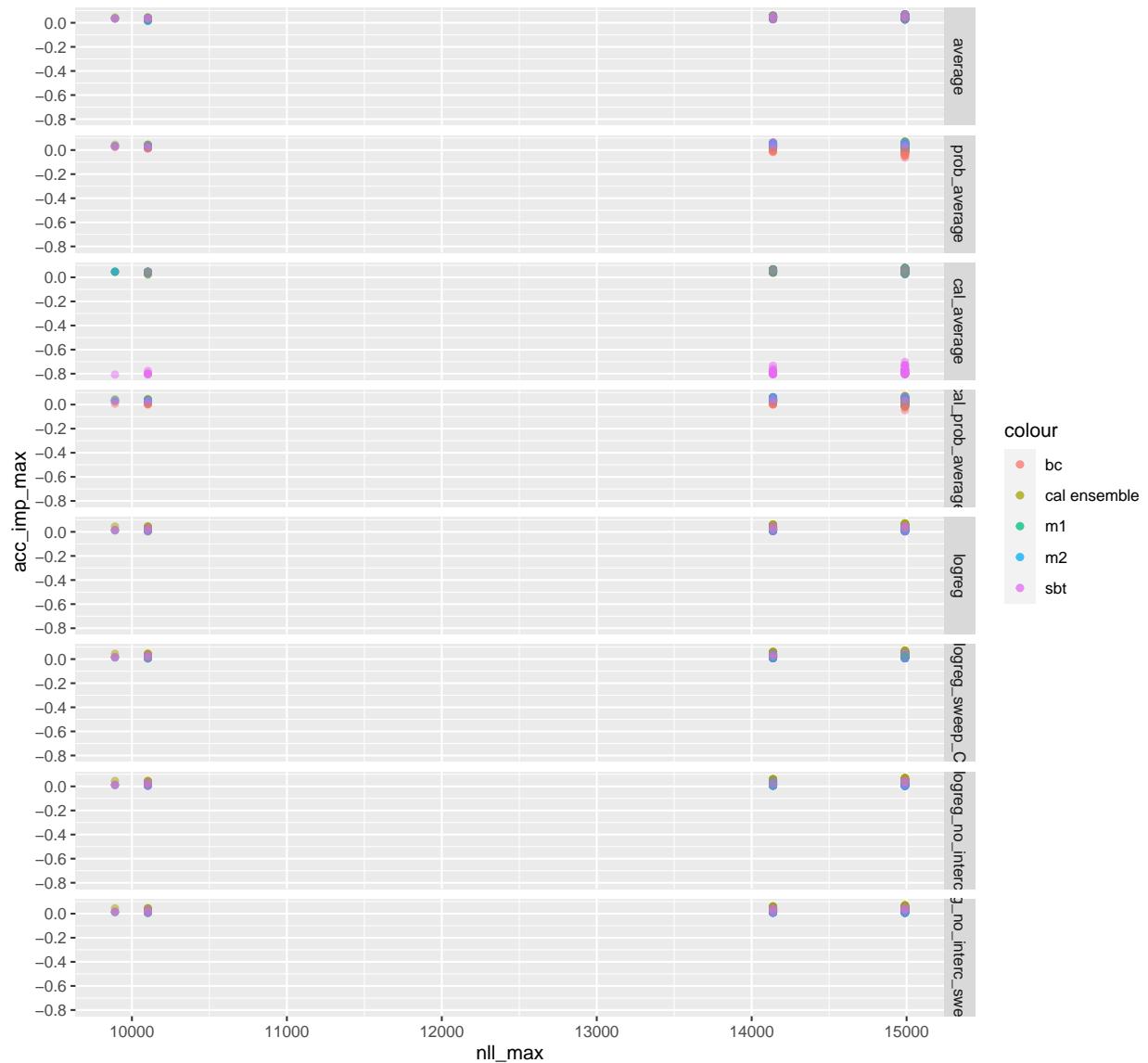
Accuracy improvement of ensemble over  
the best of networks vs nll\_min.  
Ensemble size 5



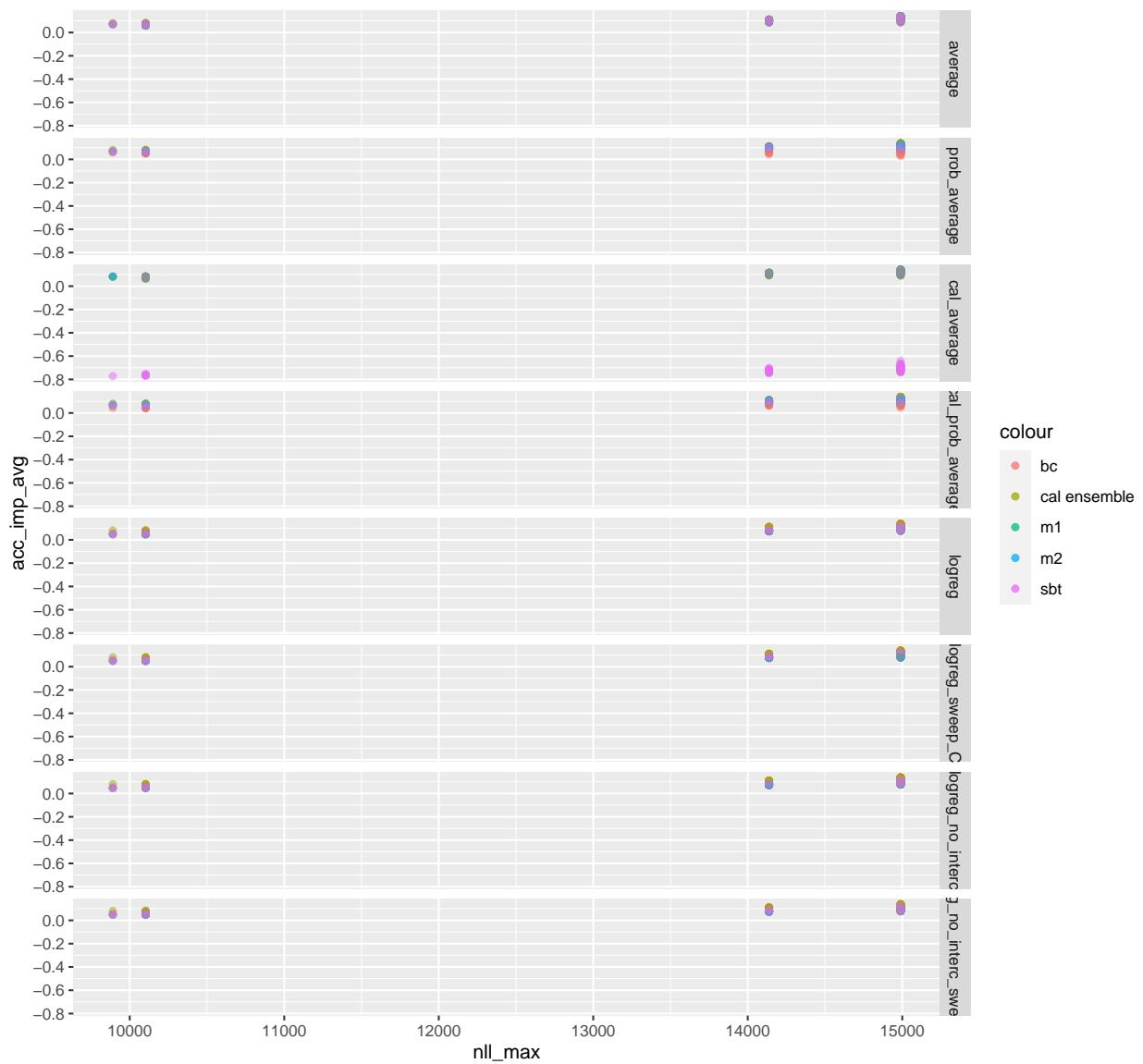
Accuracy improvement of ensemble over  
the average of networks vs nll\_min.  
Ensemble size 5



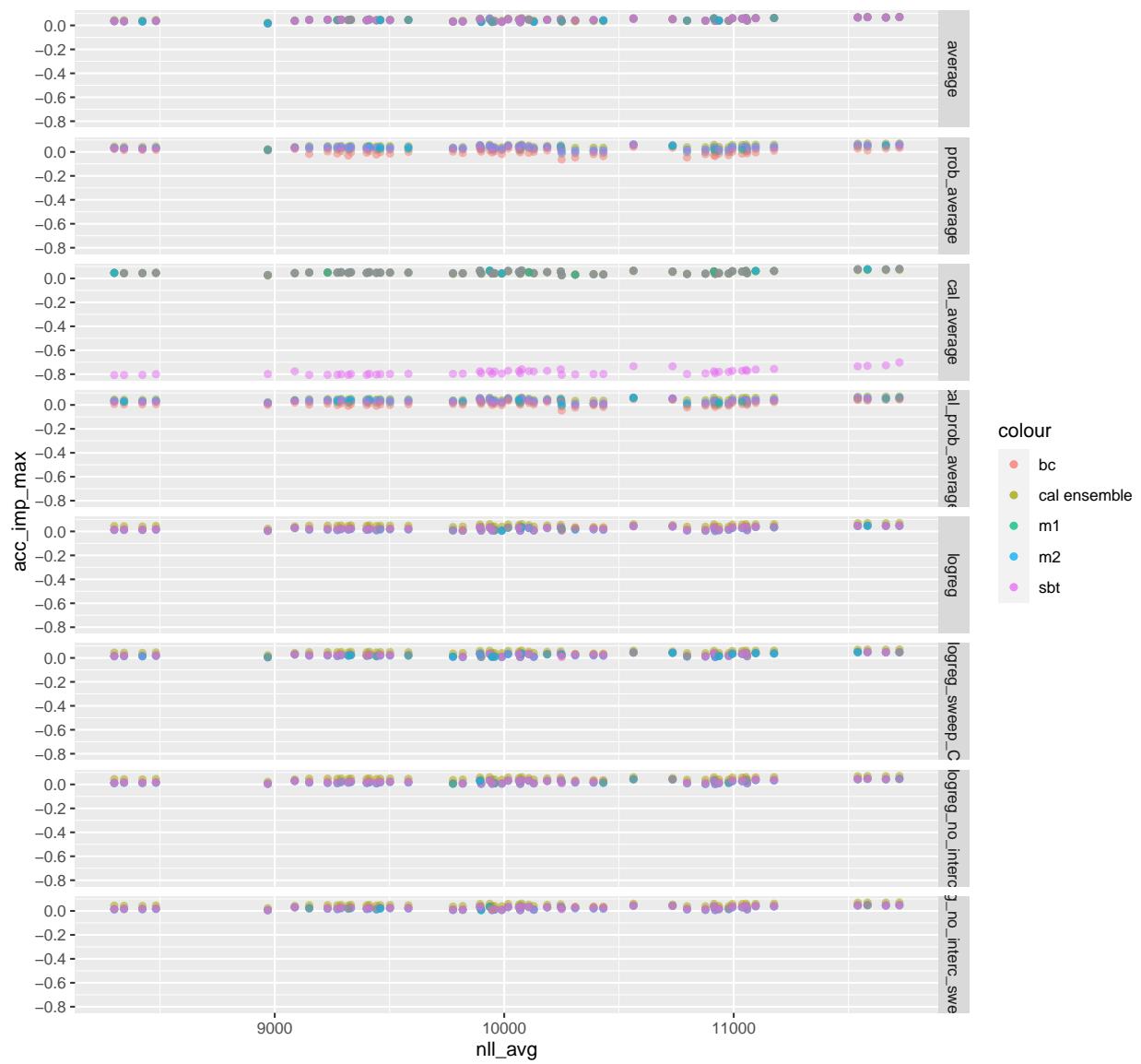
Accuracy improvement of ensemble over  
the best of networks vs nll\_max.  
Ensemble size 5



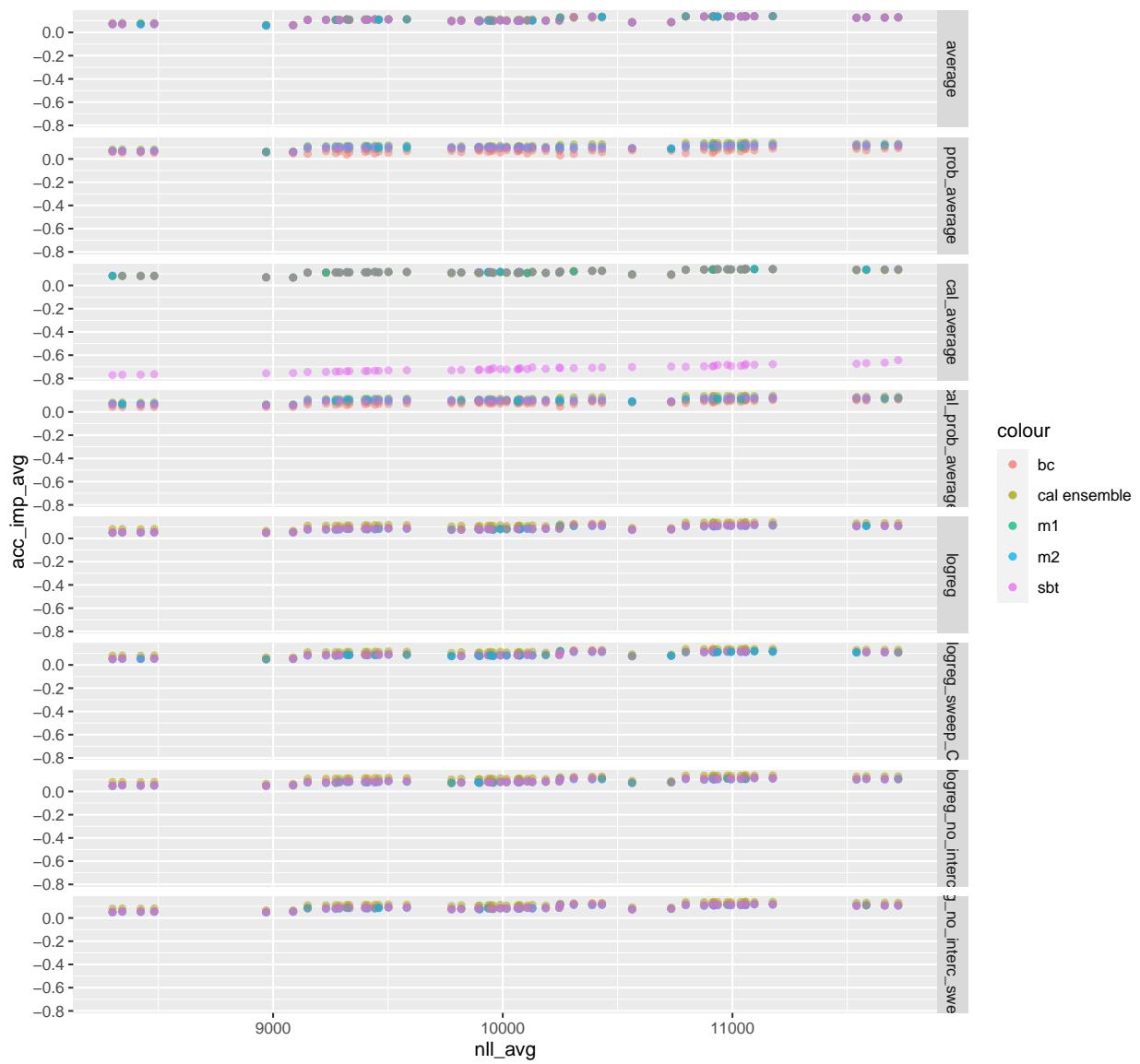
Accuracy improvement of ensemble over  
the average of networks vs nll\_max.  
Ensemble size 5



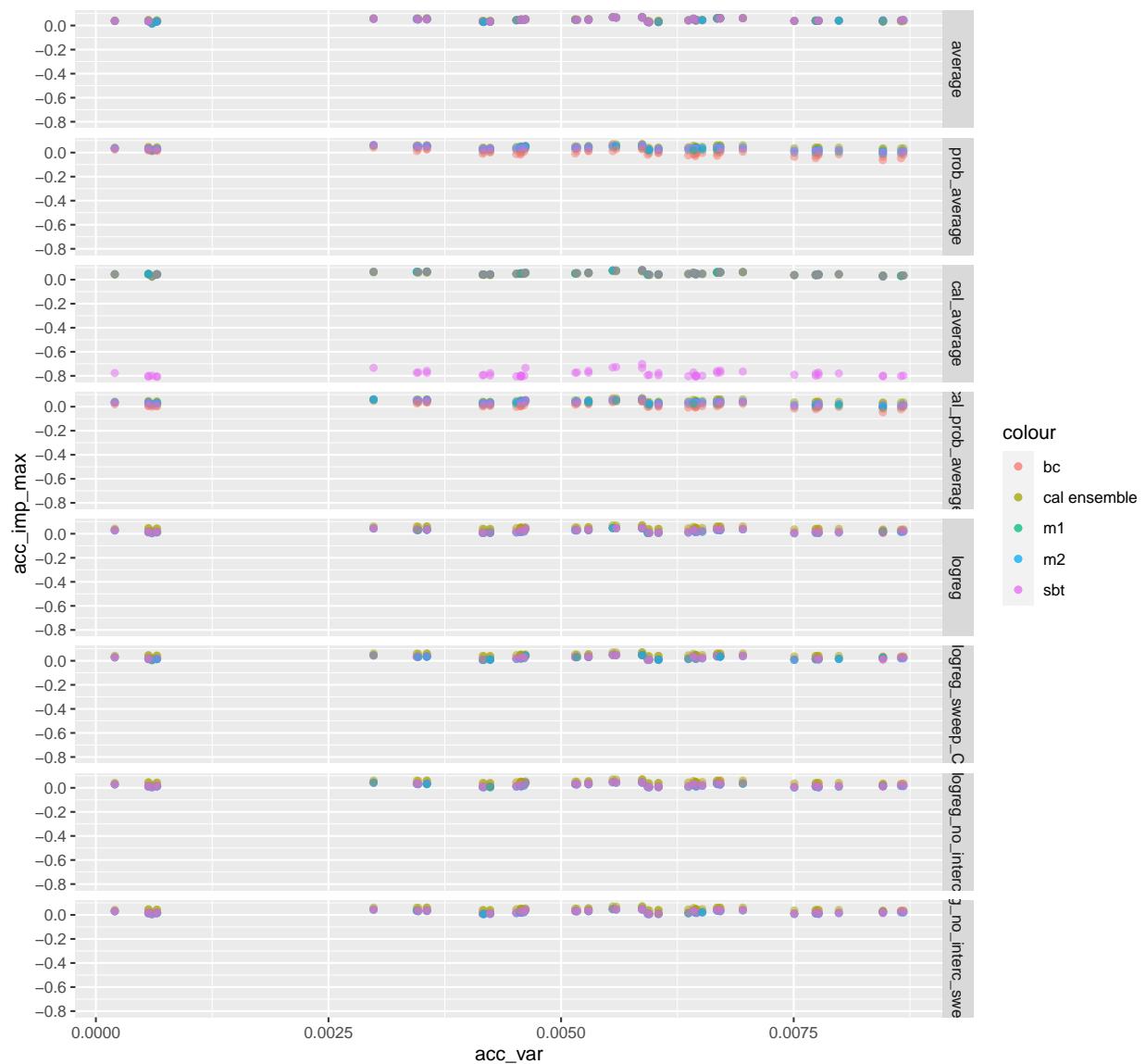
Accuracy improvement of ensemble over  
the best of networks vs nll\_avg.  
Ensemble size 5



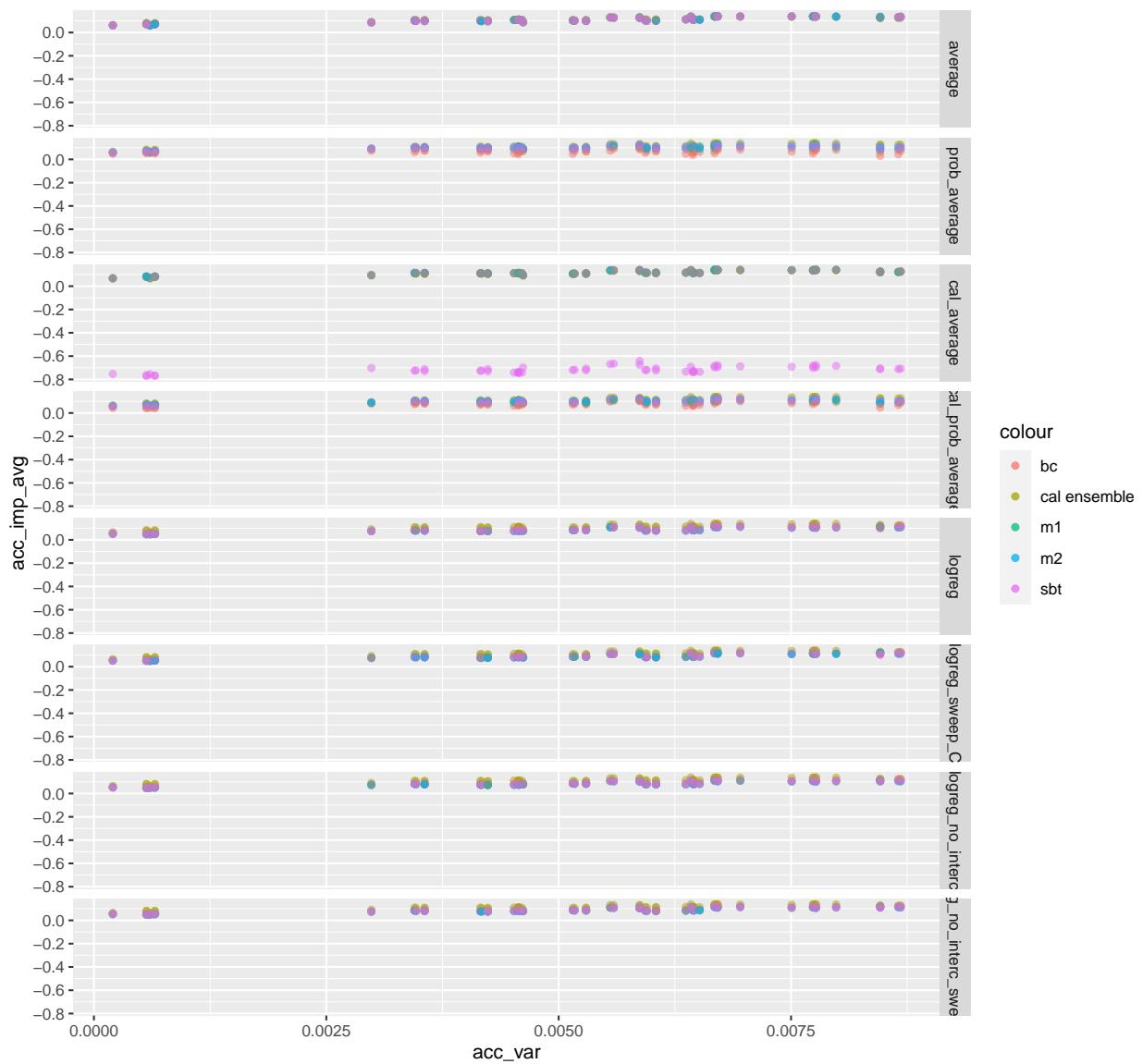
Accuracy improvement of ensemble over  
the average of networks vs nll\_avg.  
Ensemble size 5



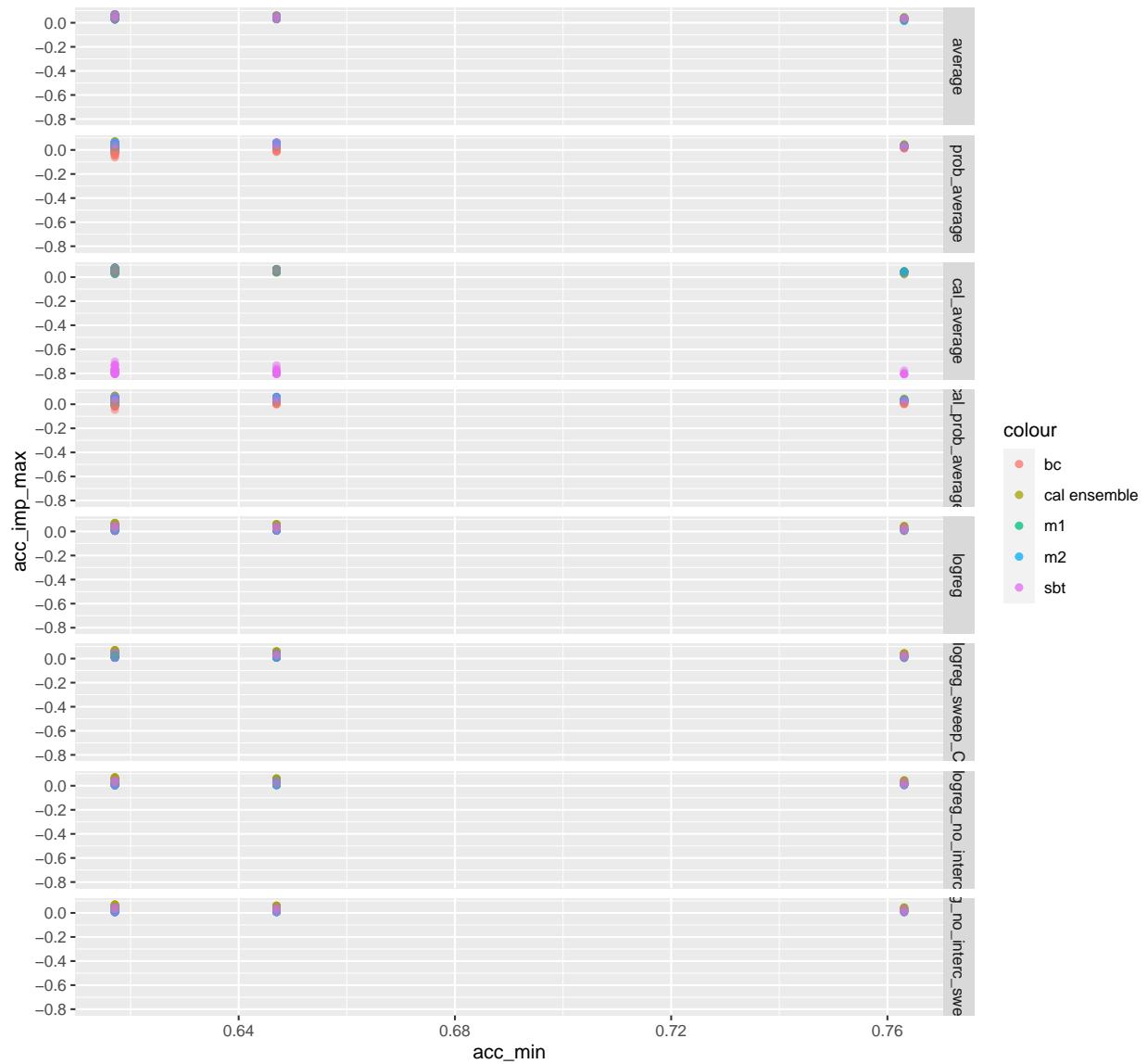
Accuracy improvement of ensemble over  
the best of networks vs acc\_var.  
Ensemble size 5



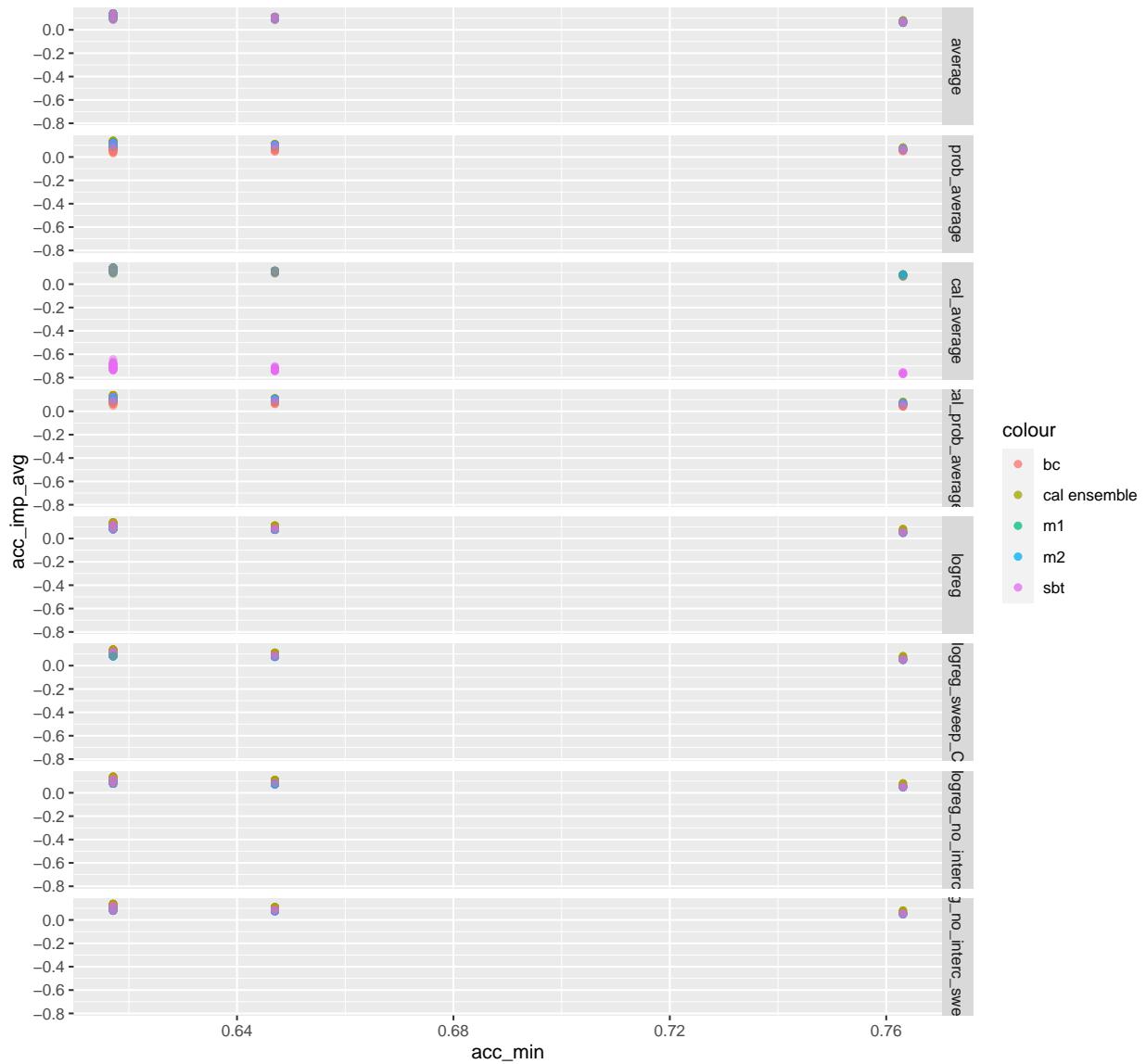
Accuracy improvement of ensemble over  
the average of networks vs acc\_var.  
Ensemble size 5



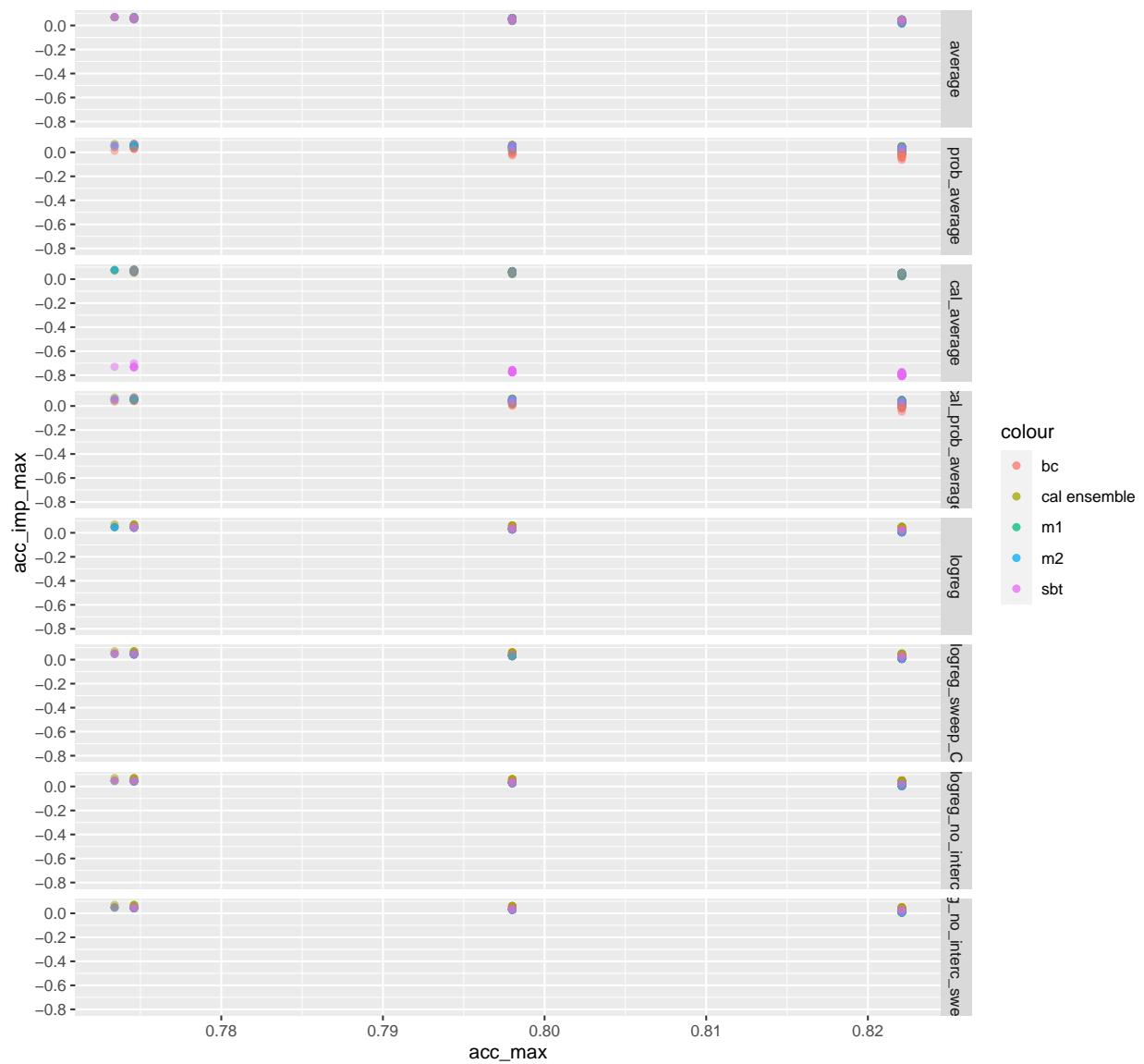
Accuracy improvement of ensemble over  
the best of networks vs acc\_min.  
Ensemble size 5



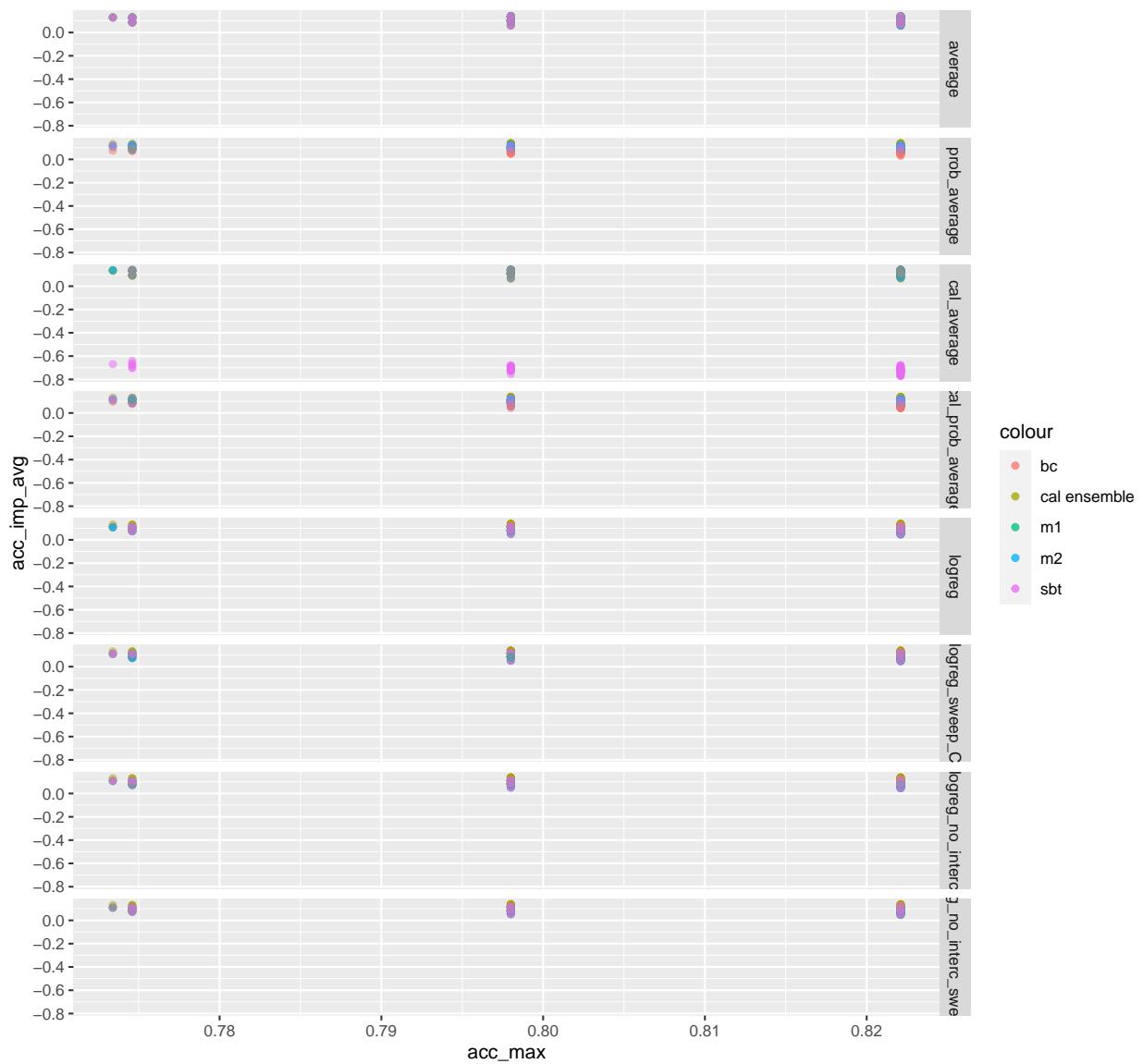
Accuracy improvement of ensemble over  
the average of networks vs acc\_min.  
Ensemble size 5



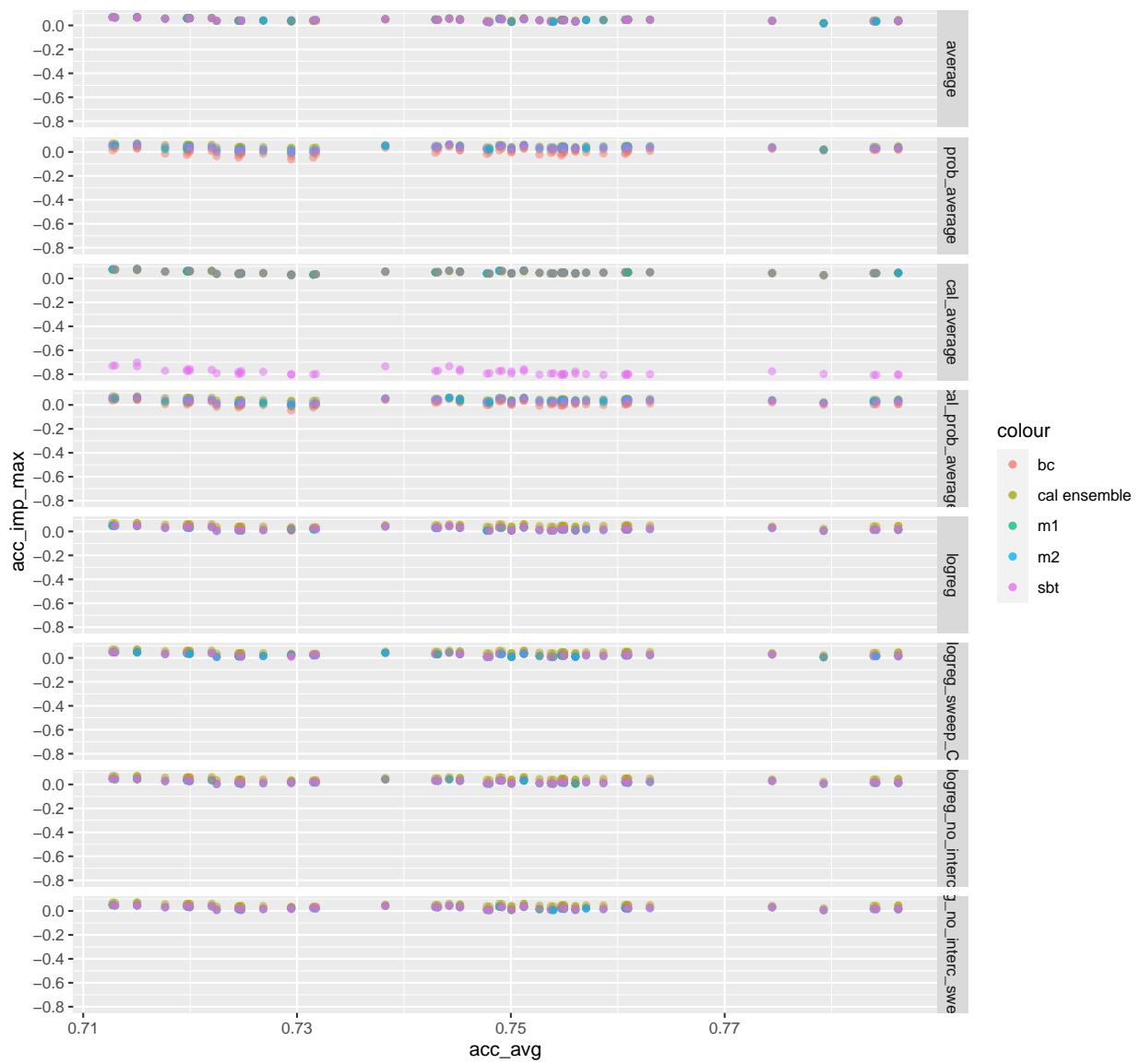
Accuracy improvement of ensemble over  
the best of networks vs acc\_max.  
Ensemble size 5



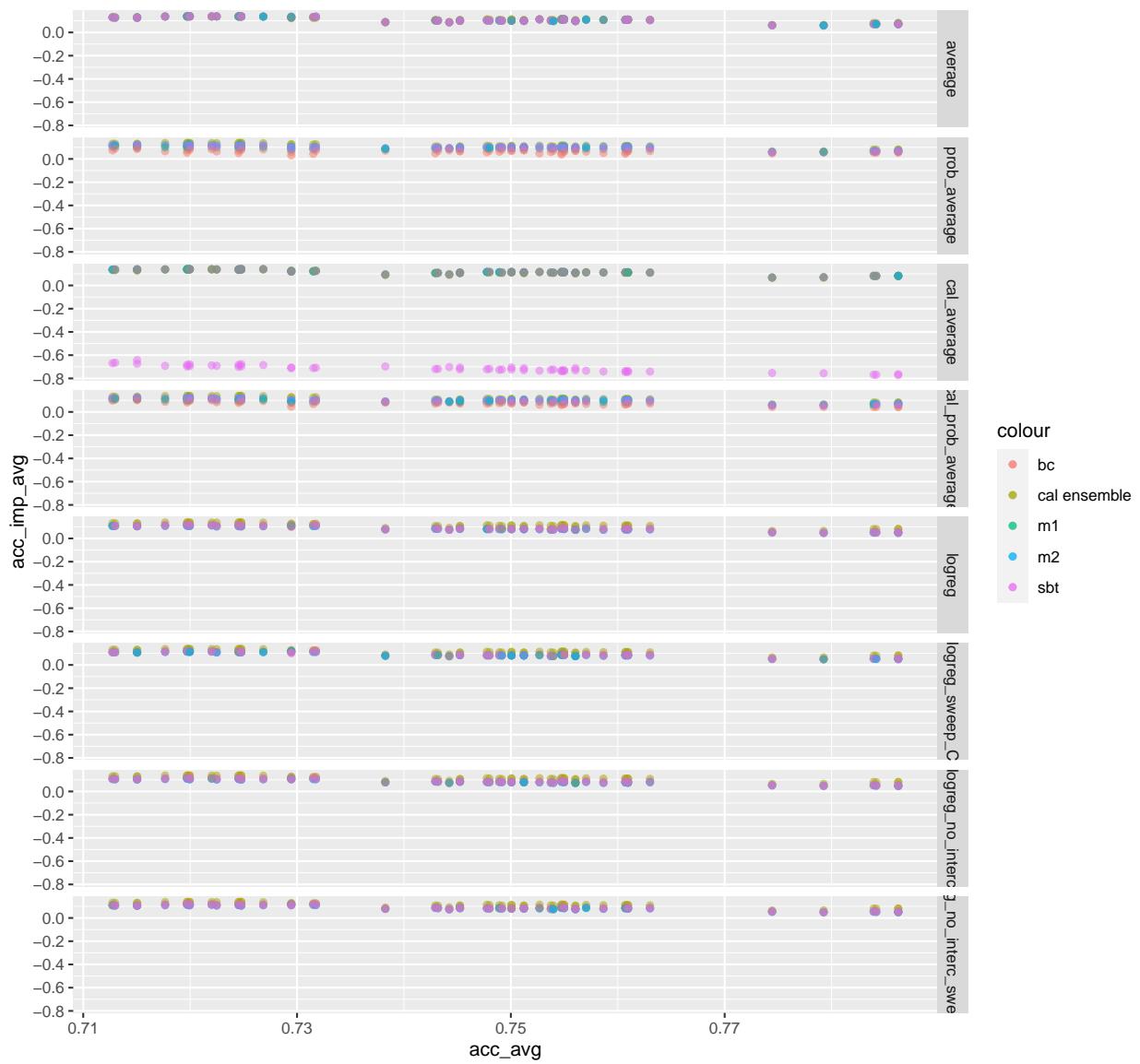
Accuracy improvement of ensemble over  
the average of networks vs acc\_max.  
Ensemble size 5



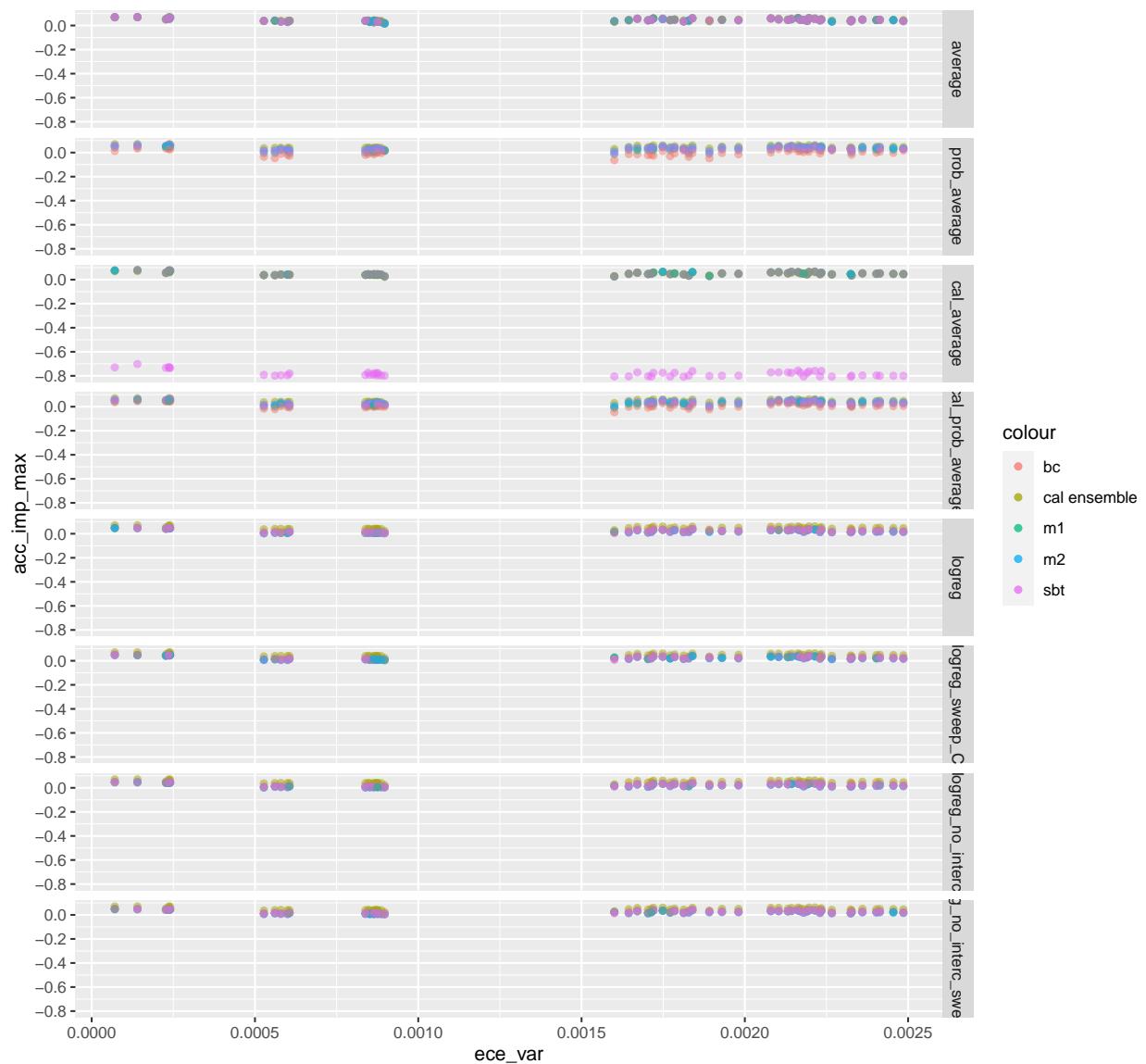
Accuracy improvement of ensemble over  
the best of networks vs acc\_avg.  
Ensemble size 5



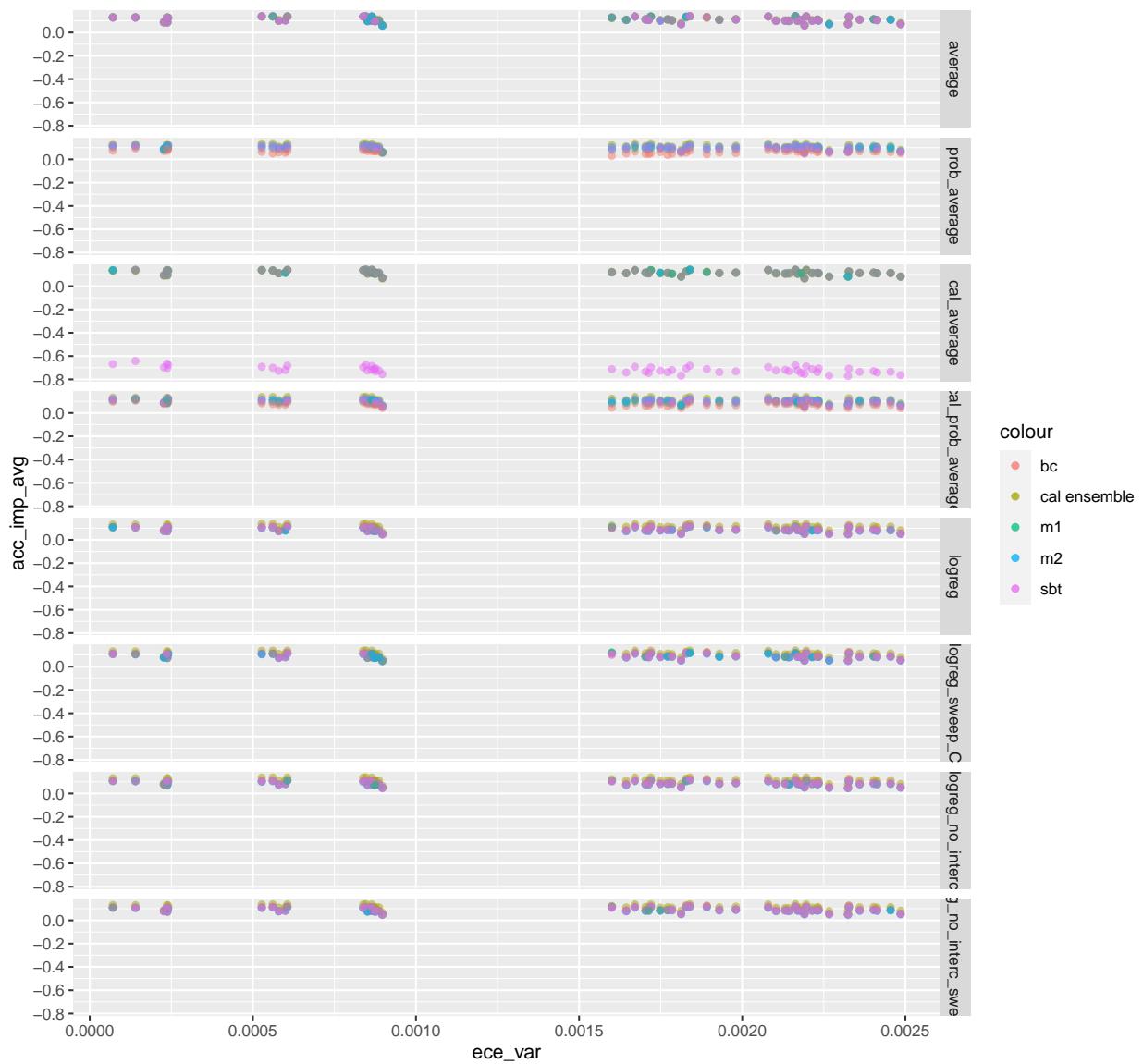
Accuracy improvement of ensemble over  
the average of networks vs acc\_avg.  
Ensemble size 5



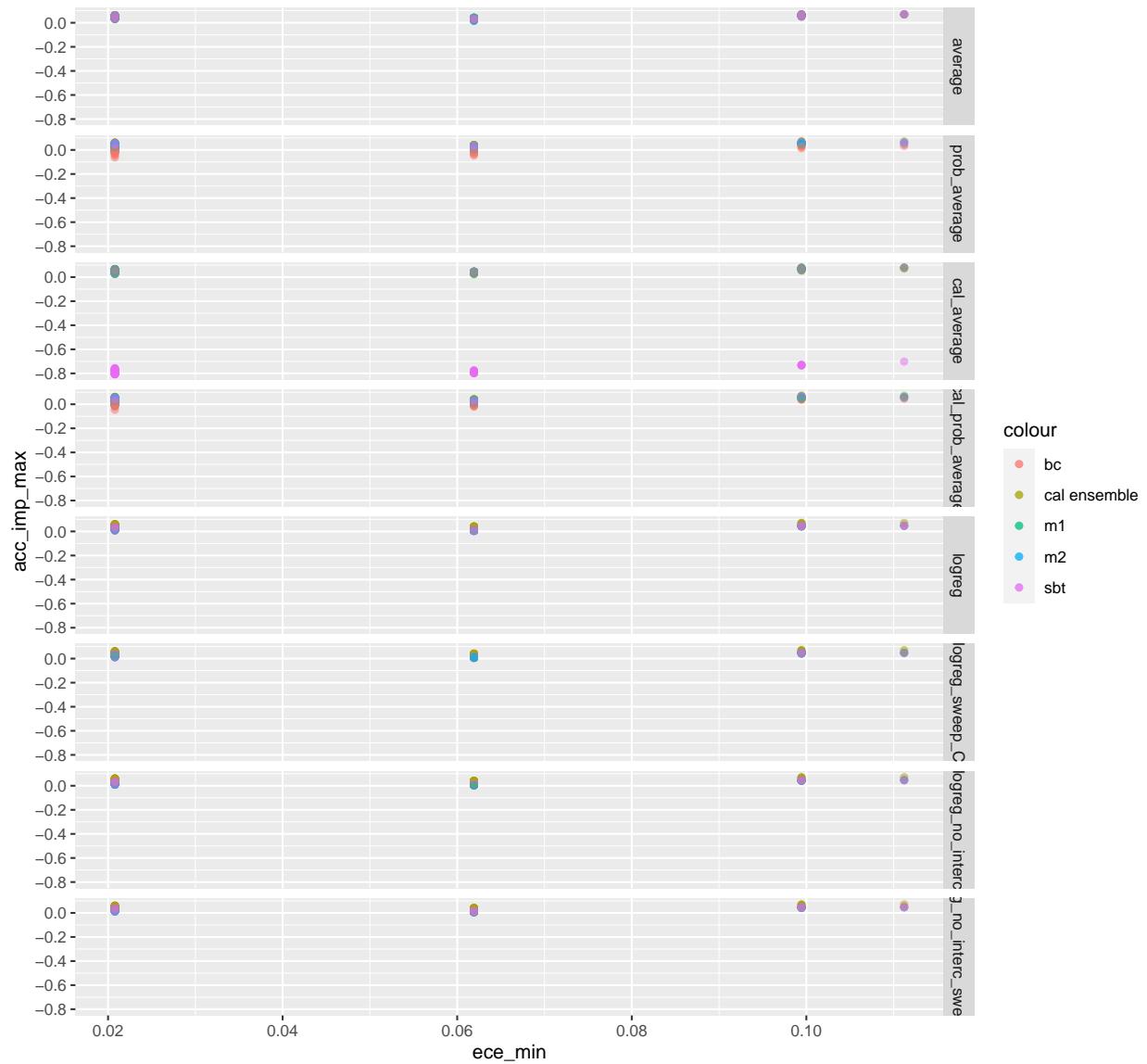
Accuracy improvement of ensemble over  
the best of networks vs ece\_var.  
Ensemble size 5



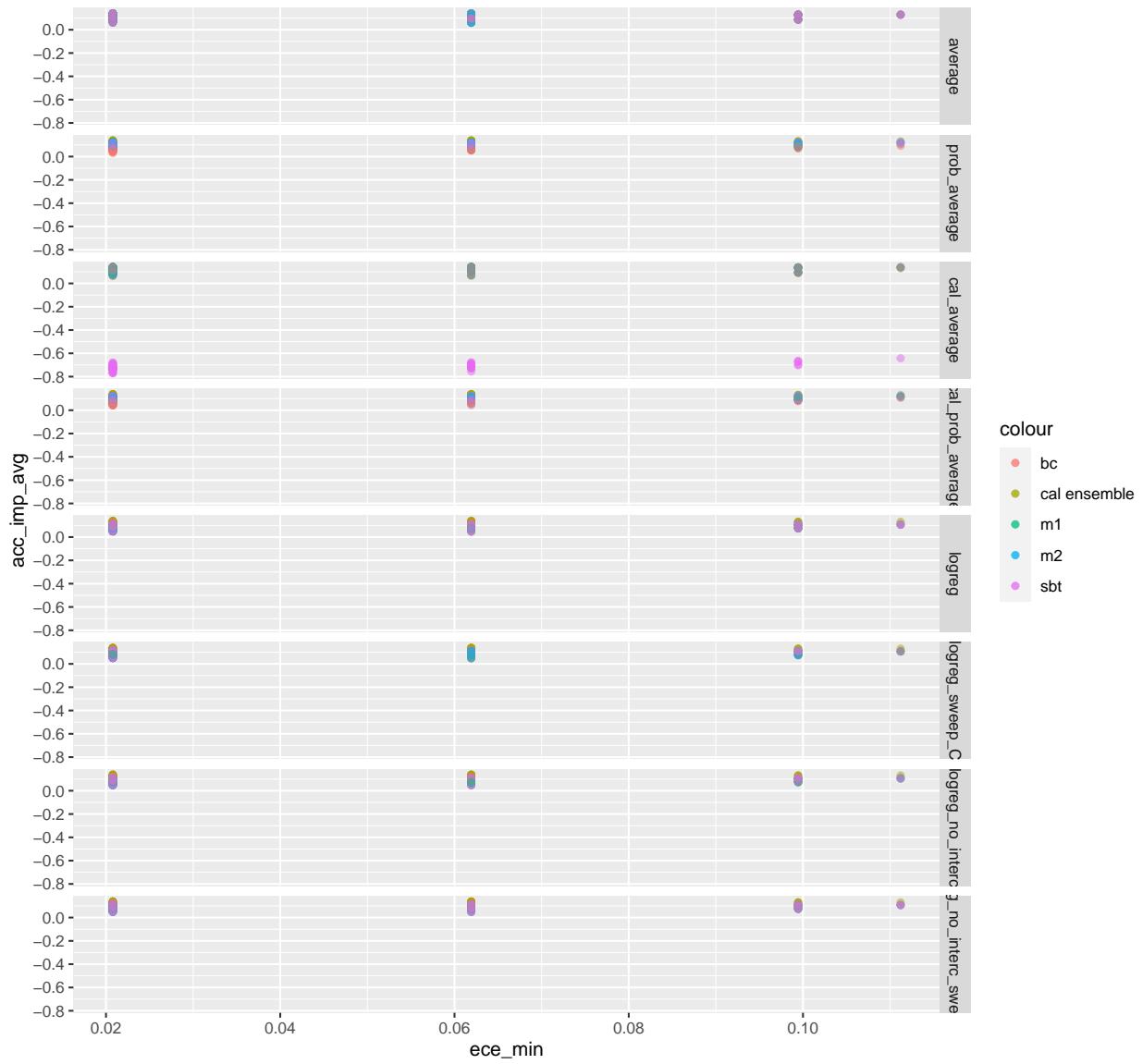
Accuracy improvement of ensemble over  
the average of networks vs ece\_var.  
Ensemble size 5



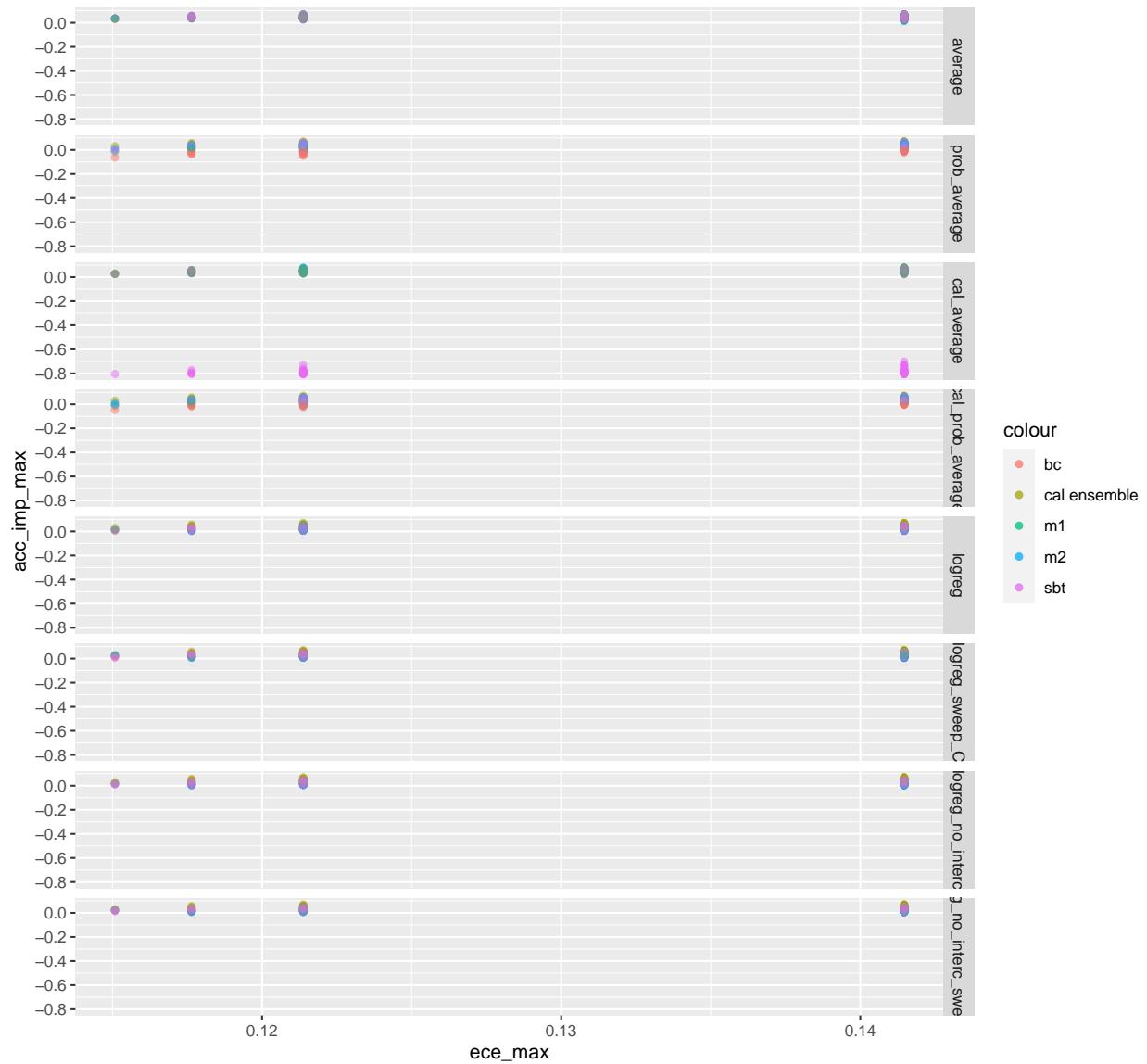
Accuracy improvement of ensemble over  
the best of networks vs ece\_min.  
Ensemble size 5



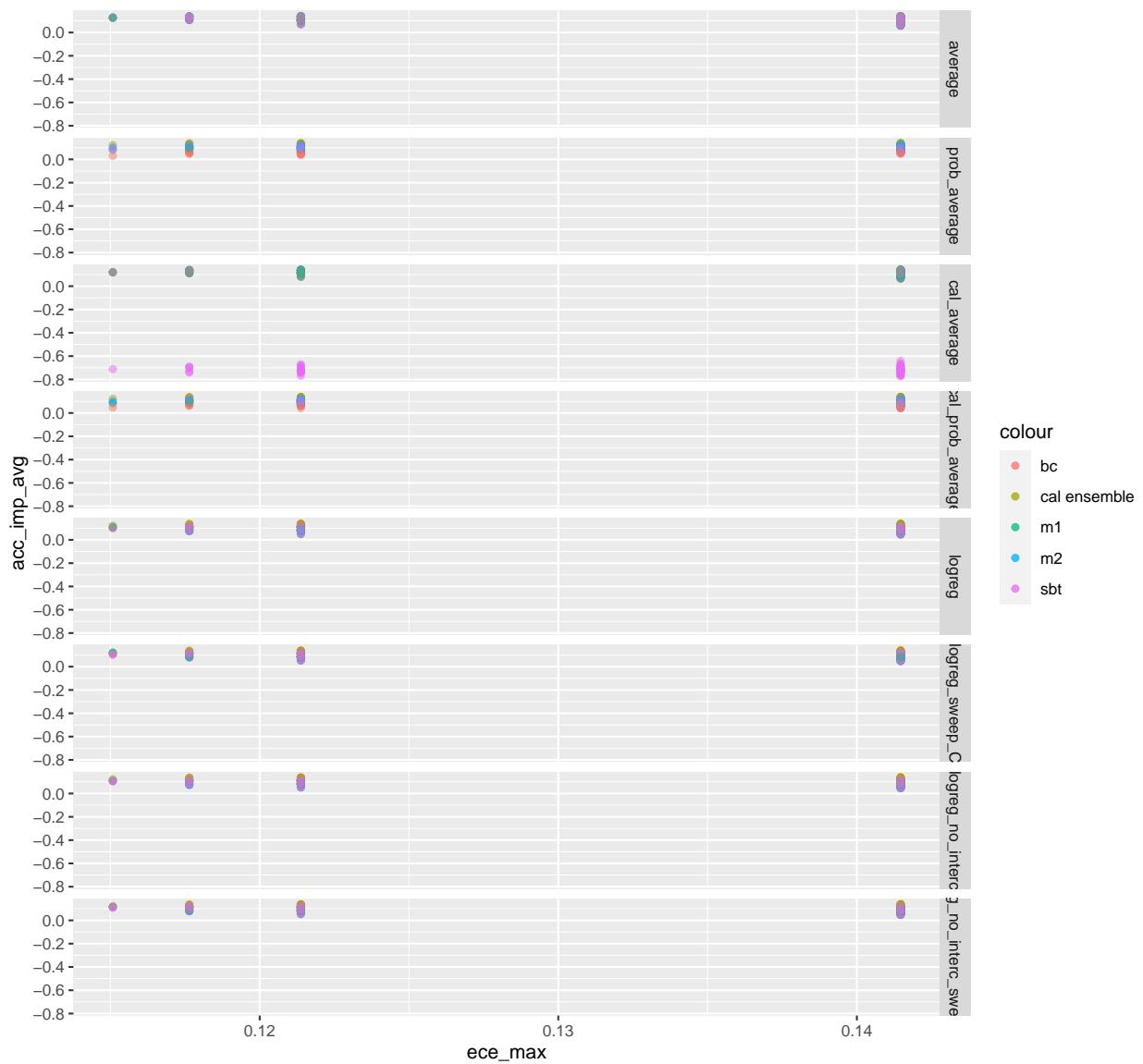
Accuracy improvement of ensemble over  
the average of networks vs ece\_min.  
Ensemble size 5



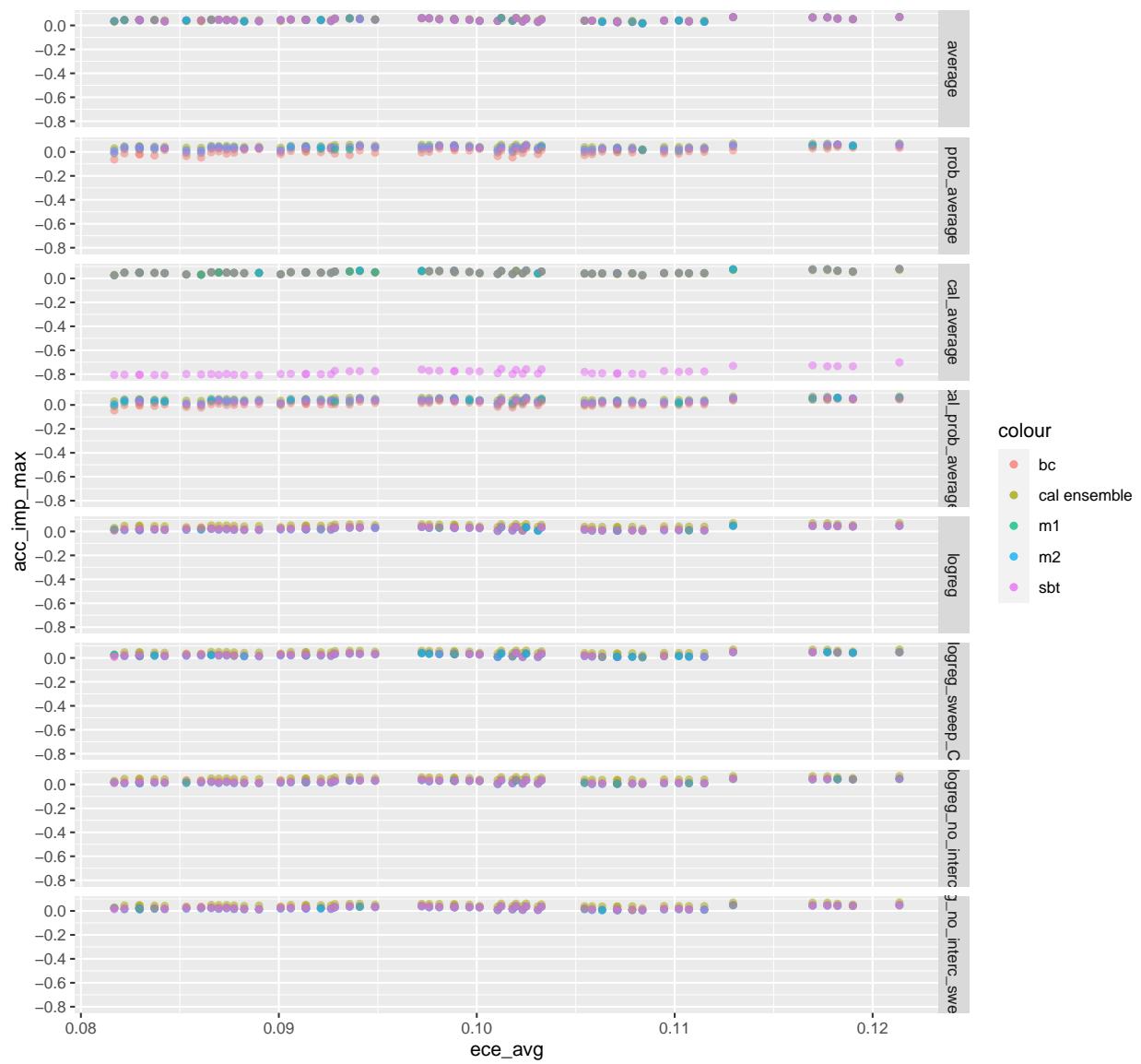
Accuracy improvement of ensemble over  
the best of networks vs ece\_max.  
Ensemble size 5



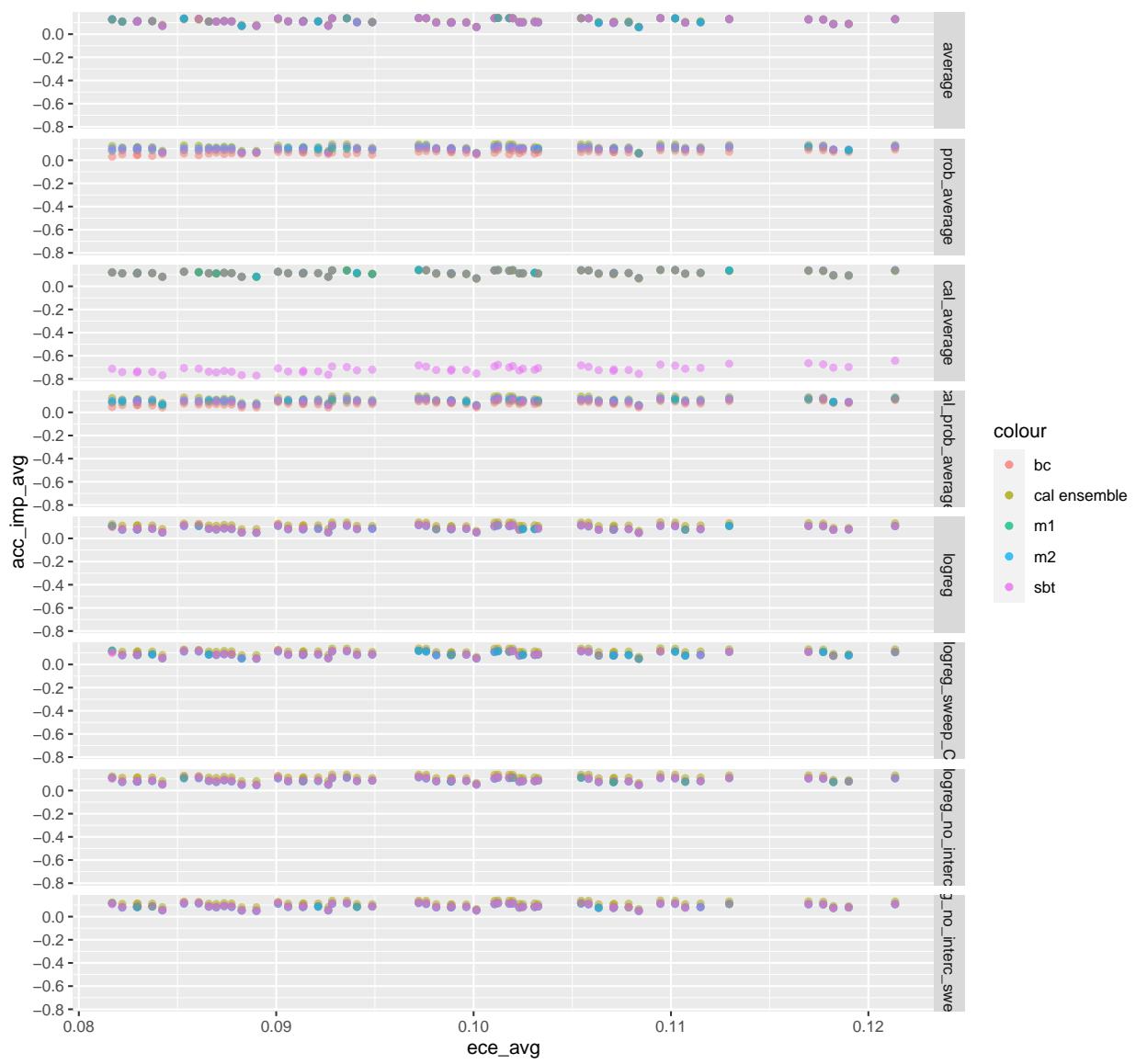
Accuracy improvement of ensemble over  
the average of networks vs ece\_max.  
Ensemble size 5



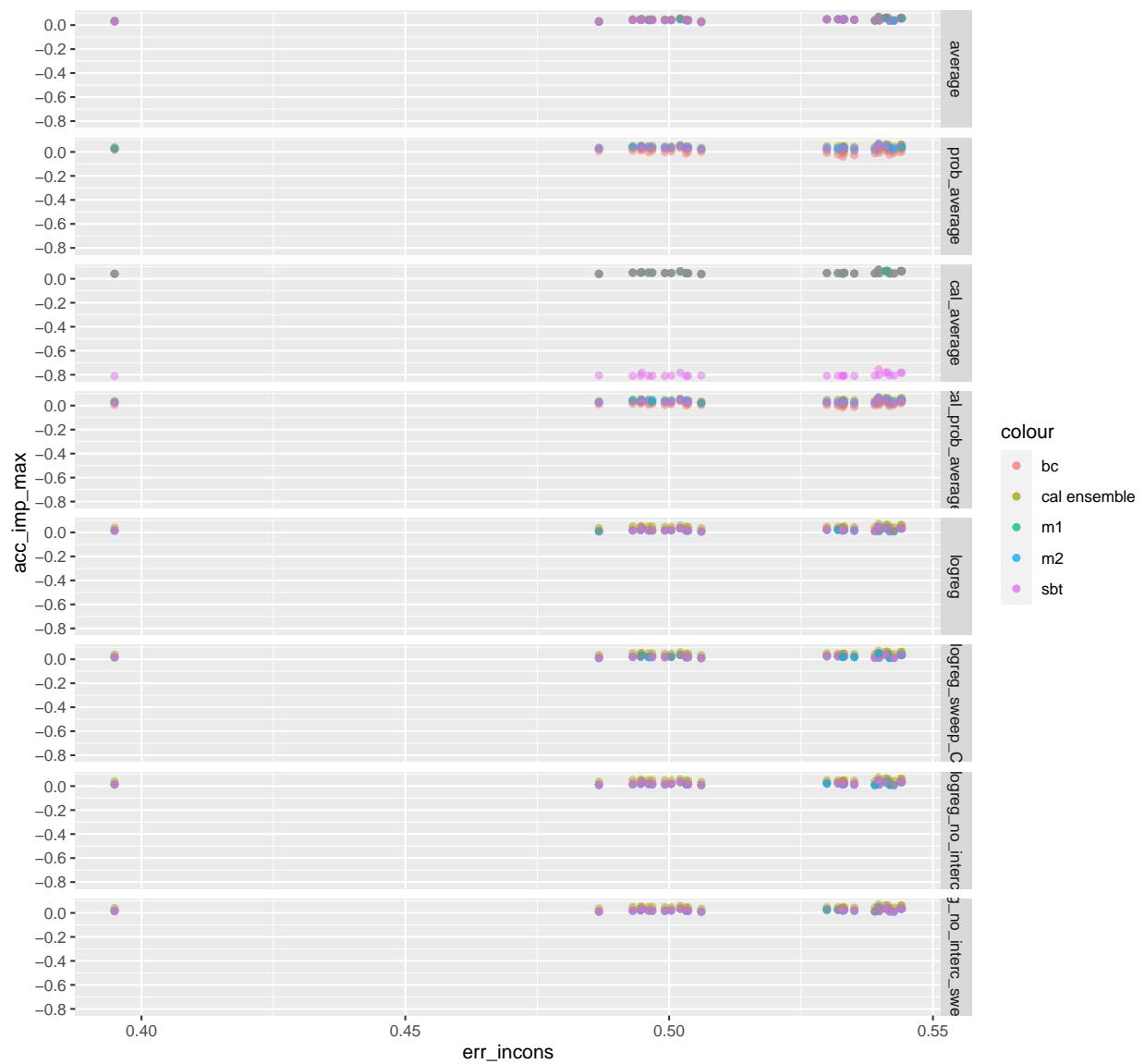
Accuracy improvement of ensemble over  
the best of networks vs ece\_avg.  
Ensemble size 5



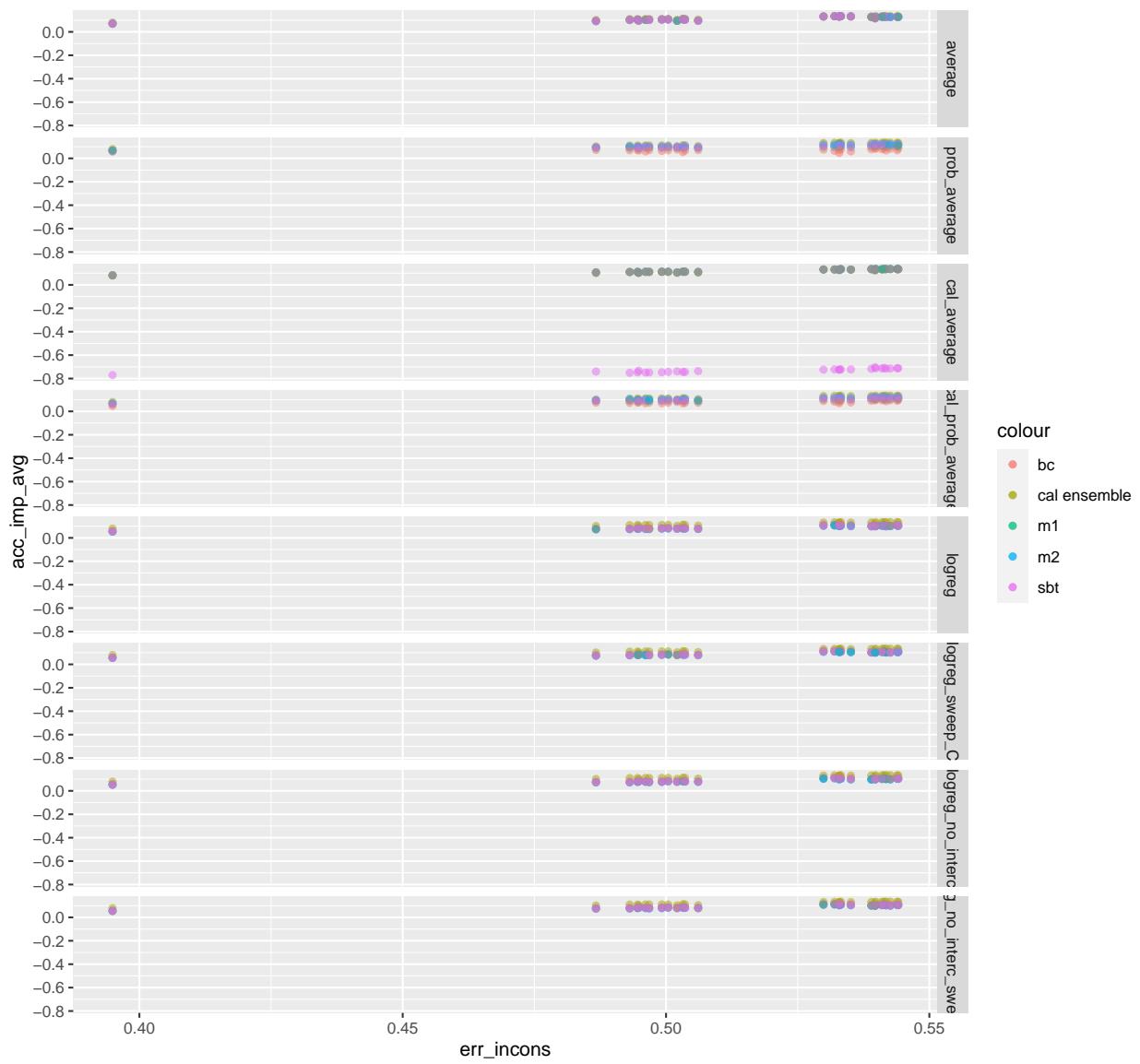
Accuracy improvement of ensemble over  
the average of networks vs ece\_avg.  
Ensemble size 5



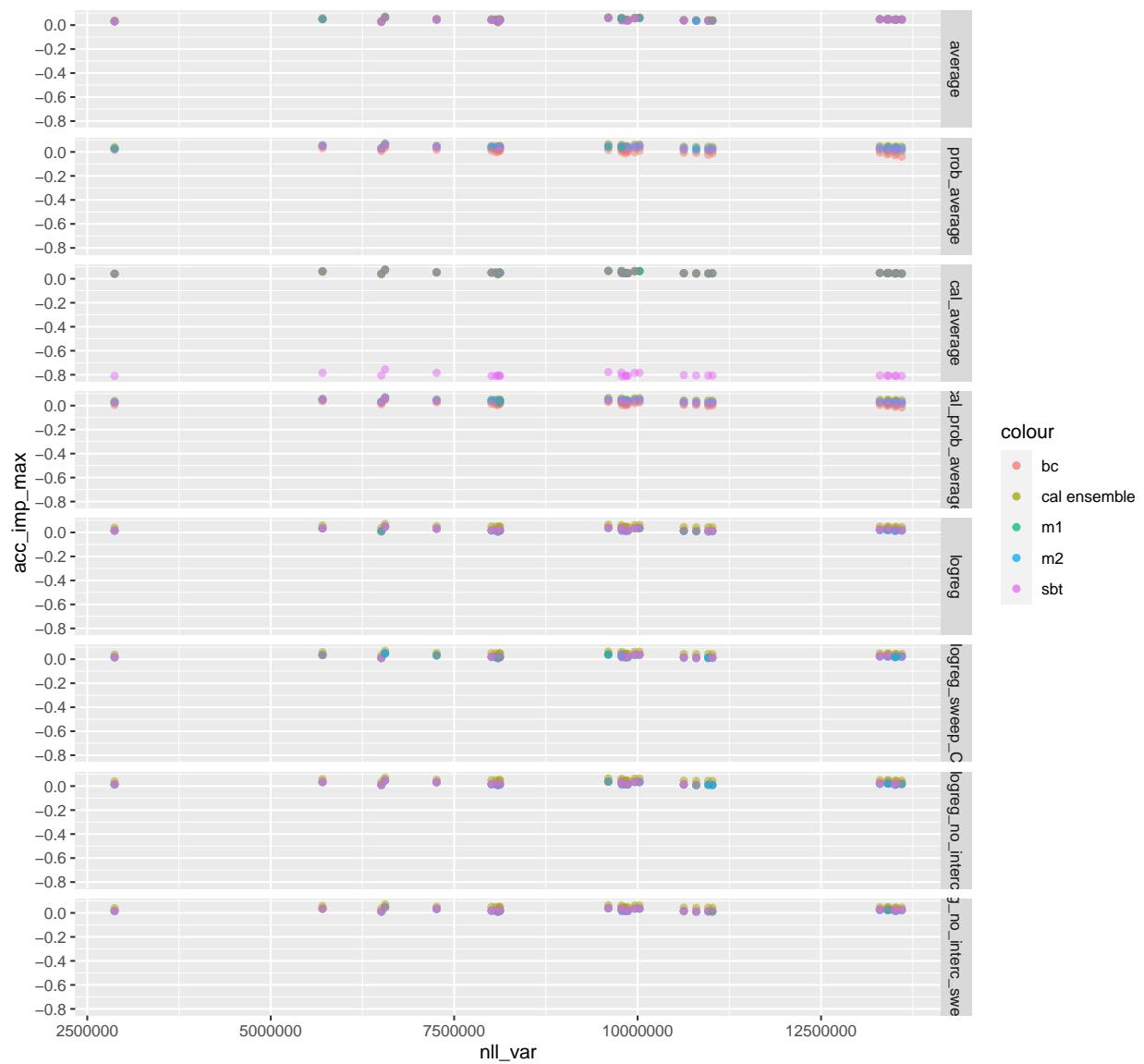
Accuracy improvement of ensemble over  
the best of networks vs err\_incons.  
Ensemble size 6



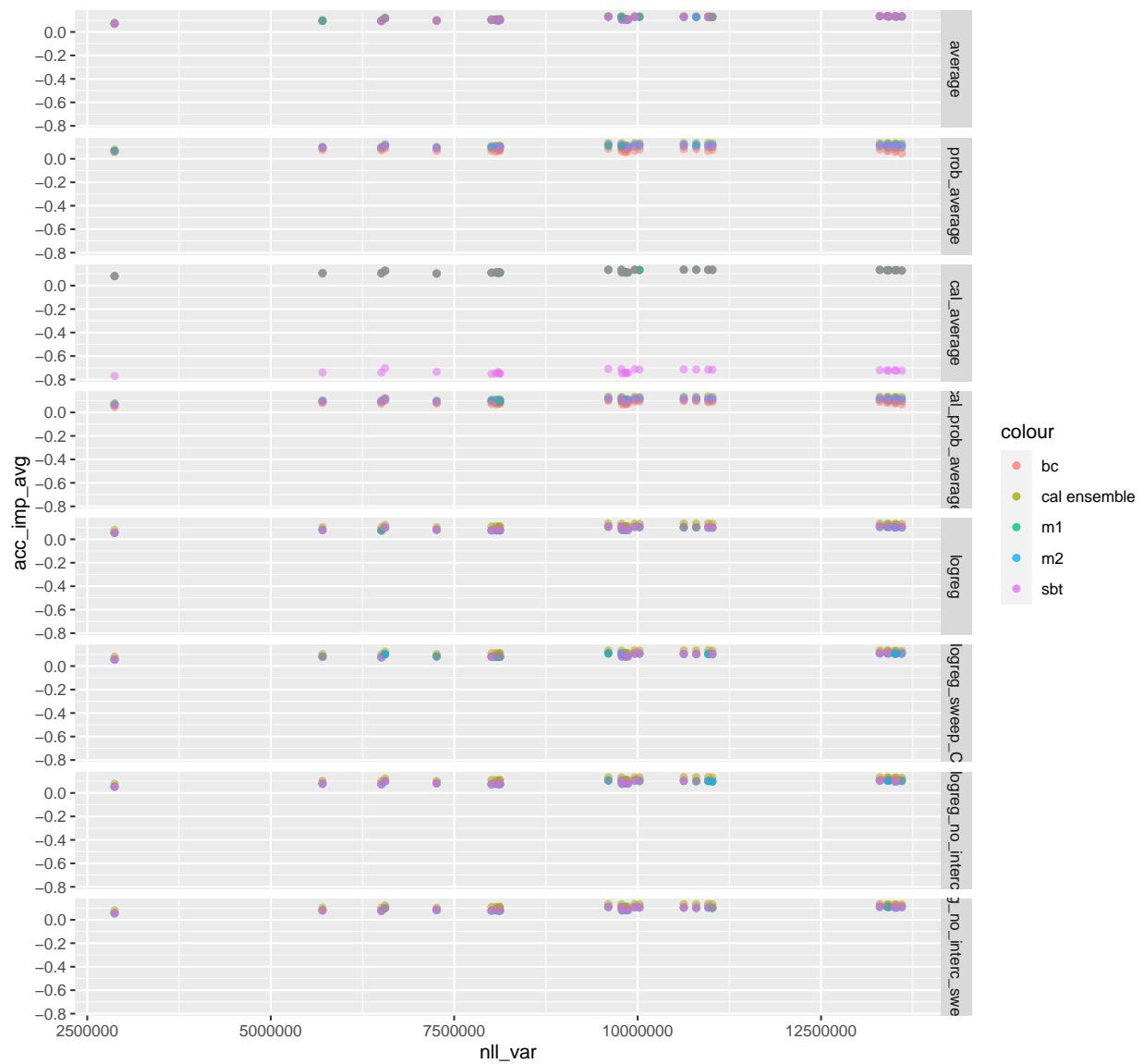
Accuracy improvement of ensemble over  
the average of networks vs err\_incons.  
Ensemble size 6



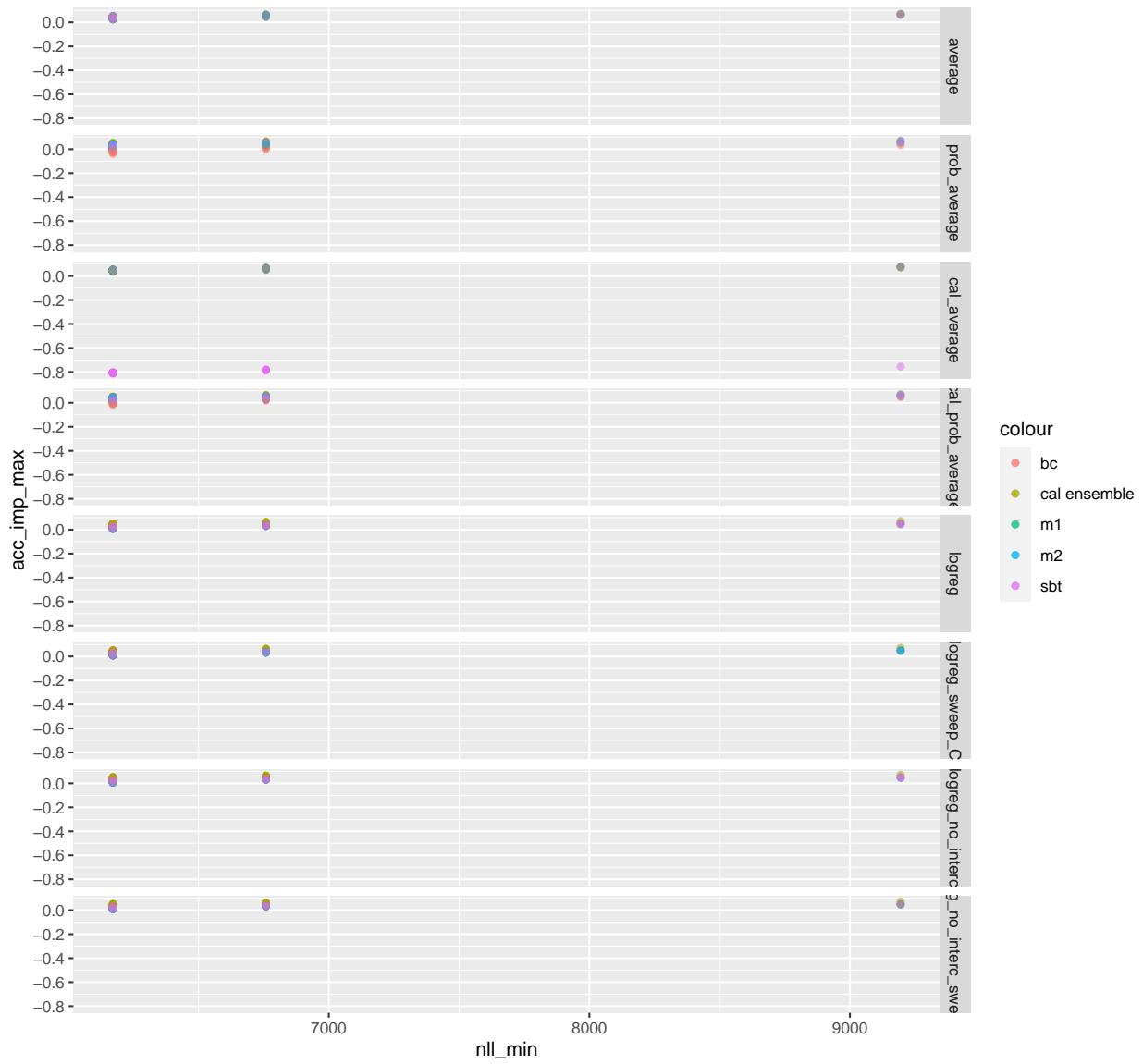
Accuracy improvement of ensemble over  
the best of networks vs nll\_var.  
Ensemble size 6



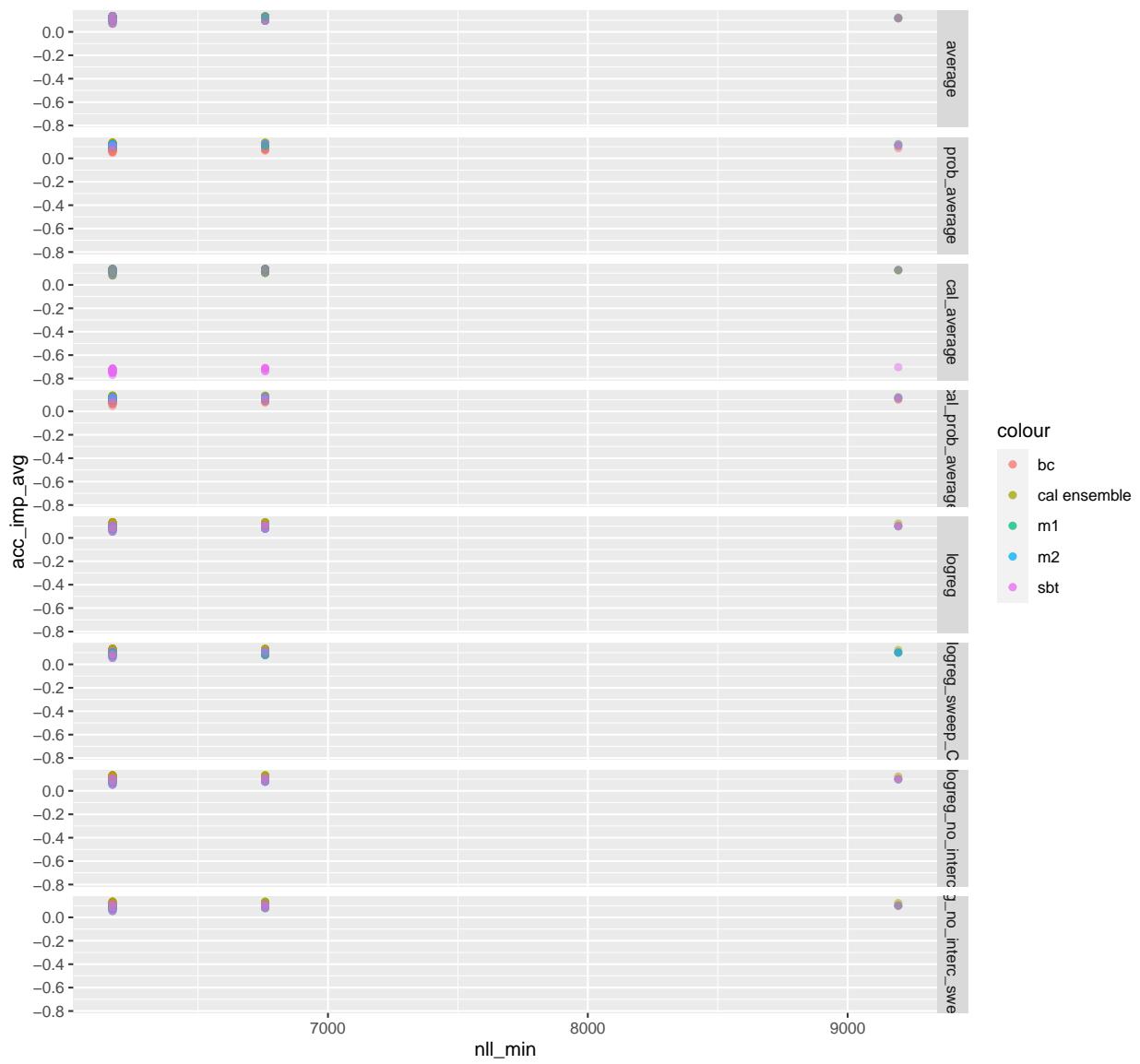
Accuracy improvement of ensemble over  
the average of networks vs nll\_var.  
Ensemble size 6



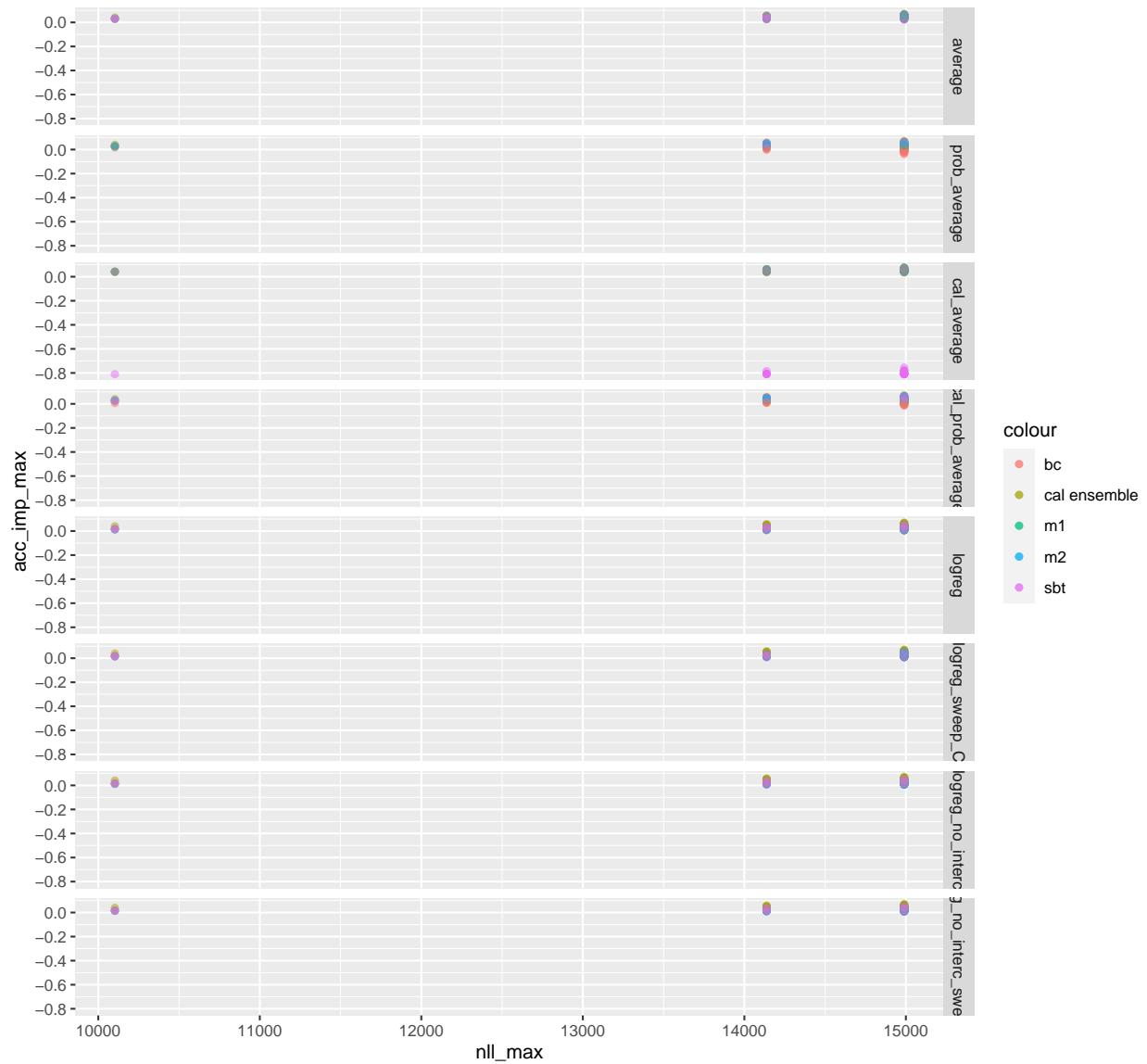
Accuracy improvement of ensemble over  
the best of networks vs nll\_min.  
Ensemble size 6



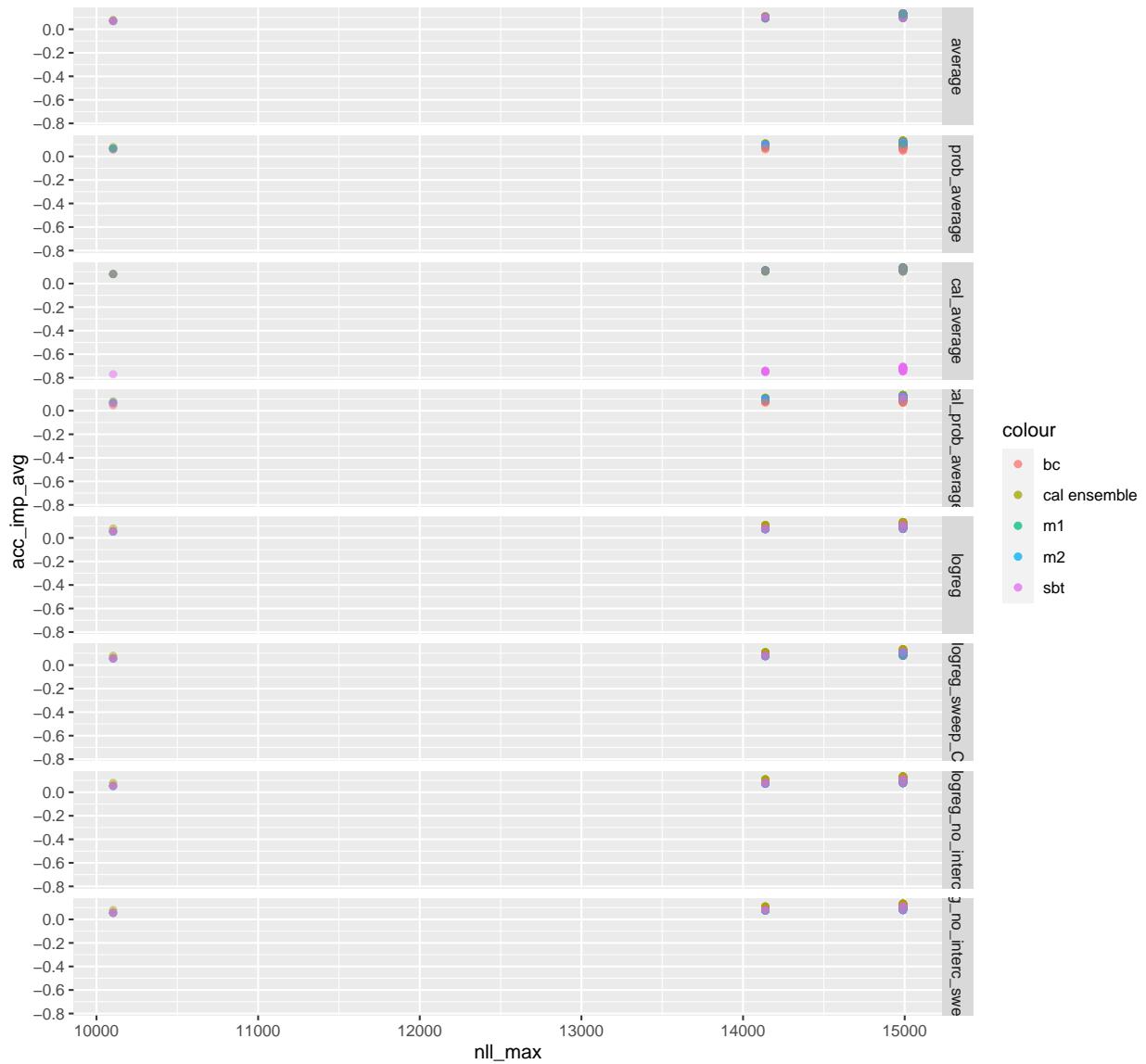
Accuracy improvement of ensemble over  
the average of networks vs nll\_min.  
Ensemble size 6



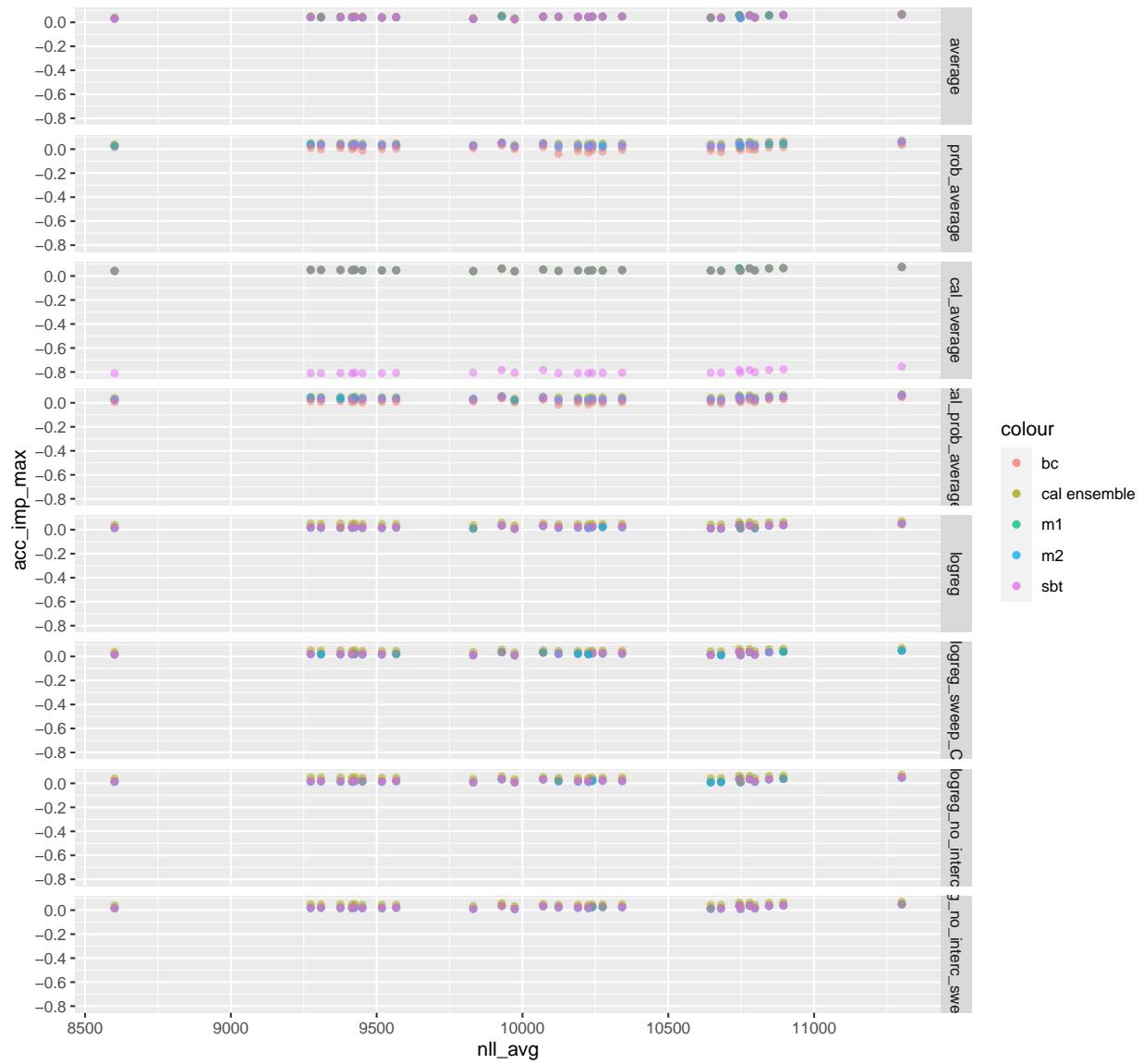
Accuracy improvement of ensemble over  
the best of networks vs nll\_max.  
Ensemble size 6



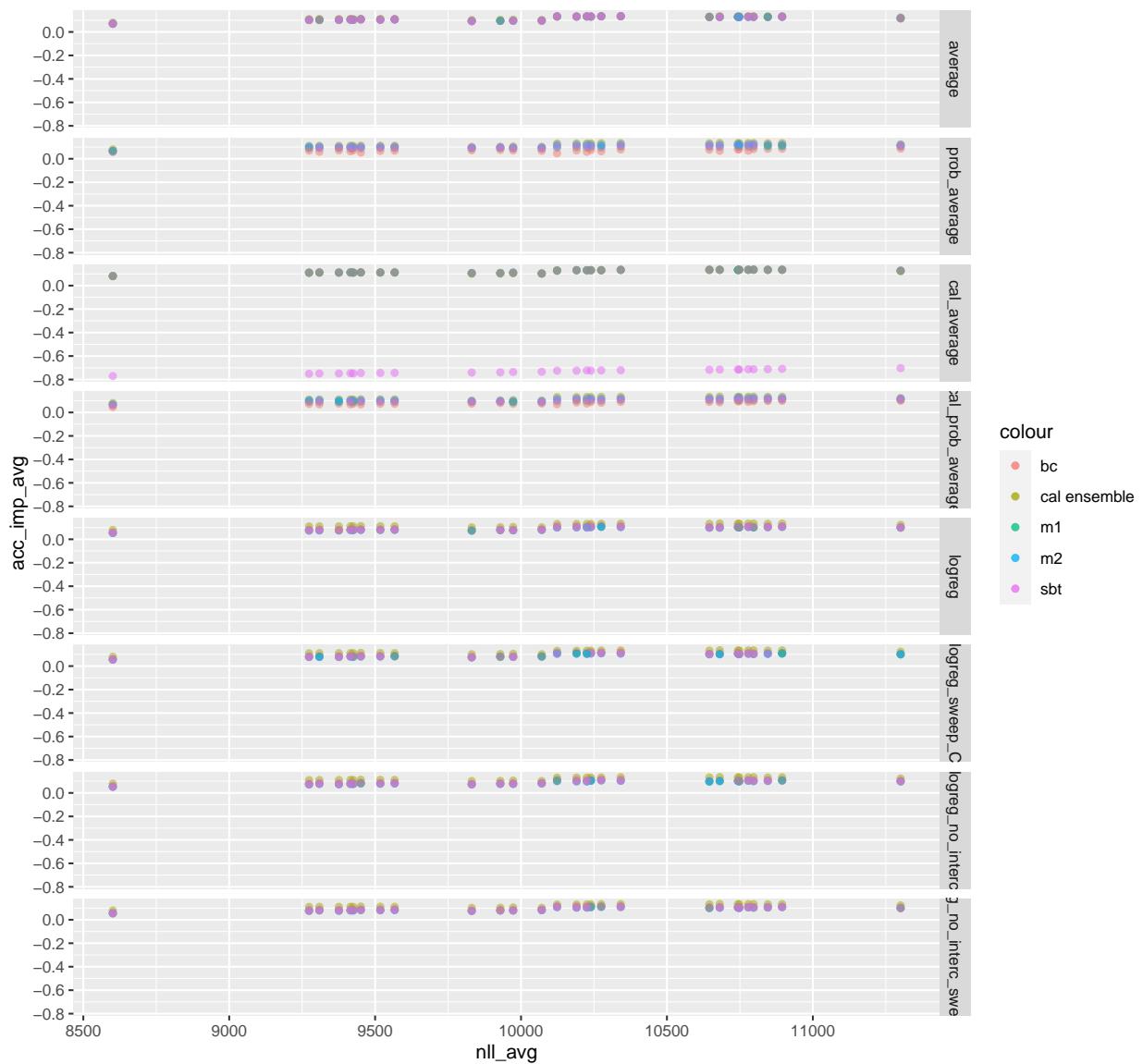
Accuracy improvement of ensemble over  
the average of networks vs nll\_max.  
Ensemble size 6



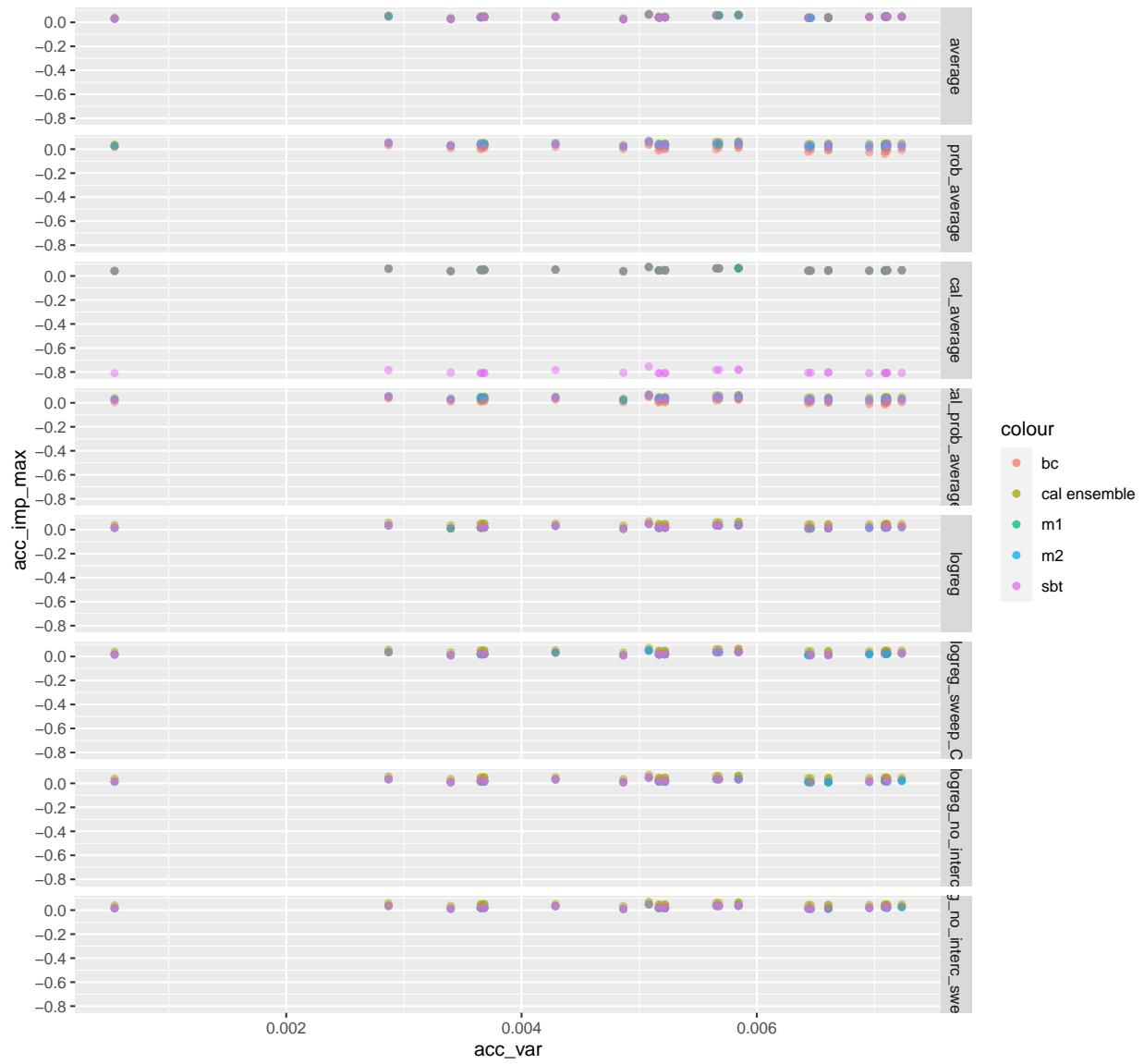
Accuracy improvement of ensemble over  
the best of networks vs nll\_avg.  
Ensemble size 6



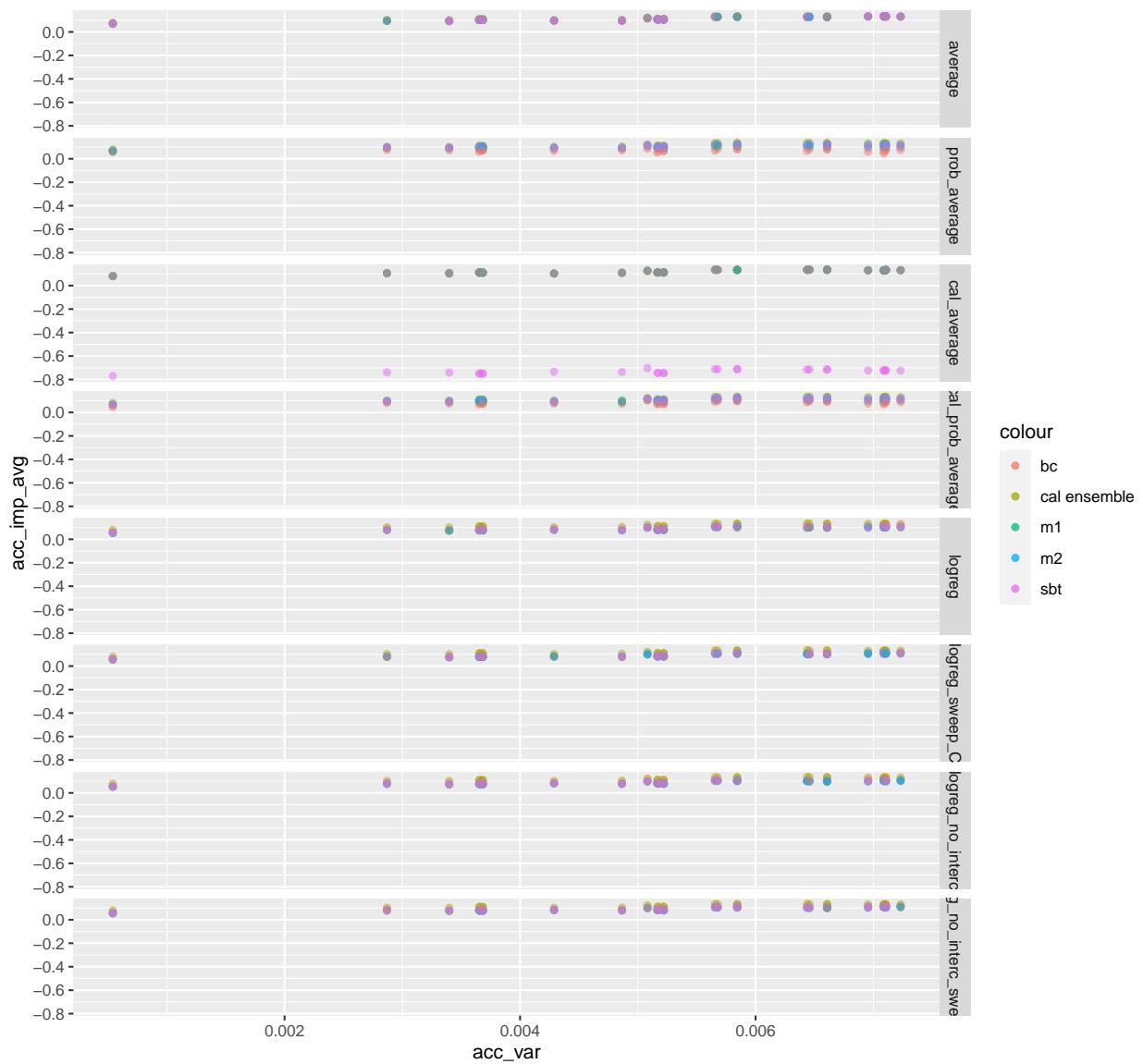
Accuracy improvement of ensemble over  
the average of networks vs nll\_avg.  
Ensemble size 6



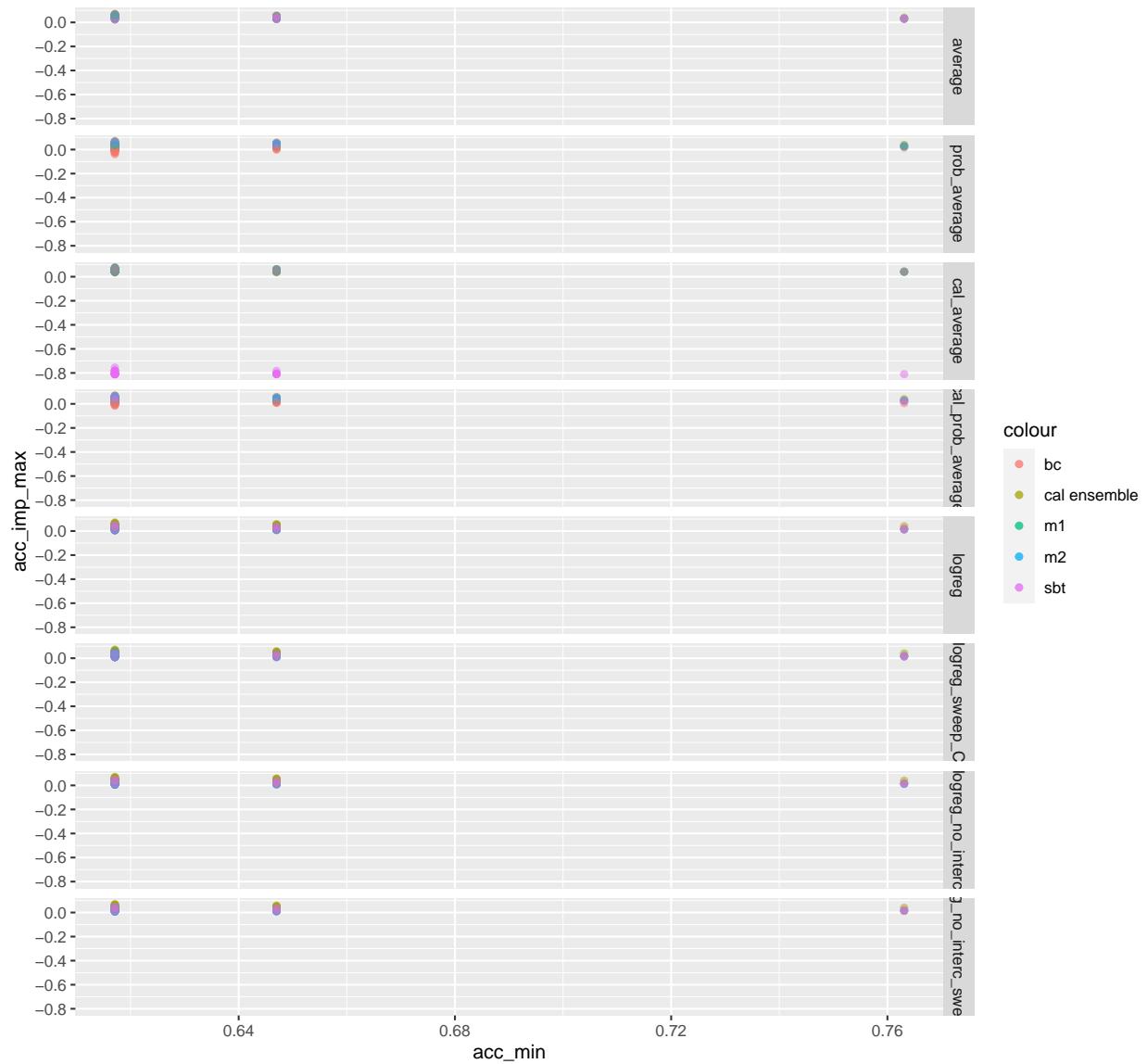
Accuracy improvement of ensemble over  
the best of networks vs acc\_var.  
Ensemble size 6



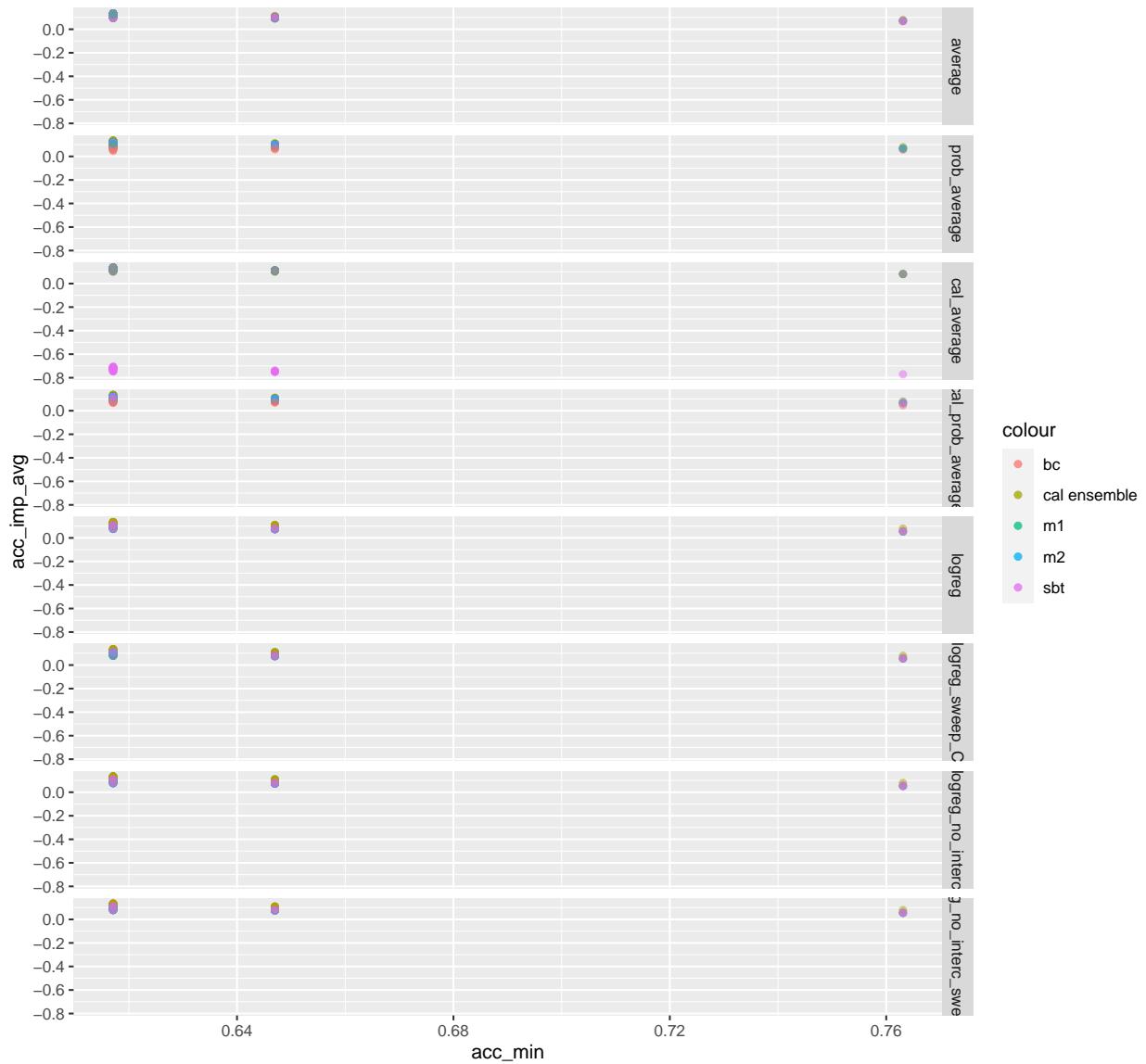
Accuracy improvement of ensemble over  
the average of networks vs acc\_var.  
Ensemble size 6



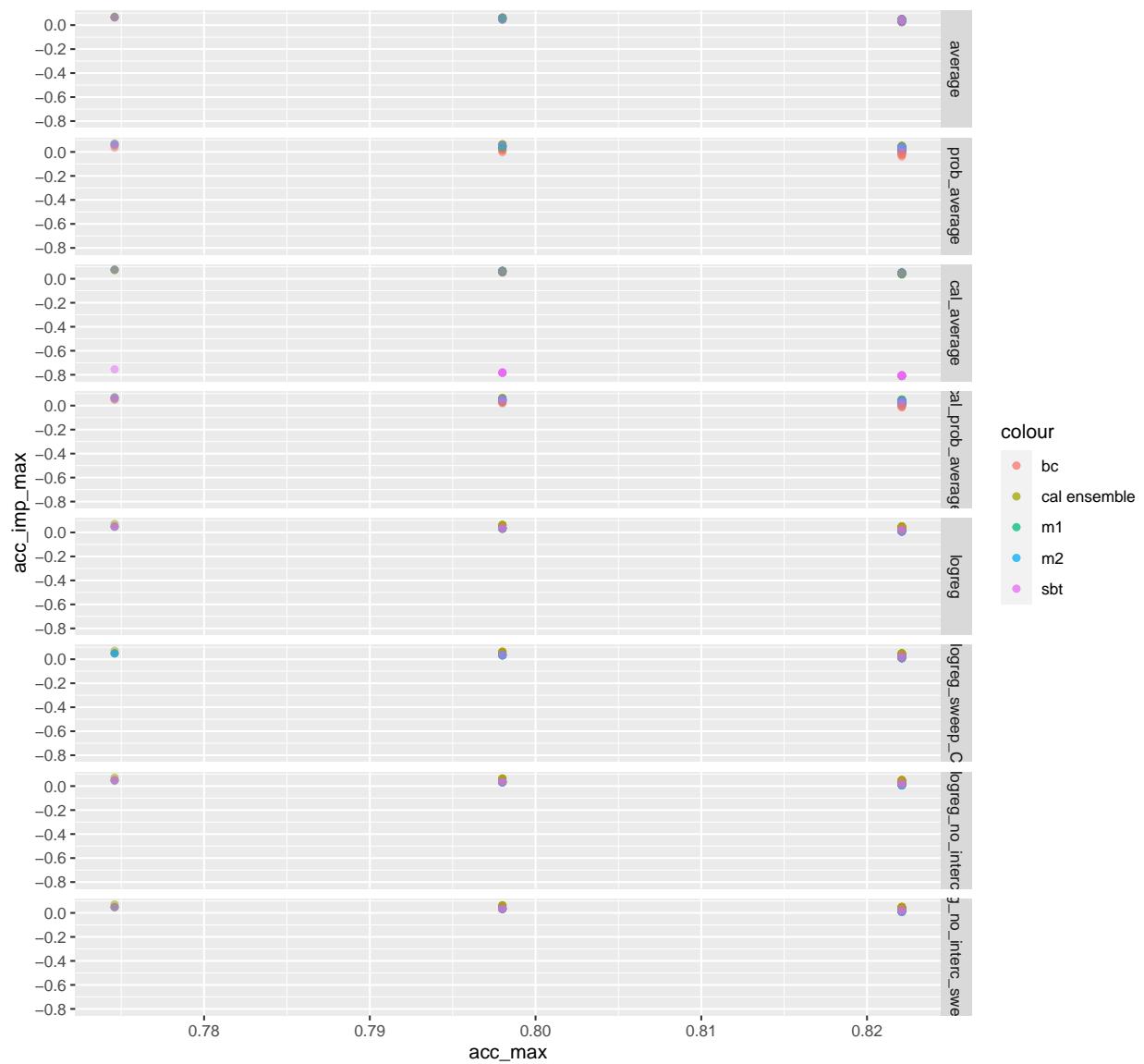
Accuracy improvement of ensemble over  
the best of networks vs acc\_min.  
Ensemble size 6



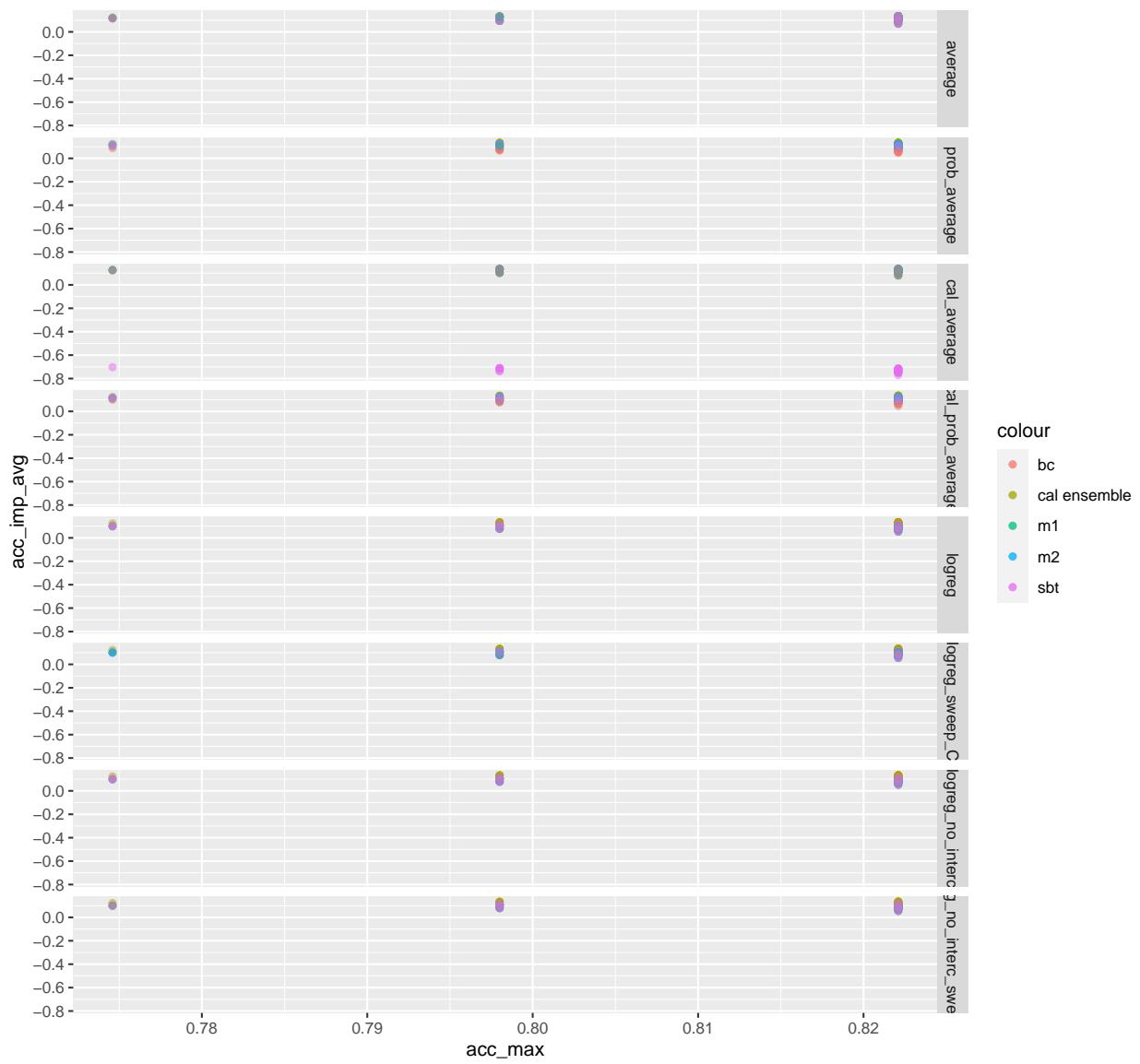
Accuracy improvement of ensemble over  
the average of networks vs acc\_min.  
Ensemble size 6



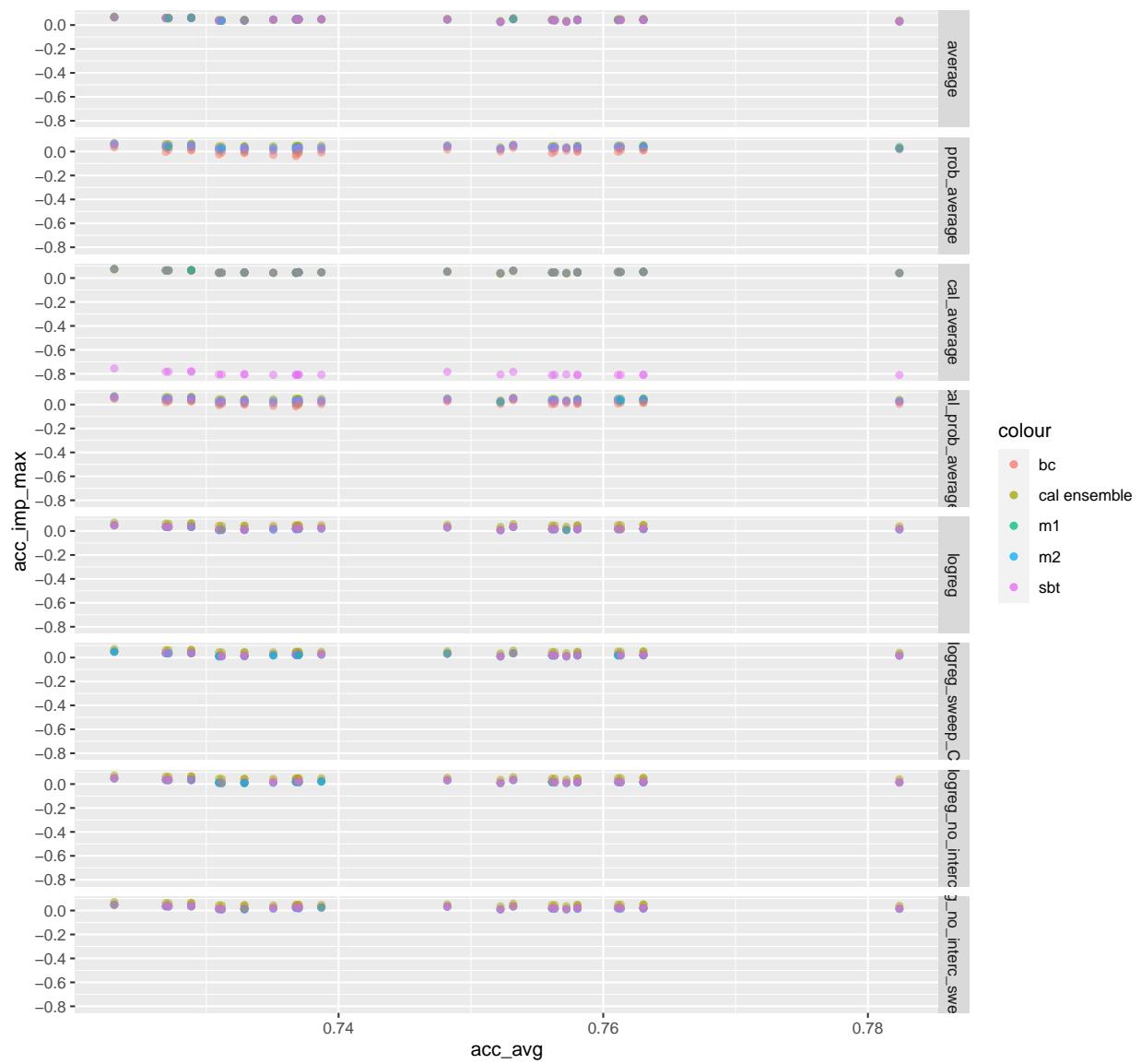
Accuracy improvement of ensemble over  
the best of networks vs acc\_max.  
Ensemble size 6



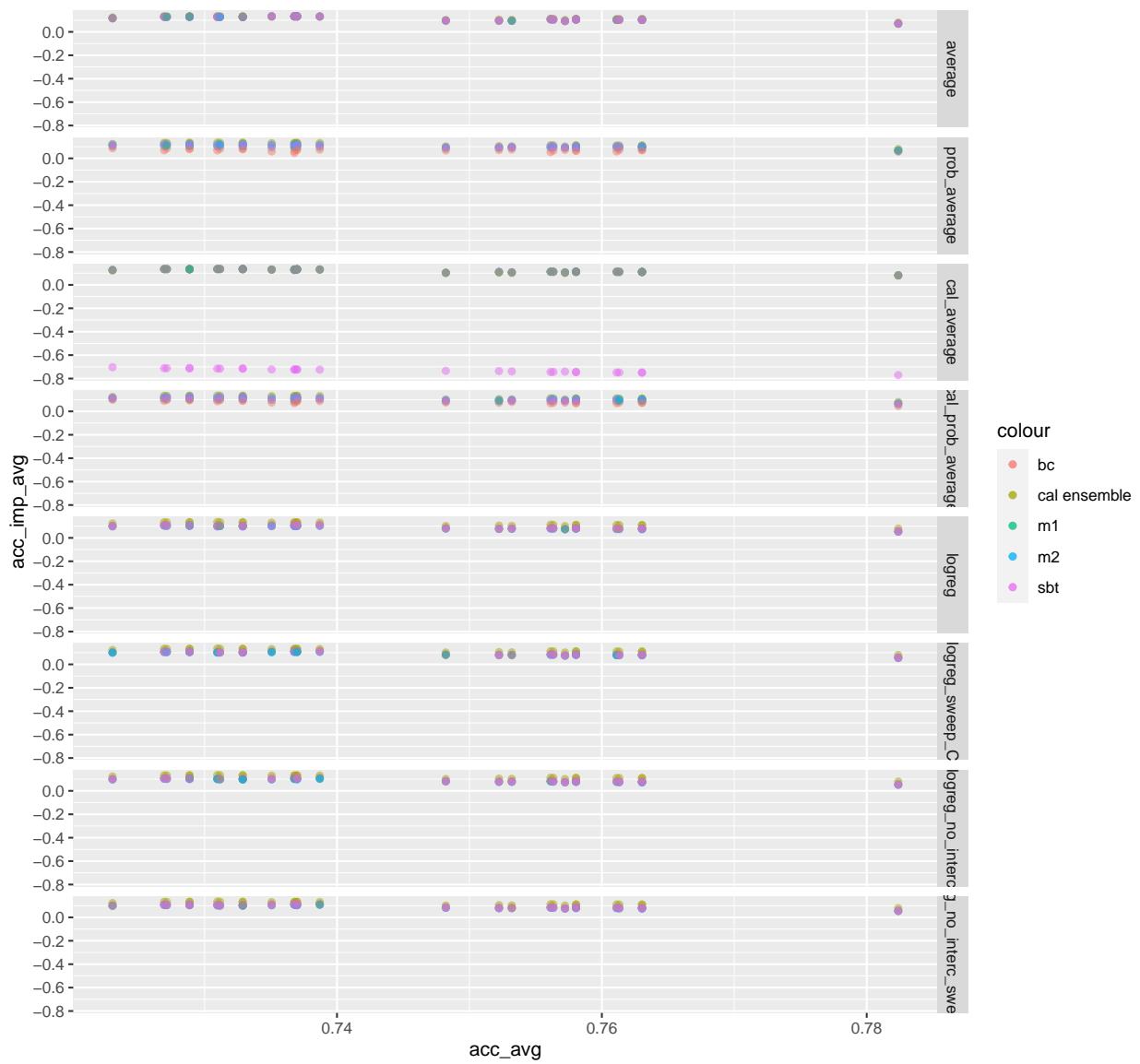
Accuracy improvement of ensemble over  
the average of networks vs acc\_max.  
Ensemble size 6



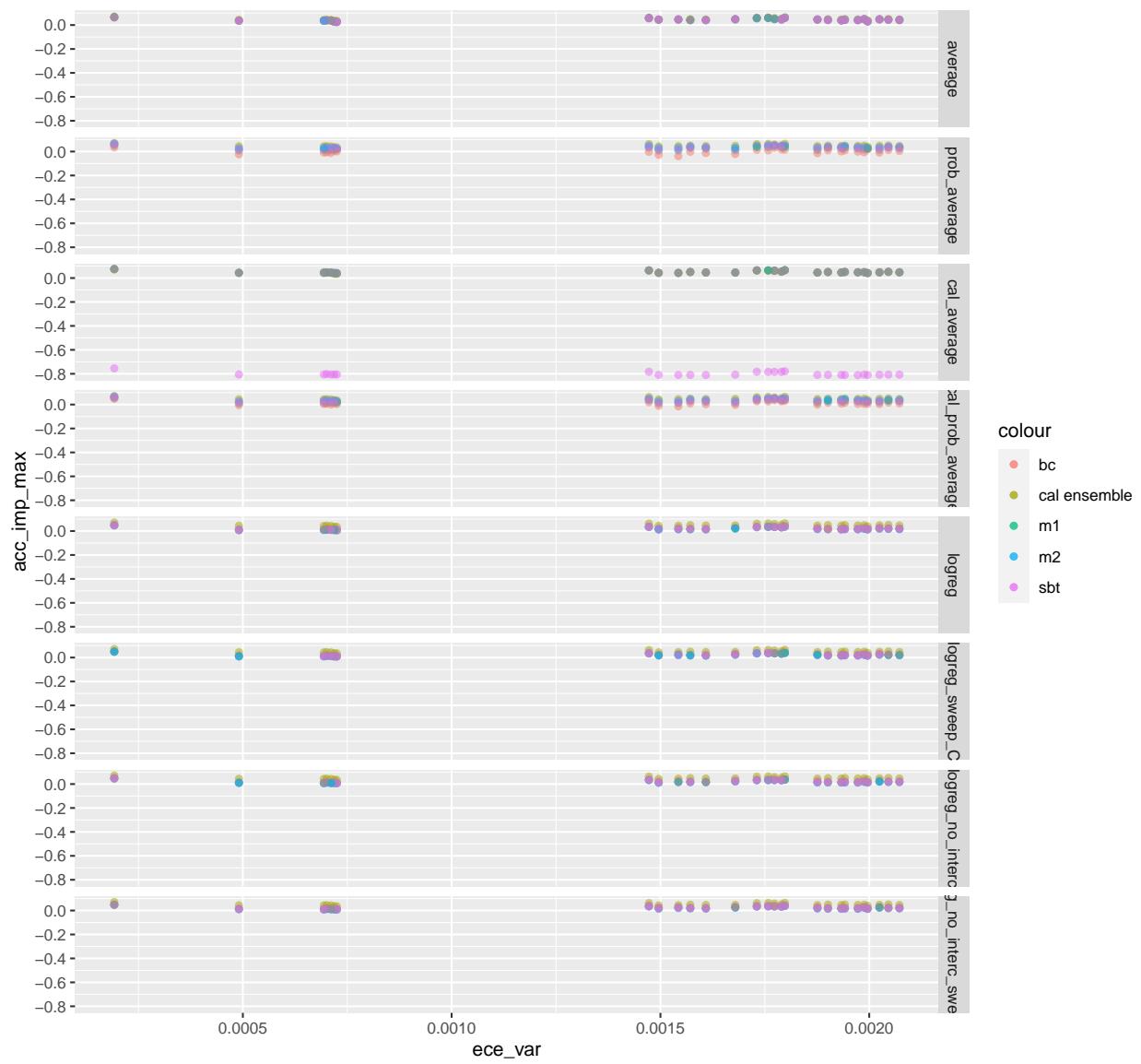
Accuracy improvement of ensemble over  
the best of networks vs acc\_avg.  
Ensemble size 6



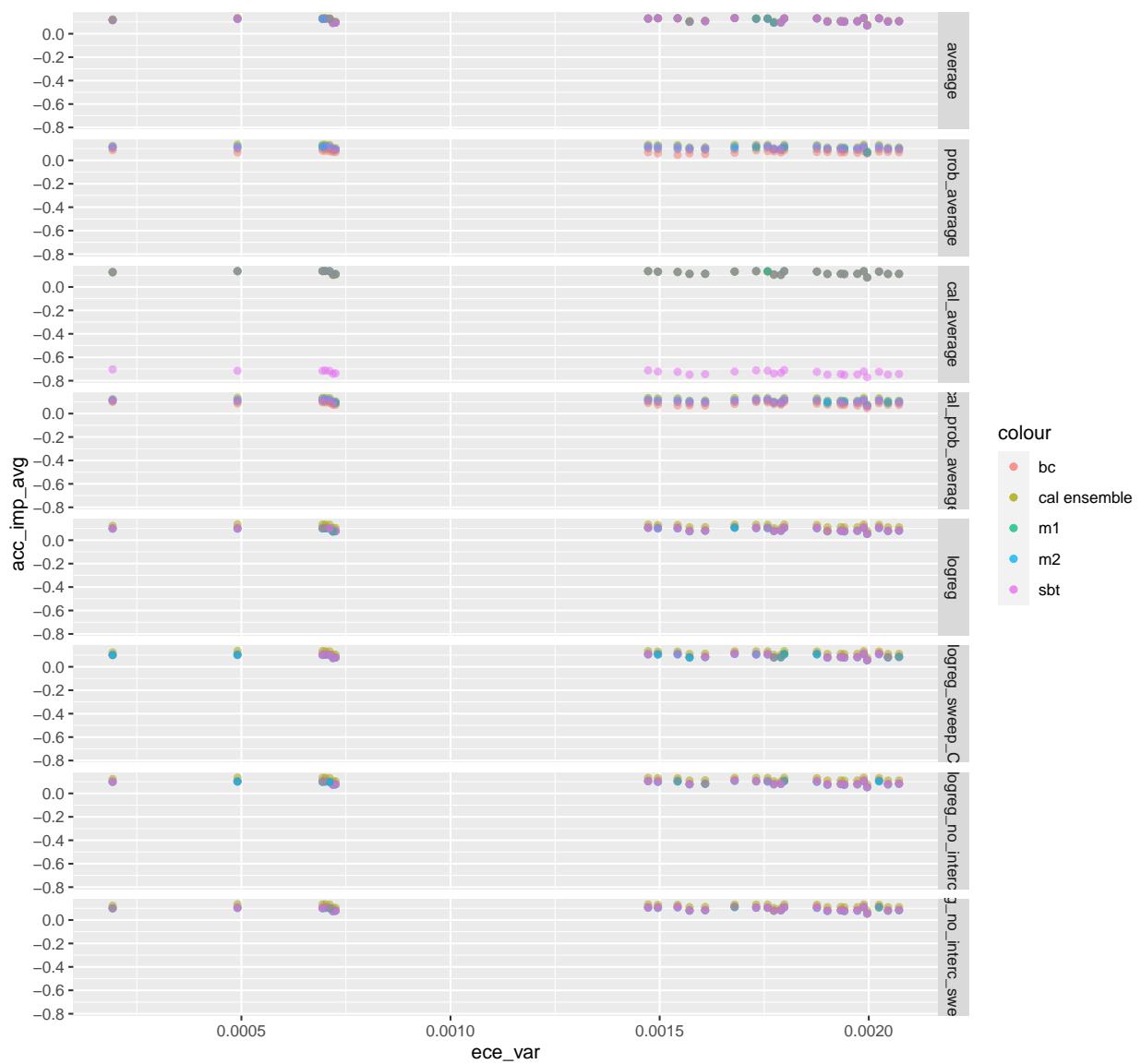
Accuracy improvement of ensemble over  
the average of networks vs acc\_avg.  
Ensemble size 6



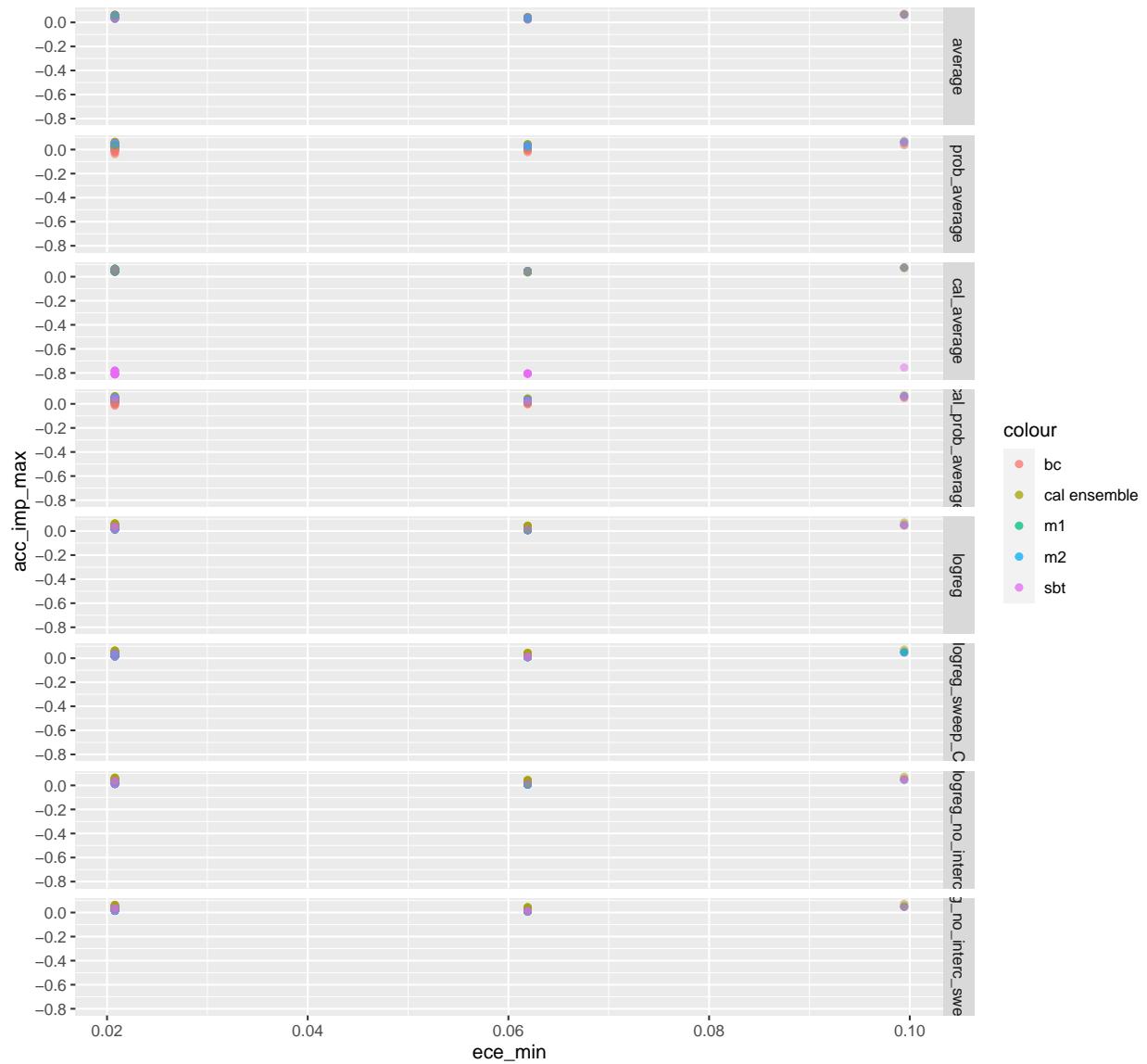
Accuracy improvement of ensemble over  
the best of networks vs ece\_var.  
Ensemble size 6



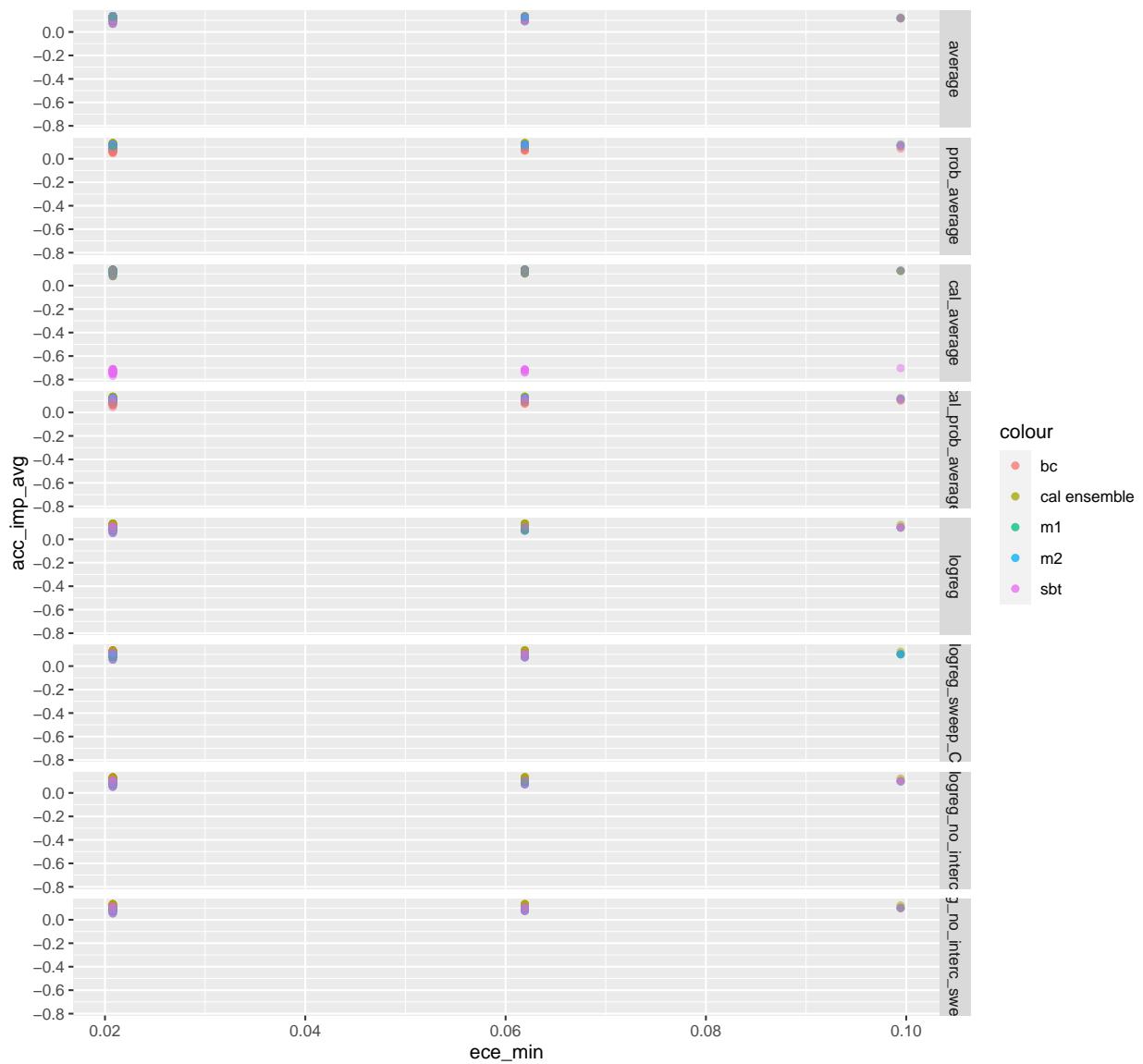
Accuracy improvement of ensemble over  
the average of networks vs ece\_var.  
Ensemble size 6



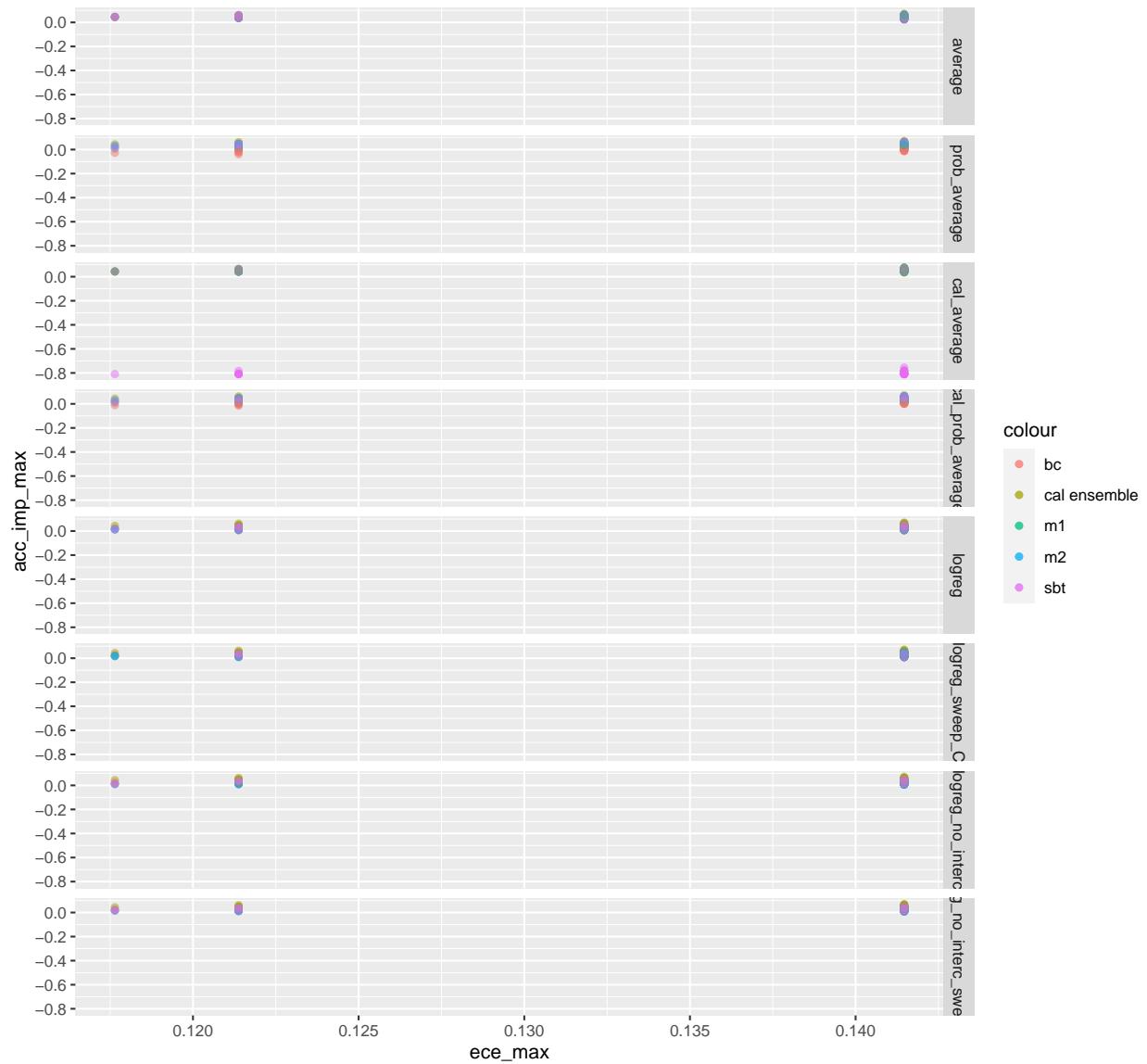
Accuracy improvement of ensemble over  
the best of networks vs ece\_min.  
Ensemble size 6



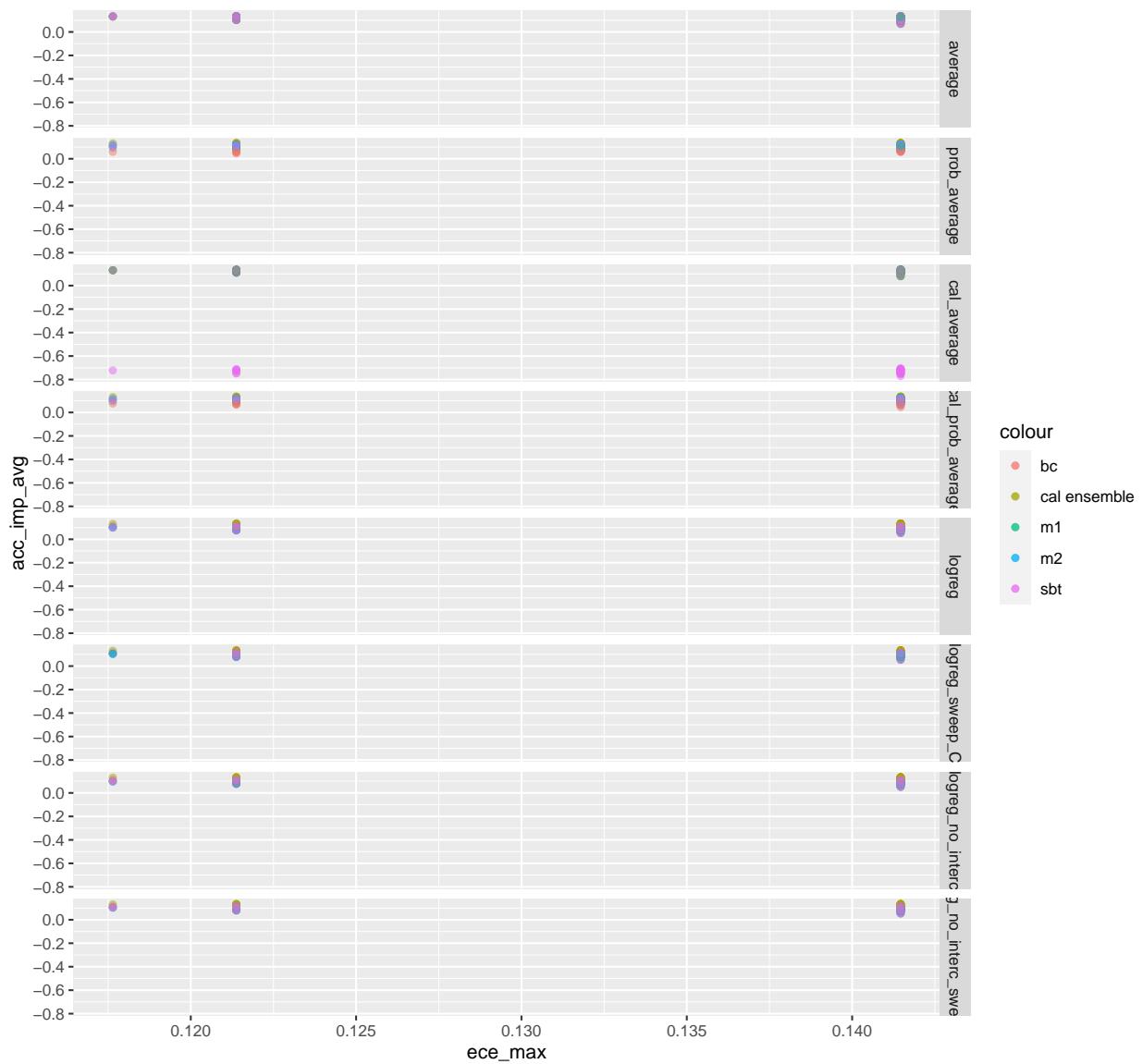
Accuracy improvement of ensemble over  
the average of networks vs ece\_min.  
Ensemble size 6



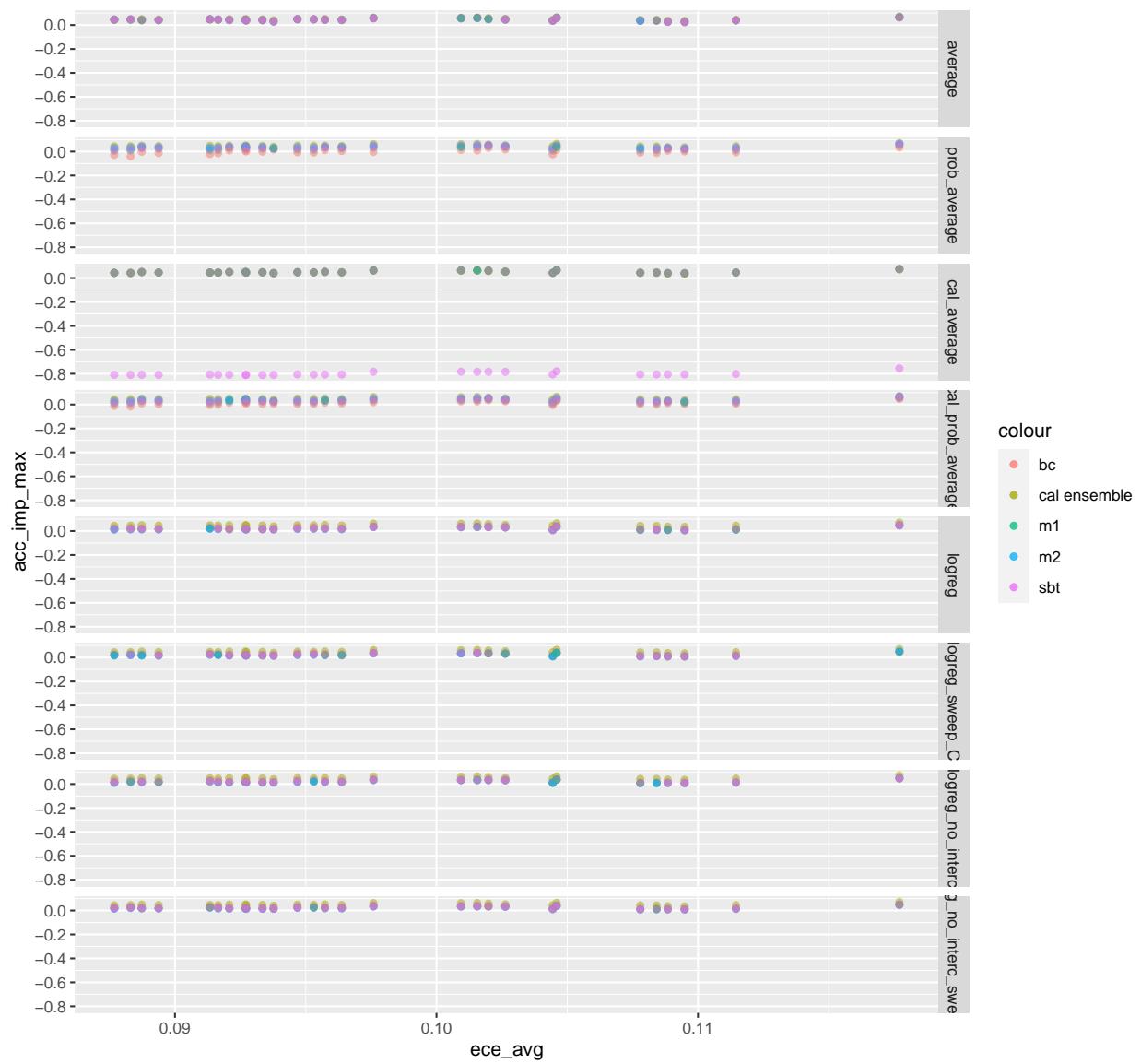
Accuracy improvement of ensemble over  
the best of networks vs ece\_max.  
Ensemble size 6



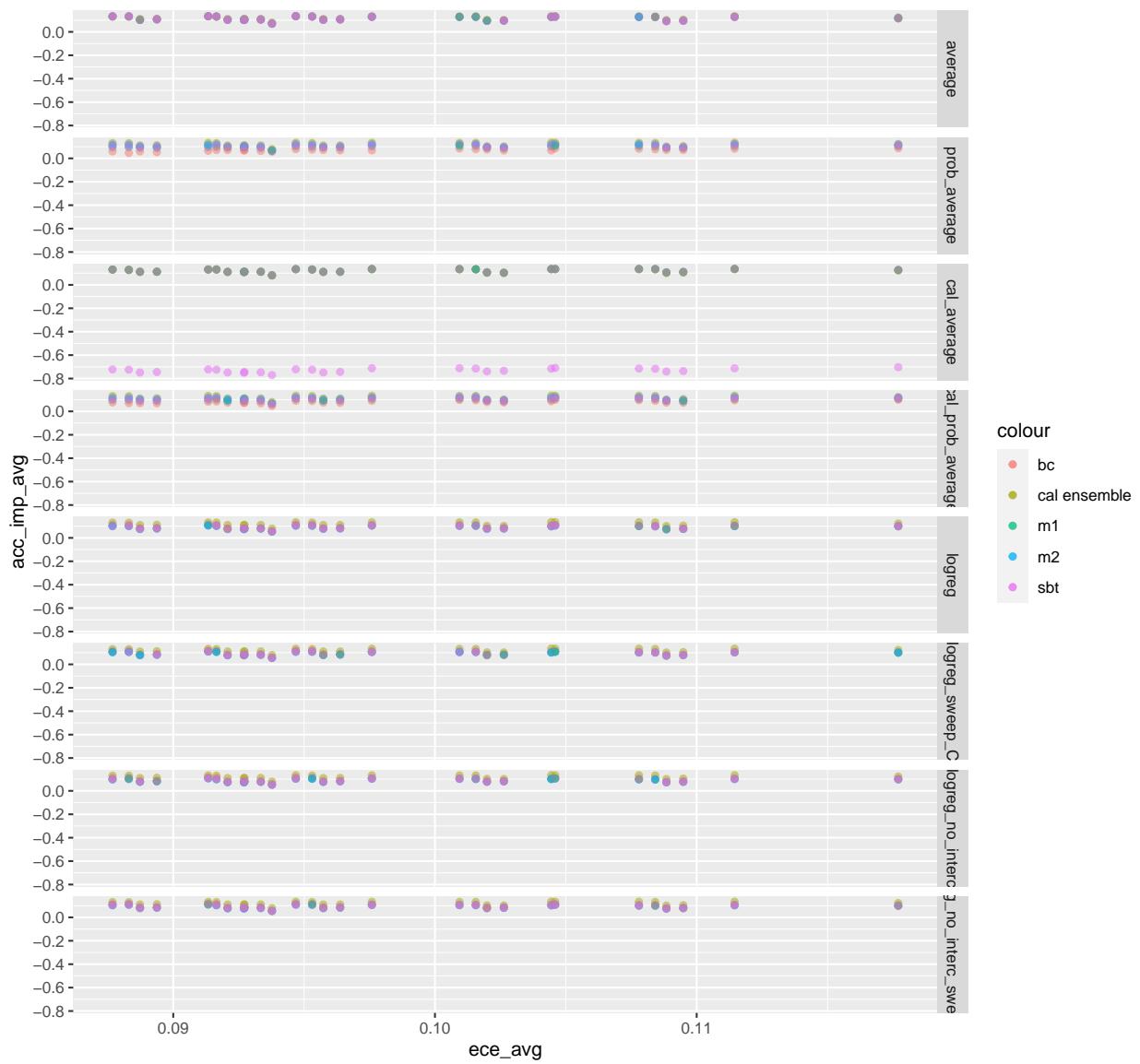
Accuracy improvement of ensemble over  
the average of networks vs ece\_max.  
Ensemble size 6



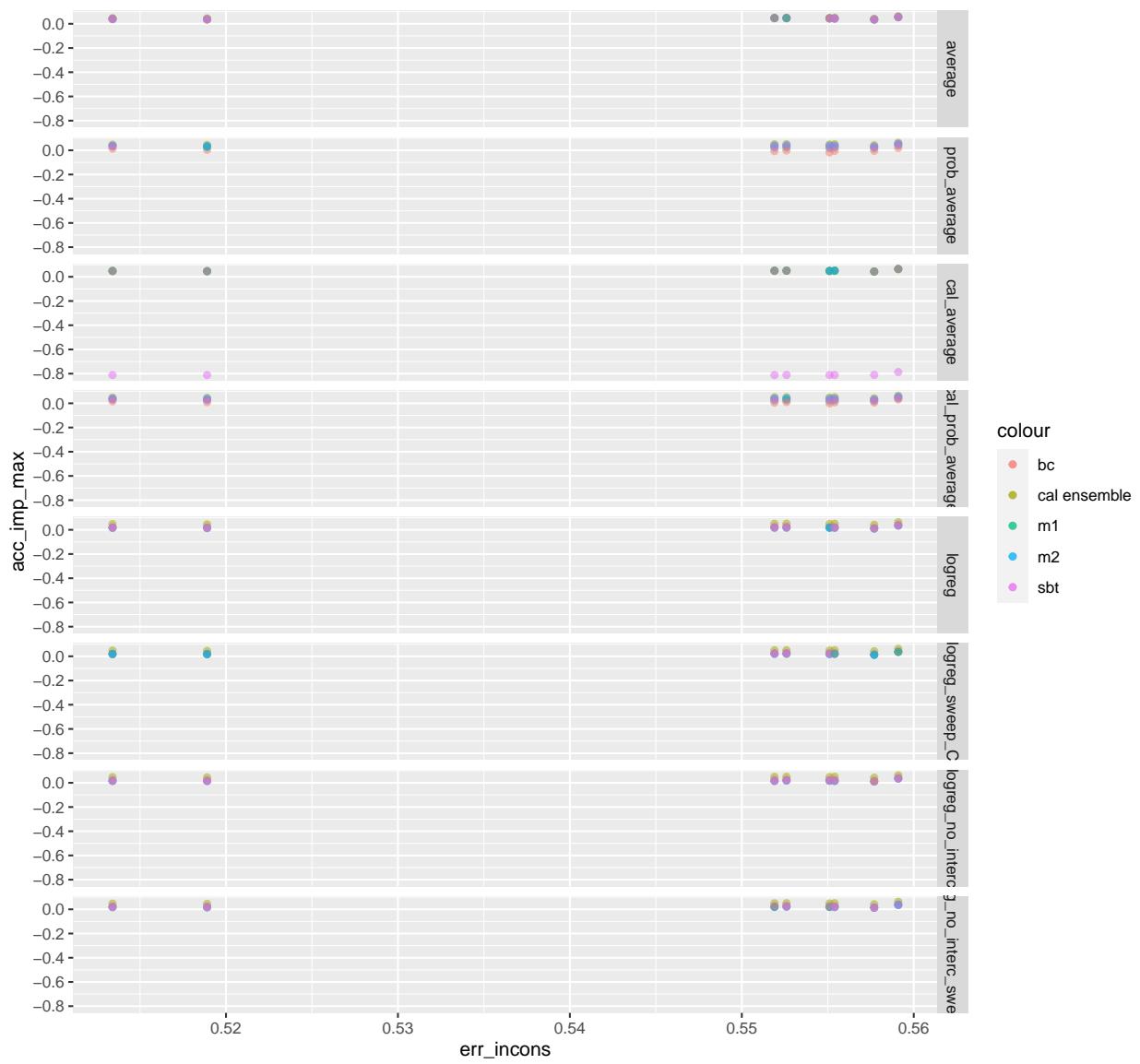
Accuracy improvement of ensemble over  
the best of networks vs ece\_avg.  
Ensemble size 6



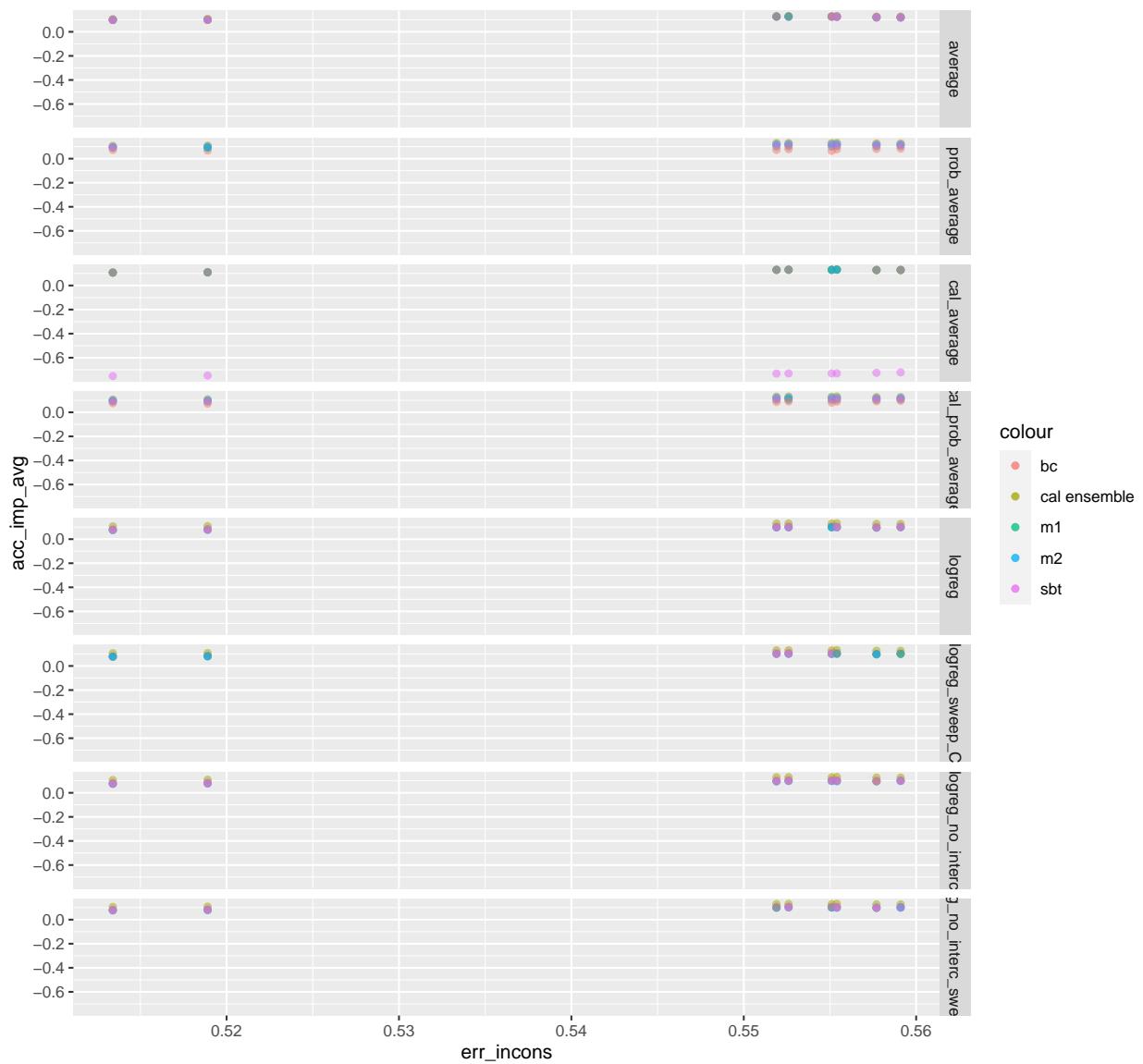
Accuracy improvement of ensemble over  
the average of networks vs ece\_avg.  
Ensemble size 6



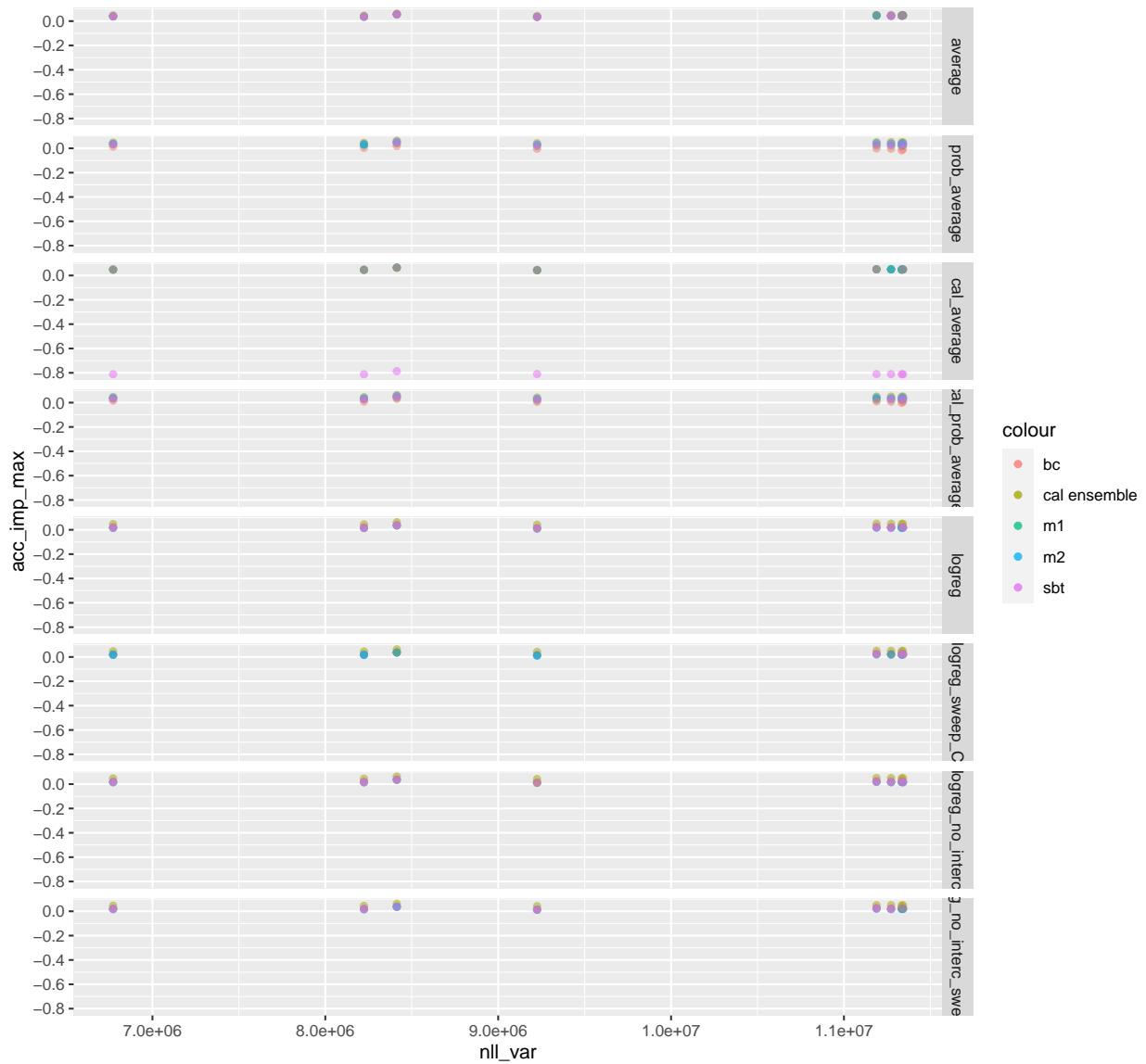
Accuracy improvement of ensemble over  
the best of networks vs err\_incons.  
Ensemble size 7



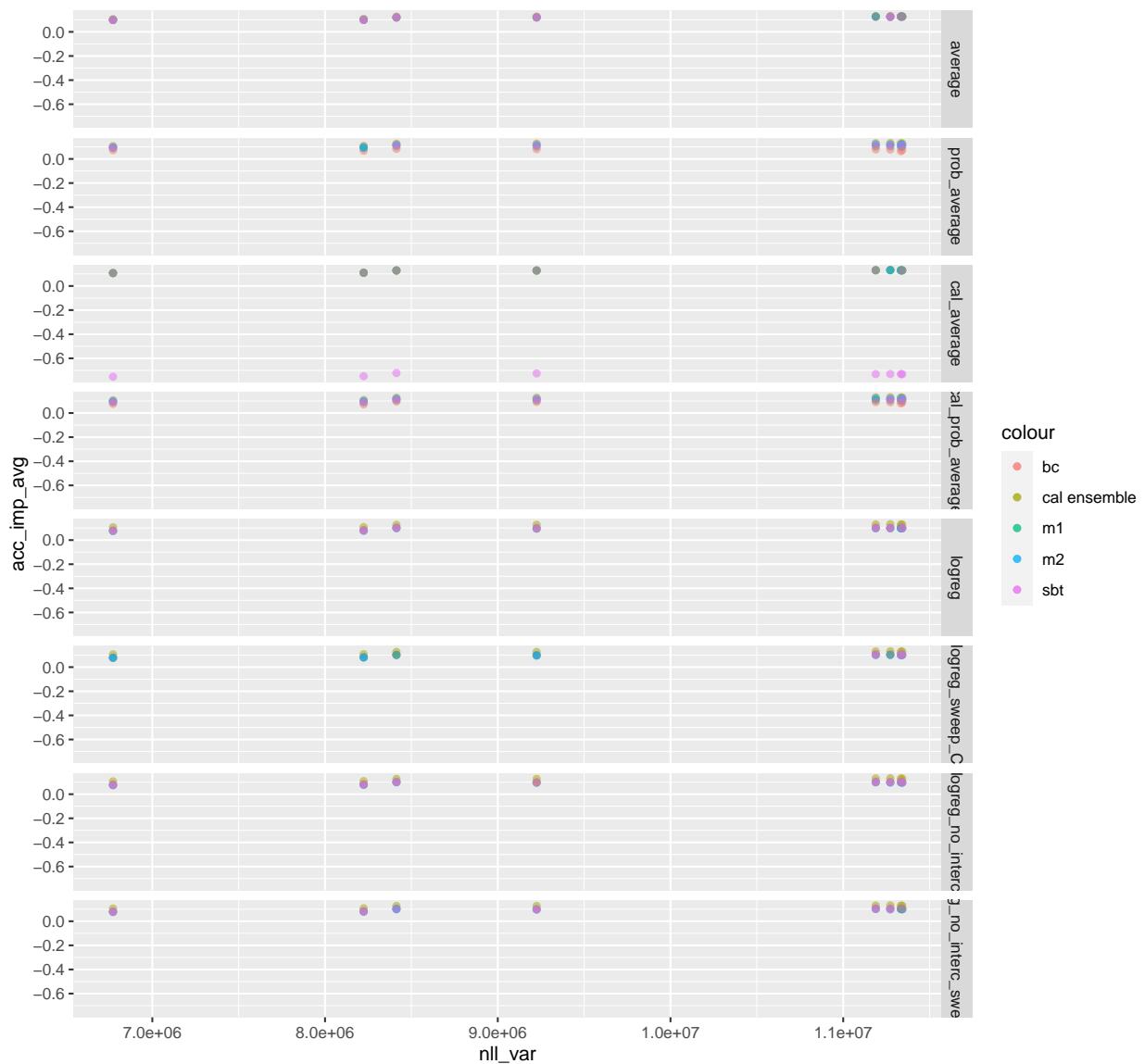
Accuracy improvement of ensemble over  
the average of networks vs err\_incons.  
Ensemble size 7



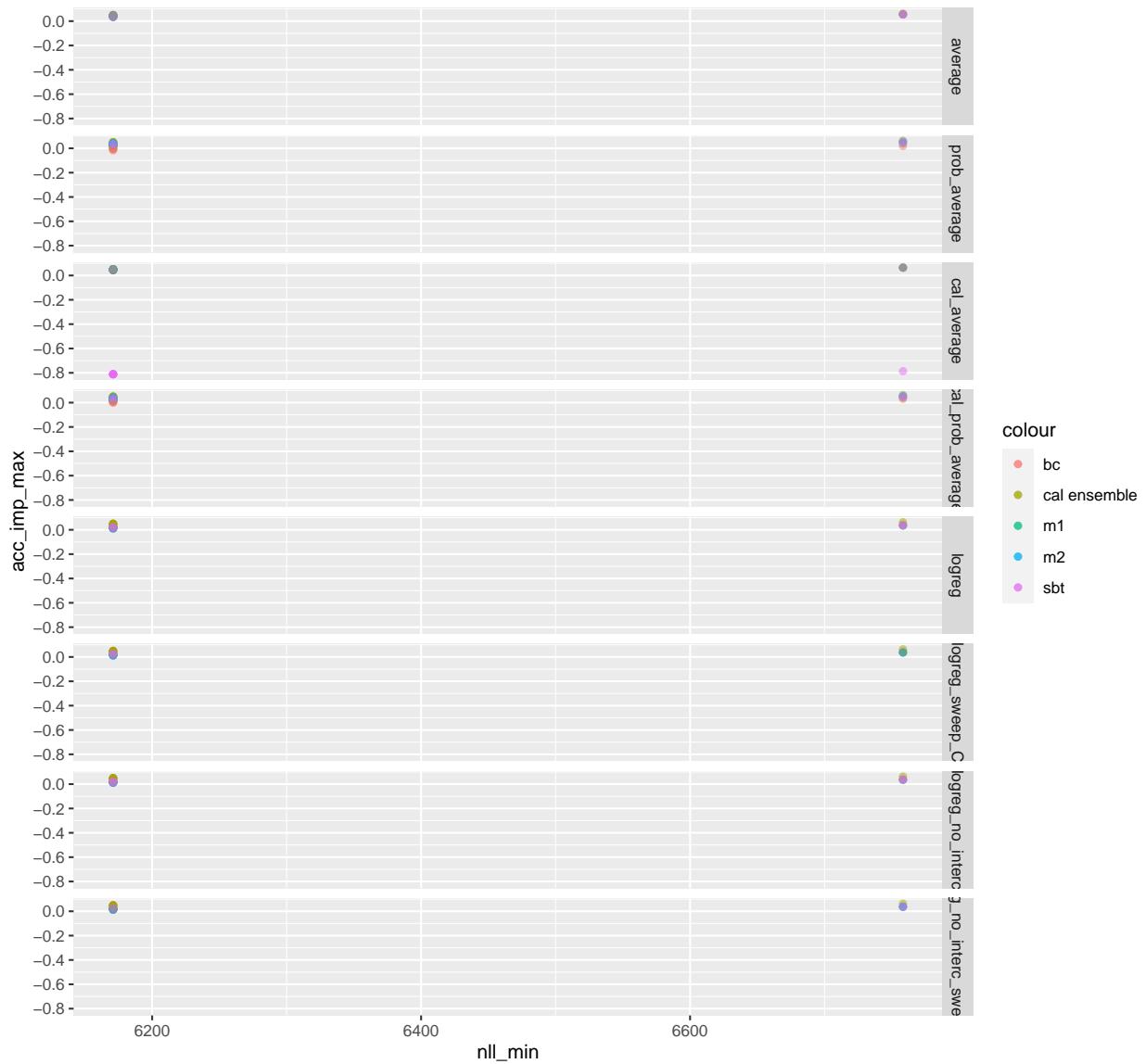
Accuracy improvement of ensemble over  
the best of networks vs nll\_var.  
Ensemble size 7



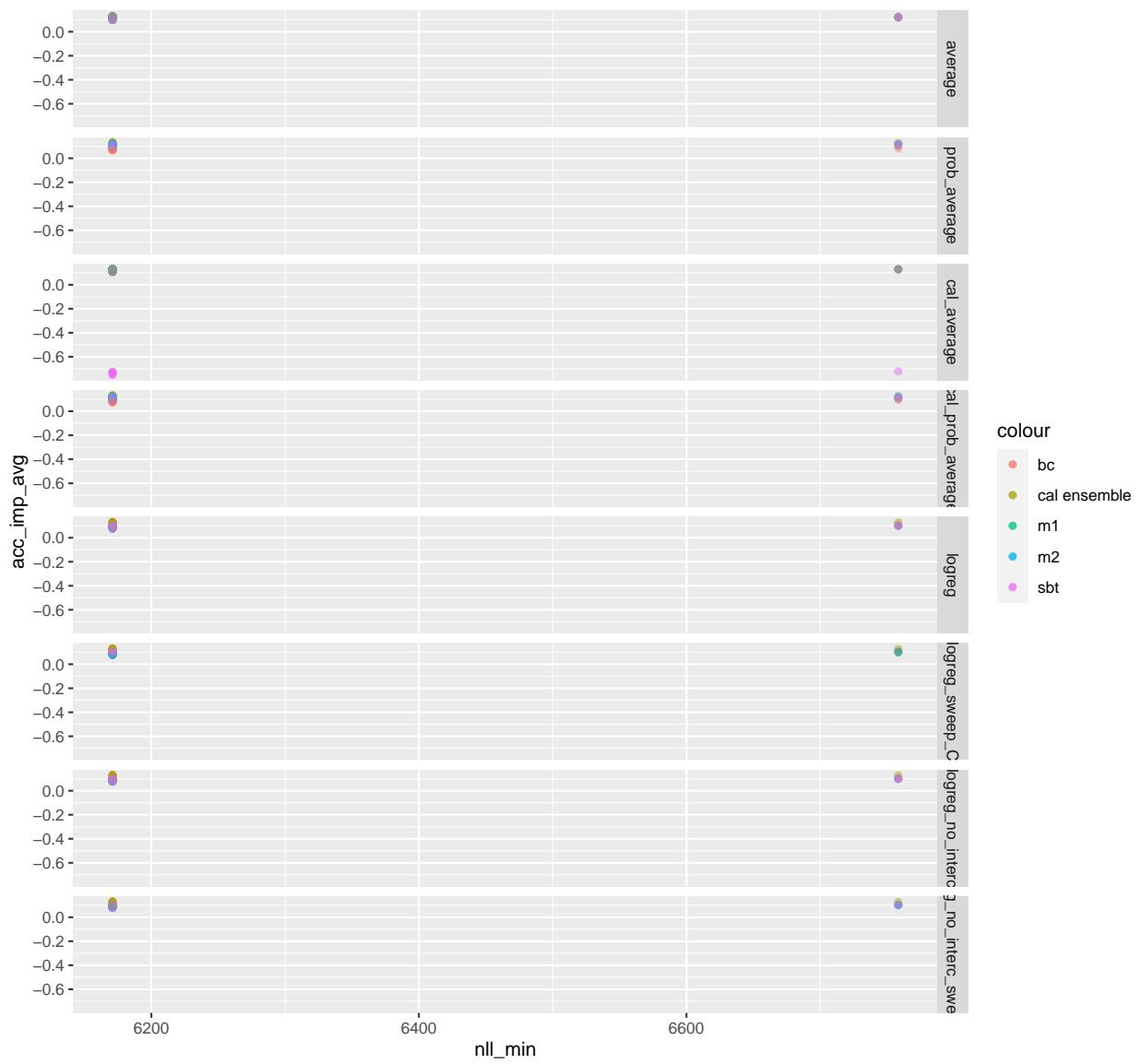
Accuracy improvement of ensemble over  
the average of networks vs nll\_var.  
Ensemble size 7



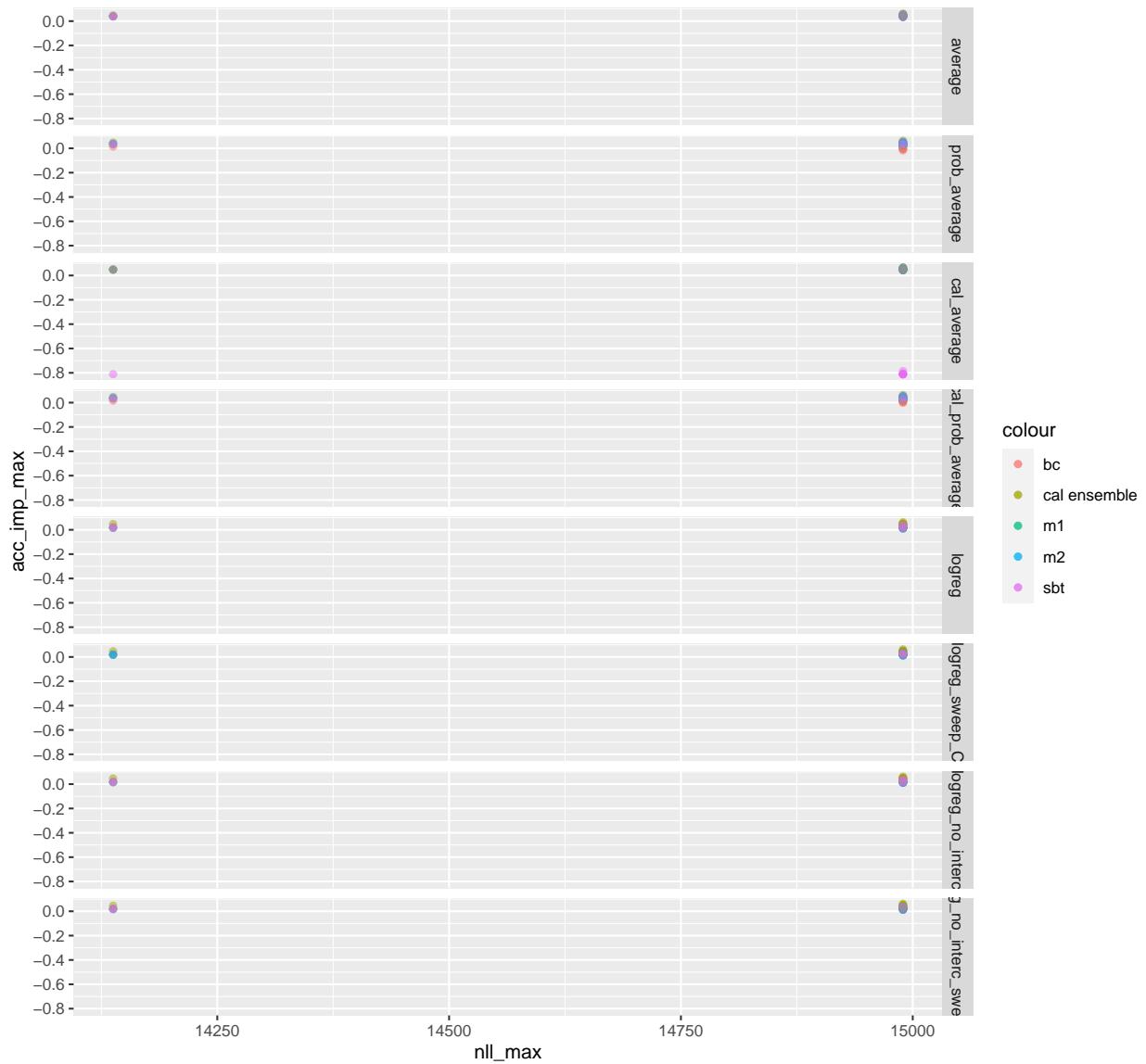
Accuracy improvement of ensemble over  
the best of networks vs nll\_min.  
Ensemble size 7



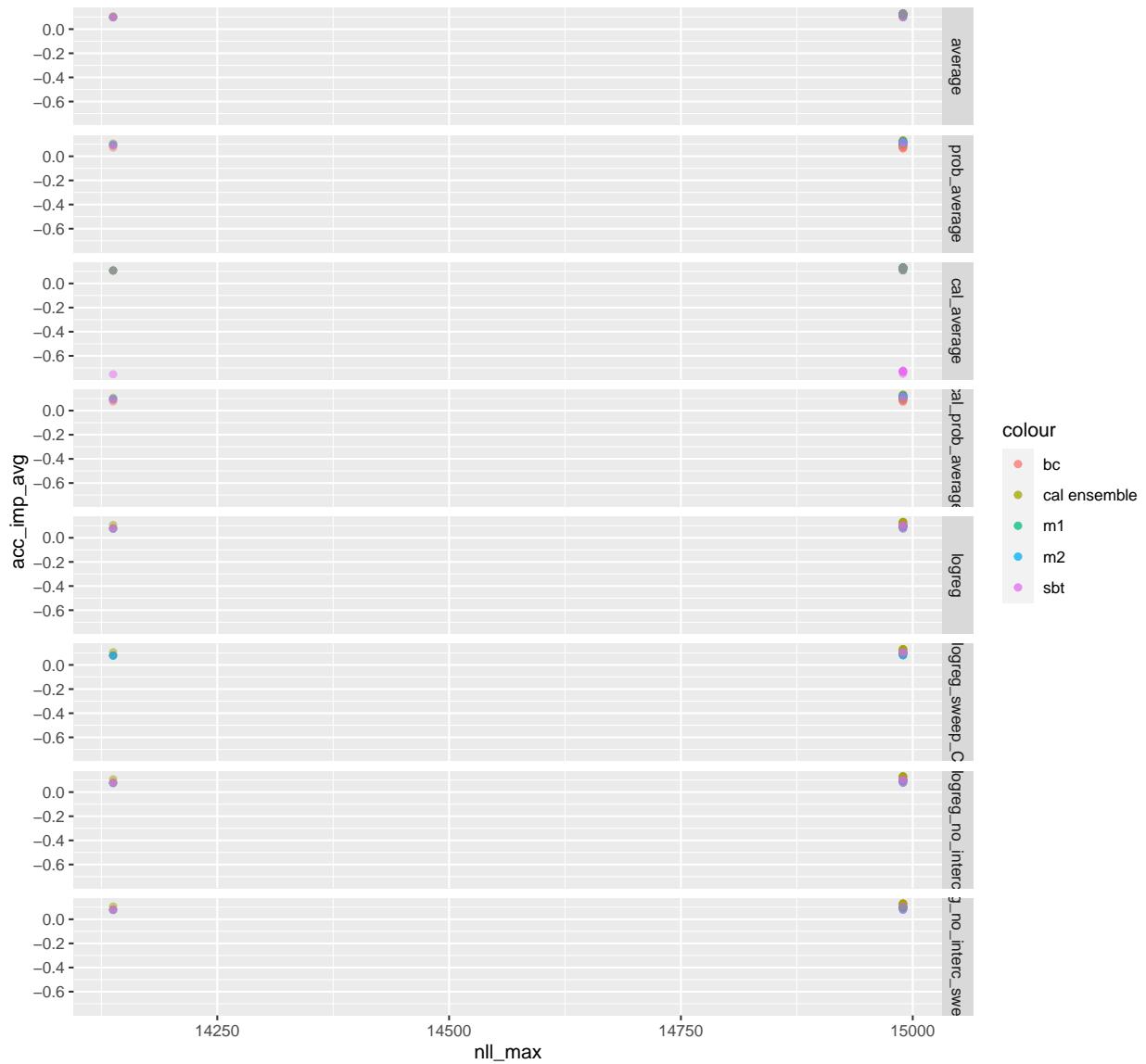
Accuracy improvement of ensemble over  
the average of networks vs nll\_min.  
Ensemble size 7



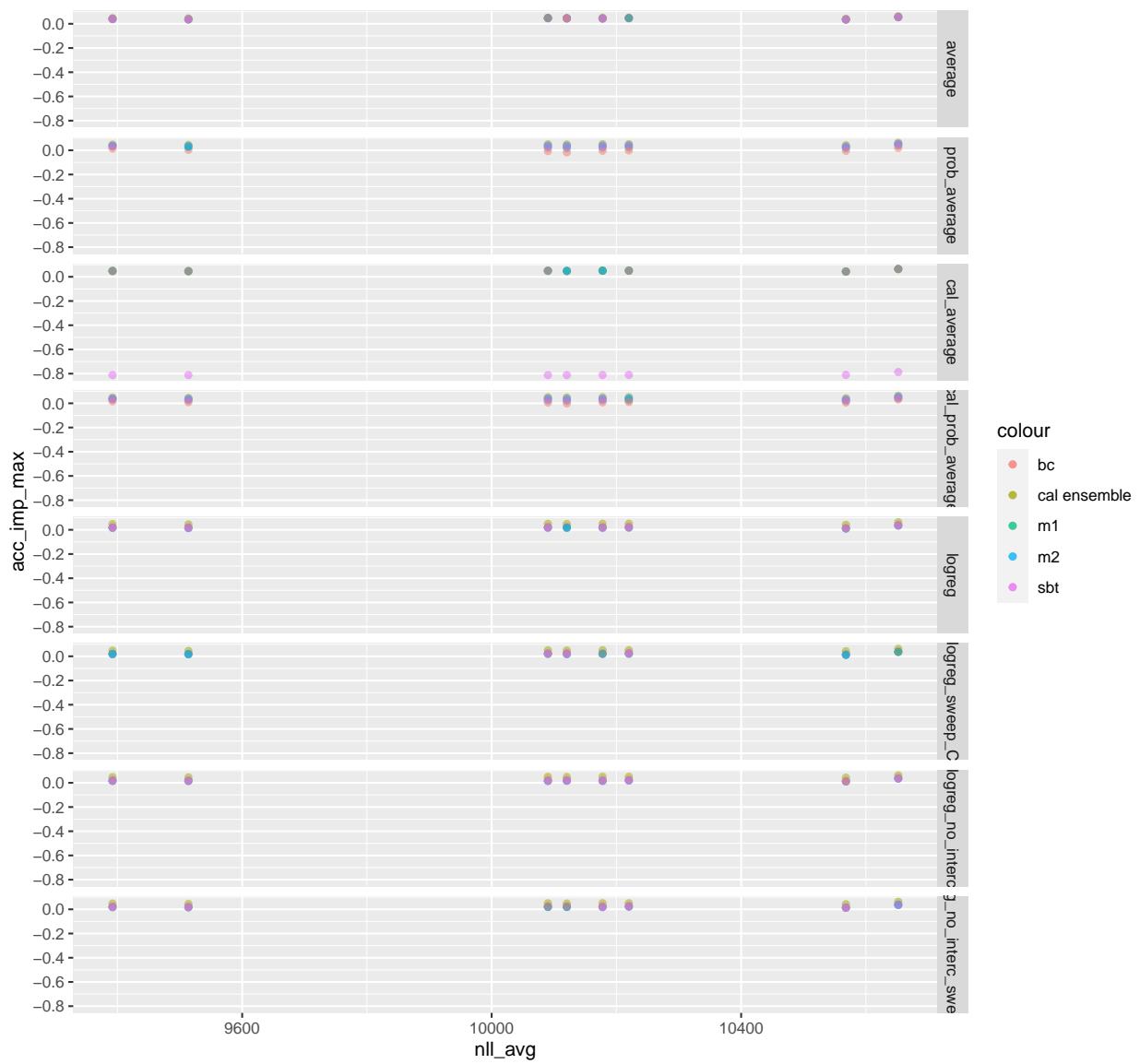
Accuracy improvement of ensemble over  
the best of networks vs nll\_max.  
Ensemble size 7



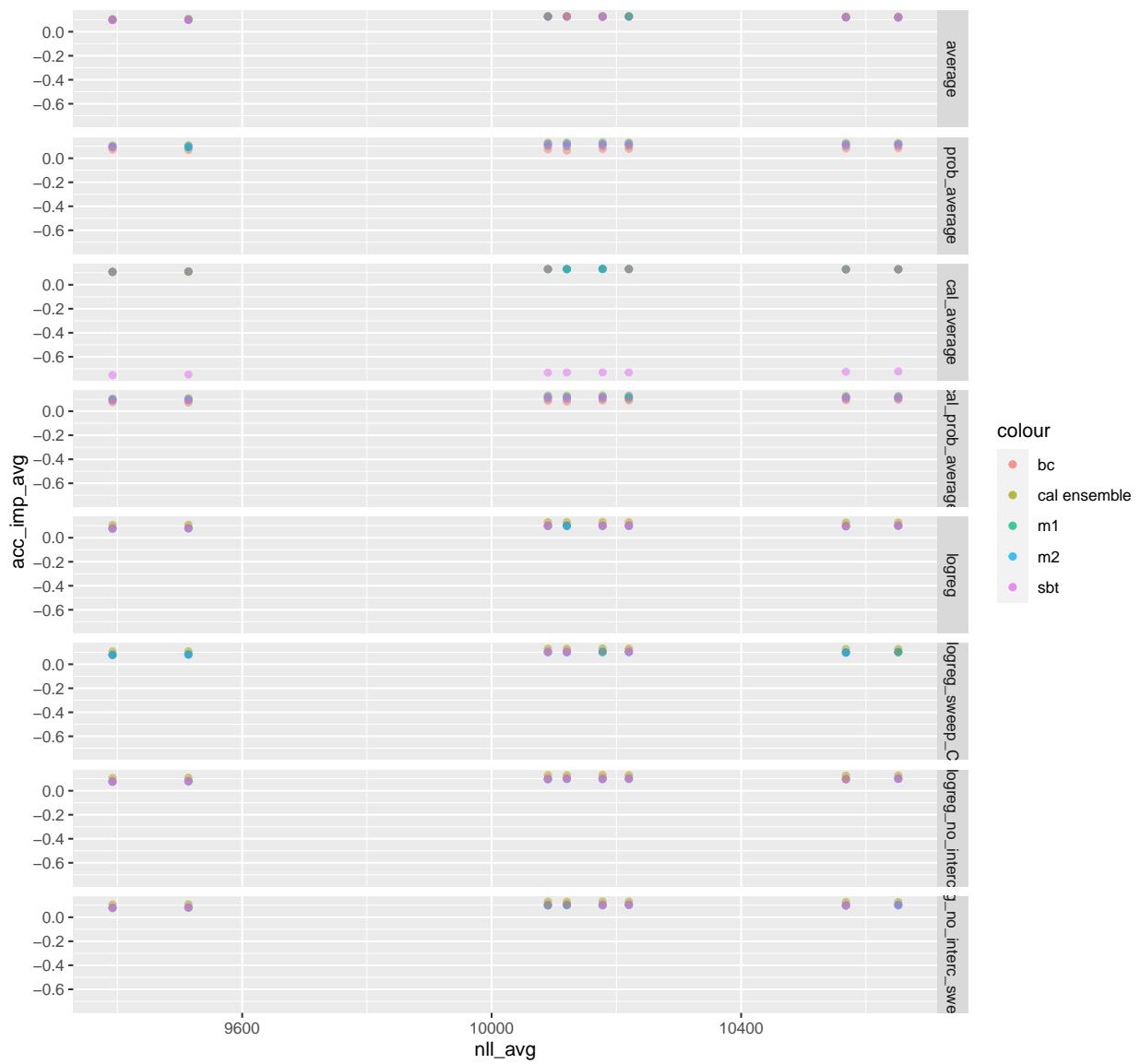
Accuracy improvement of ensemble over  
the average of networks vs nll\_max.  
Ensemble size 7



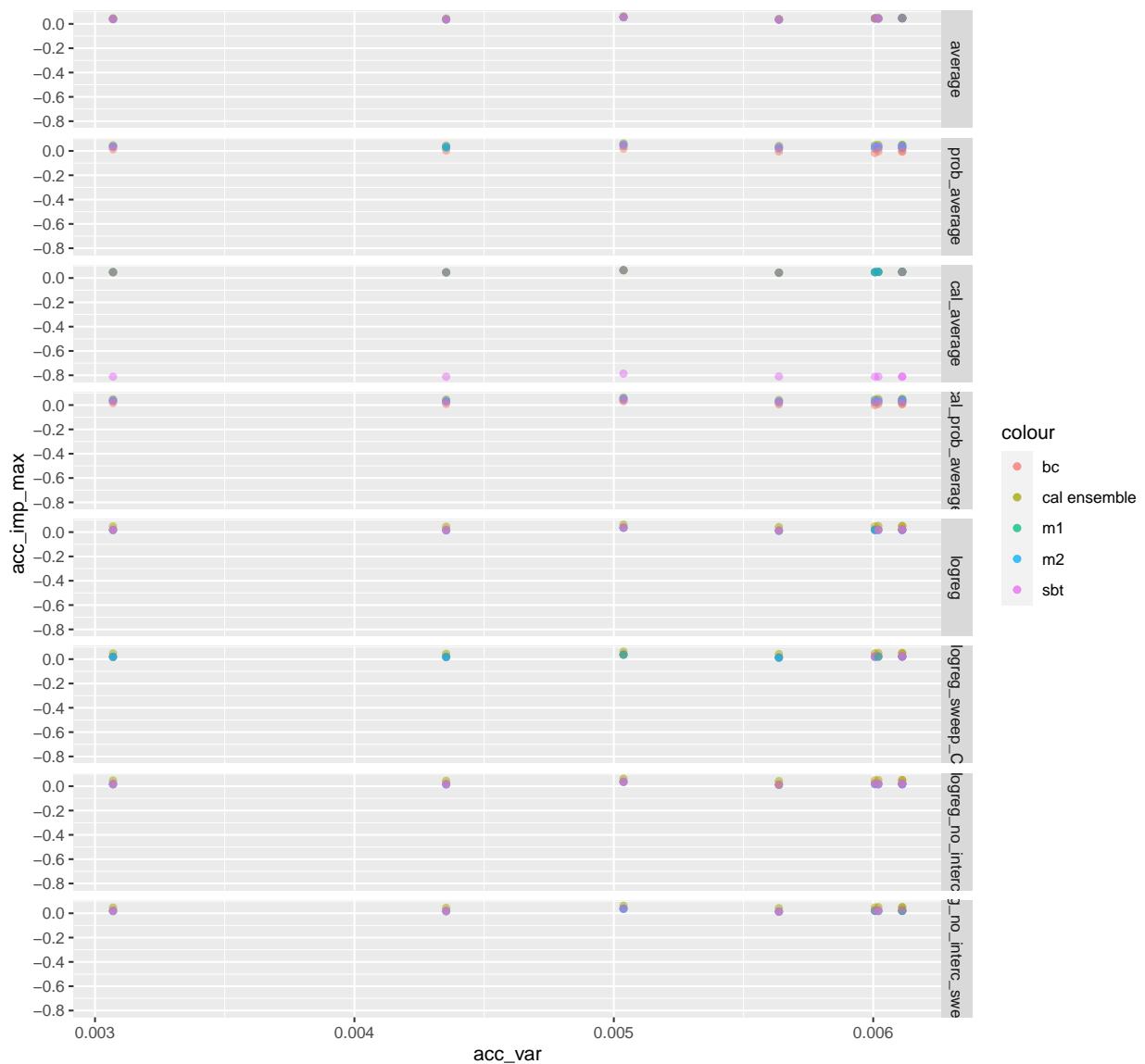
Accuracy improvement of ensemble over  
the best of networks vs nll\_avg.  
Ensemble size 7



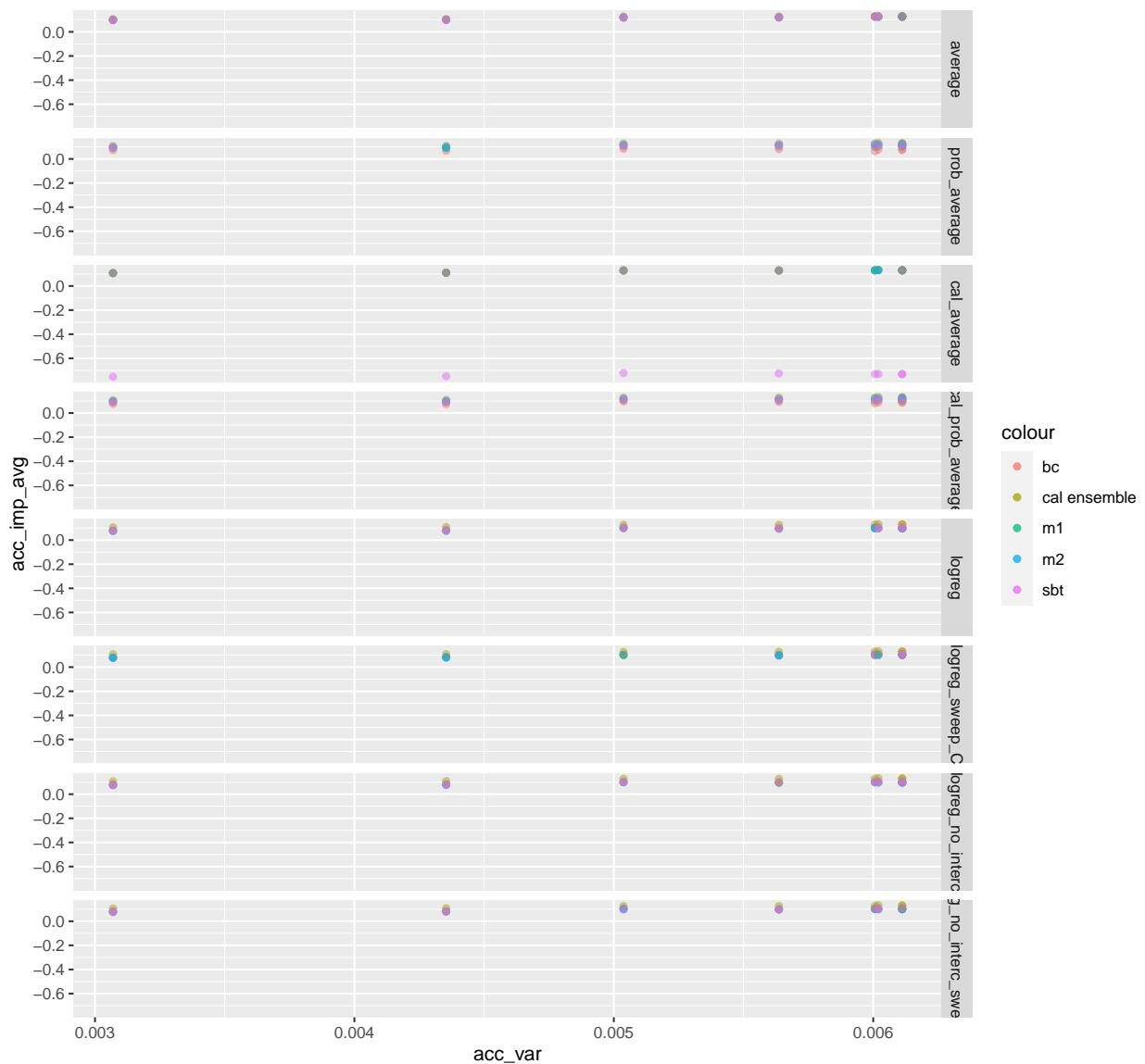
Accuracy improvement of ensemble over  
the average of networks vs nll\_avg.  
Ensemble size 7



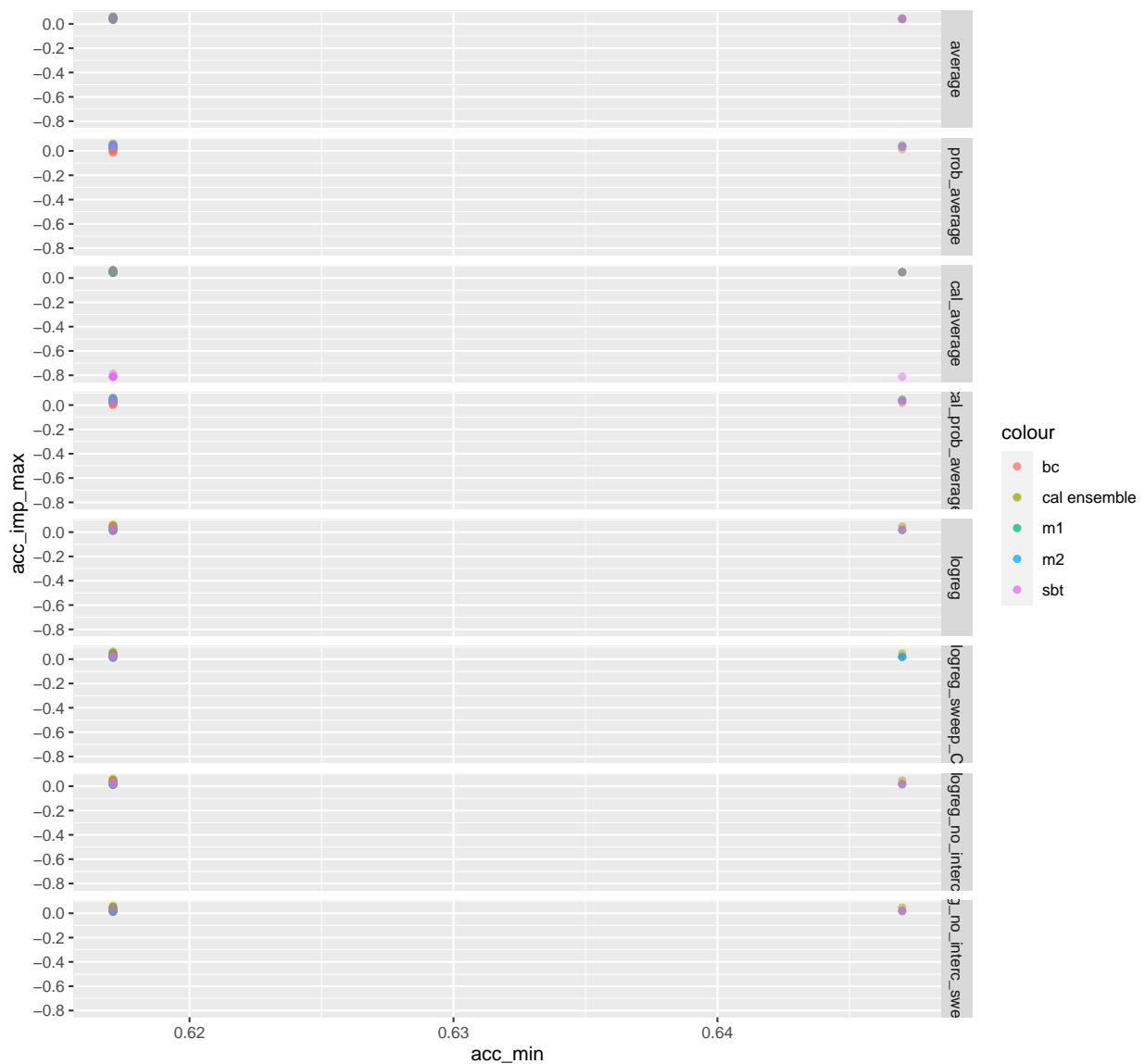
Accuracy improvement of ensemble over  
the best of networks vs acc\_var.  
Ensemble size 7



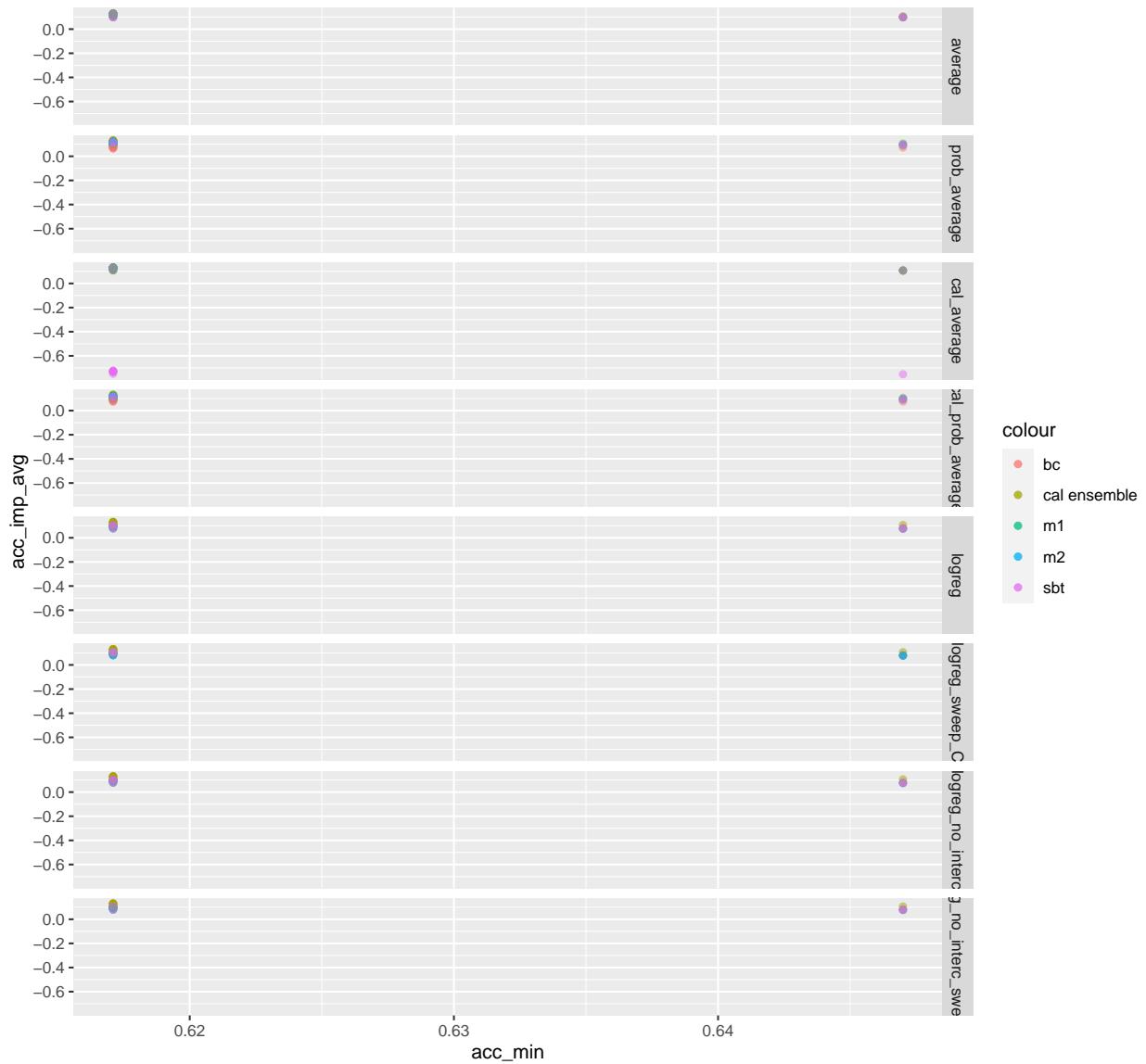
Accuracy improvement of ensemble over  
the average of networks vs acc\_var.  
Ensemble size 7



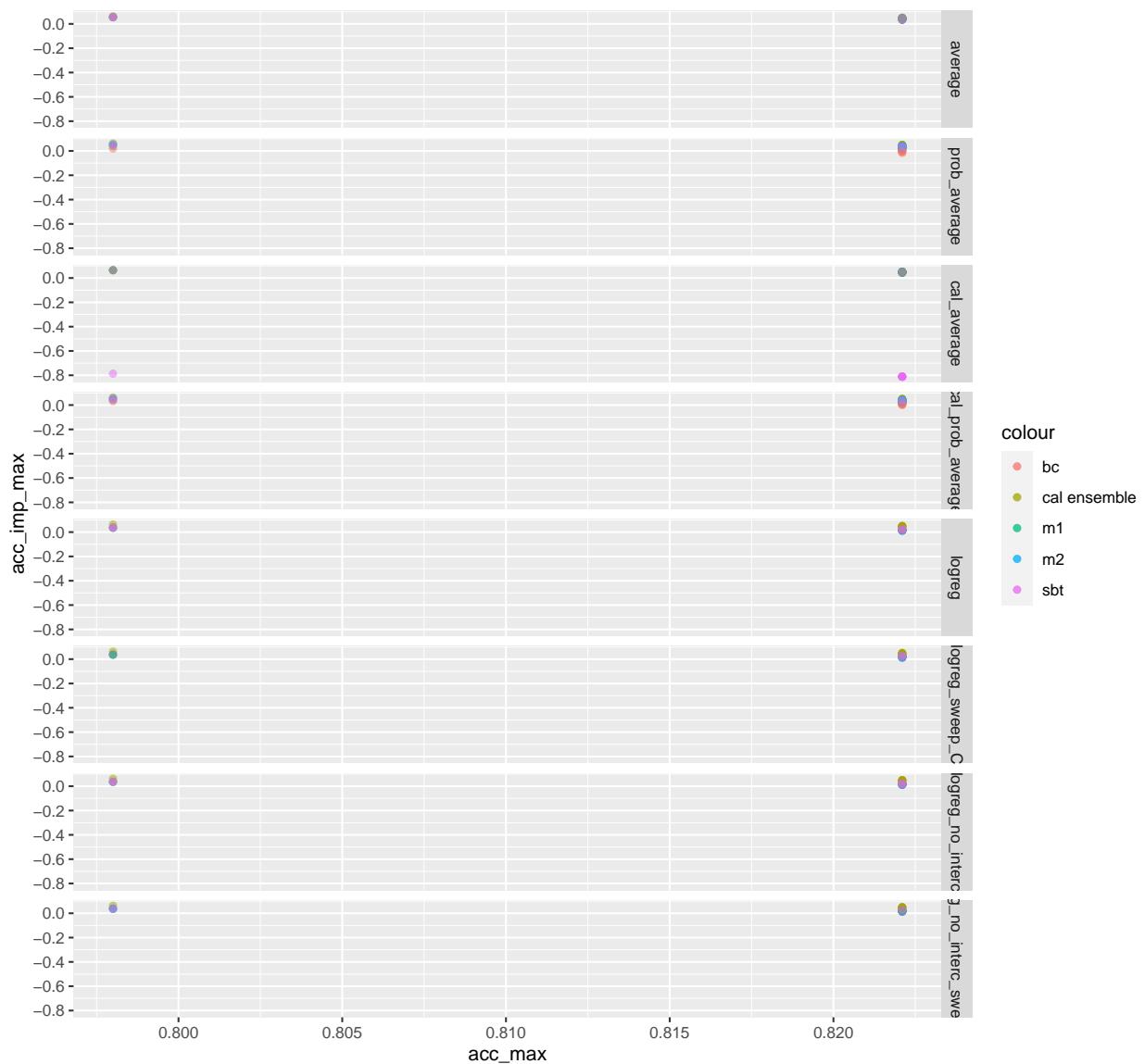
Accuracy improvement of ensemble over  
the best of networks vs acc\_min.  
Ensemble size 7



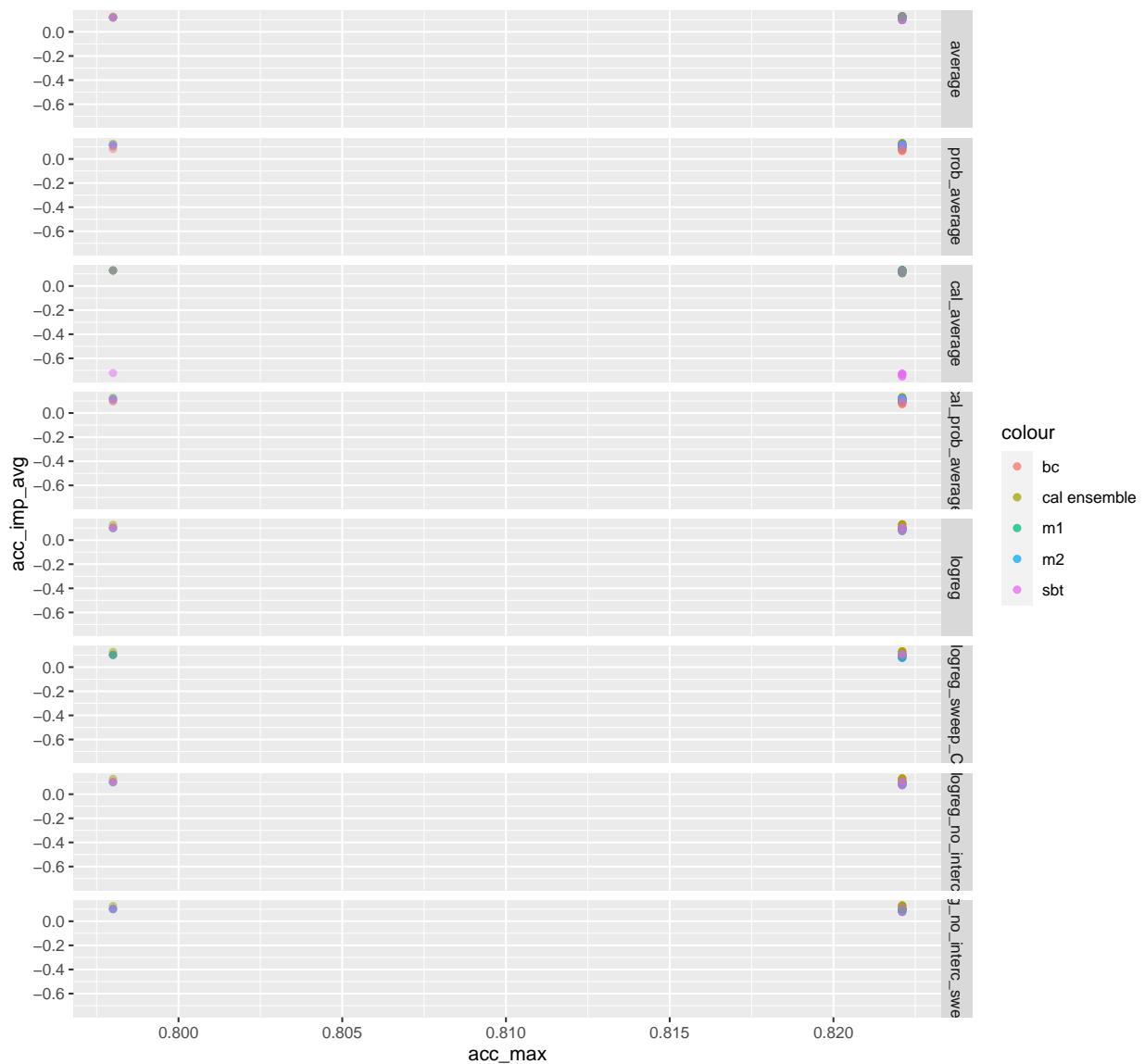
Accuracy improvement of ensemble over  
the average of networks vs acc\_min.  
Ensemble size 7



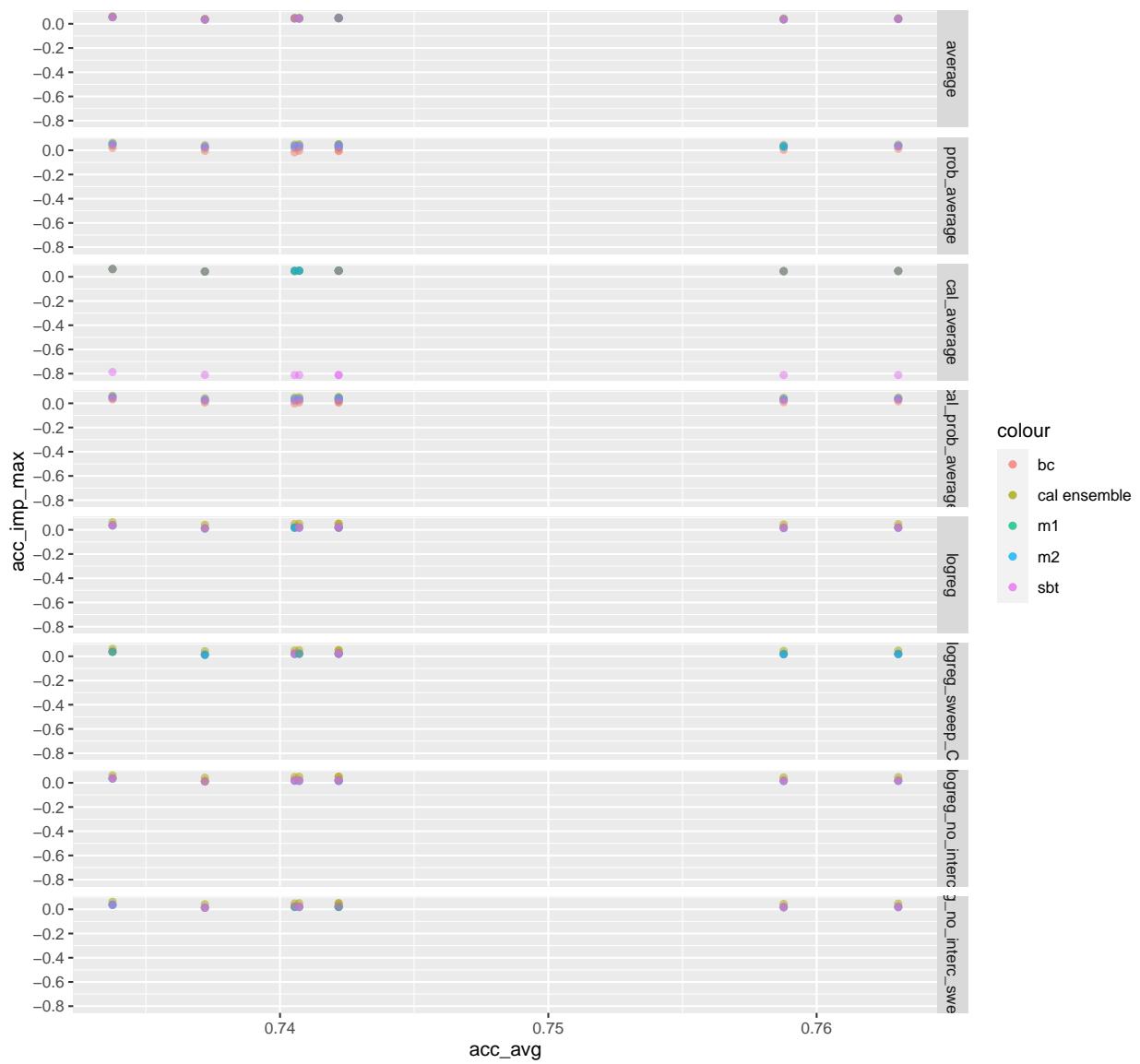
Accuracy improvement of ensemble over  
the best of networks vs acc\_max.  
Ensemble size 7



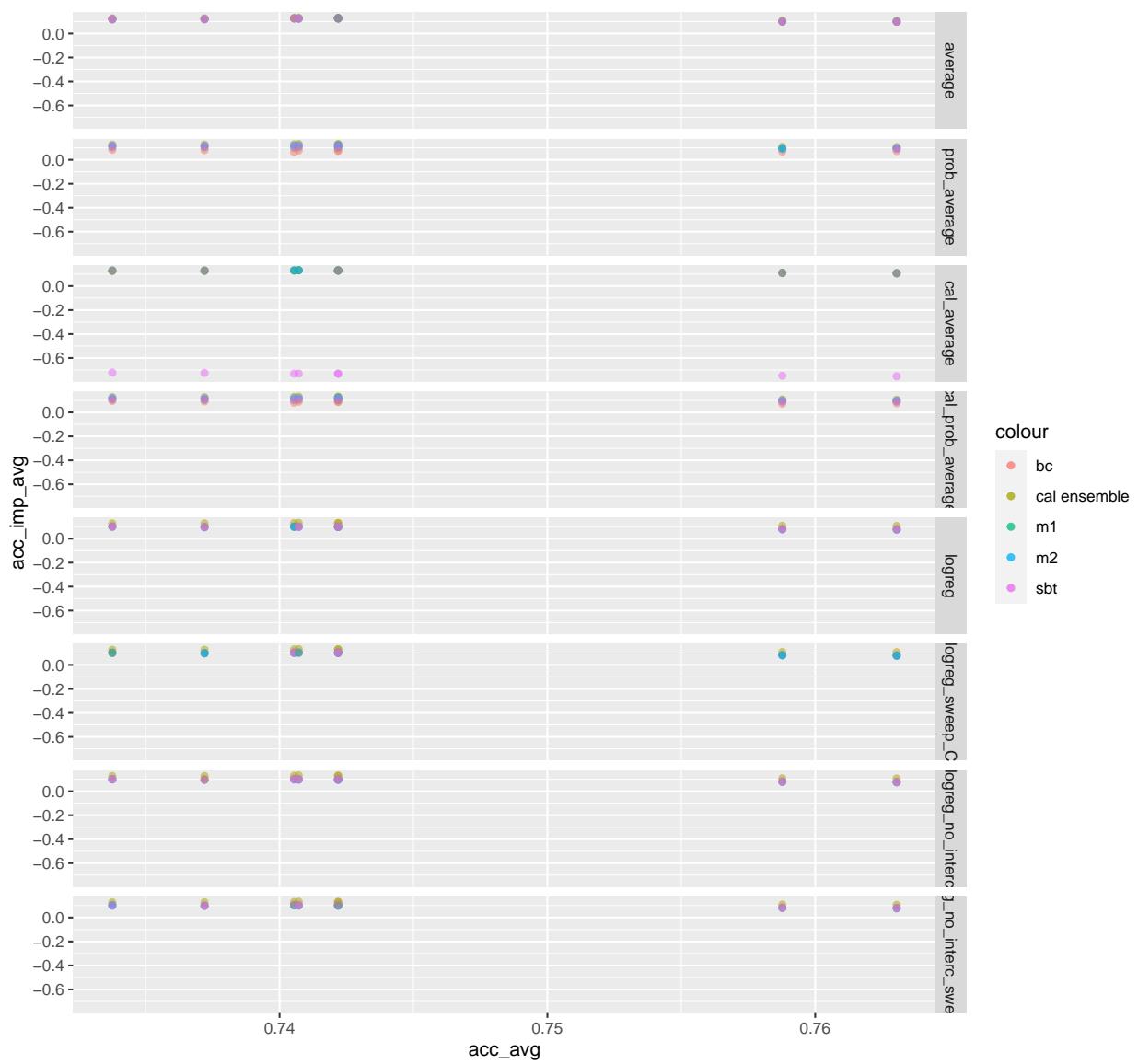
Accuracy improvement of ensemble over  
the average of networks vs acc\_max.  
Ensemble size 7



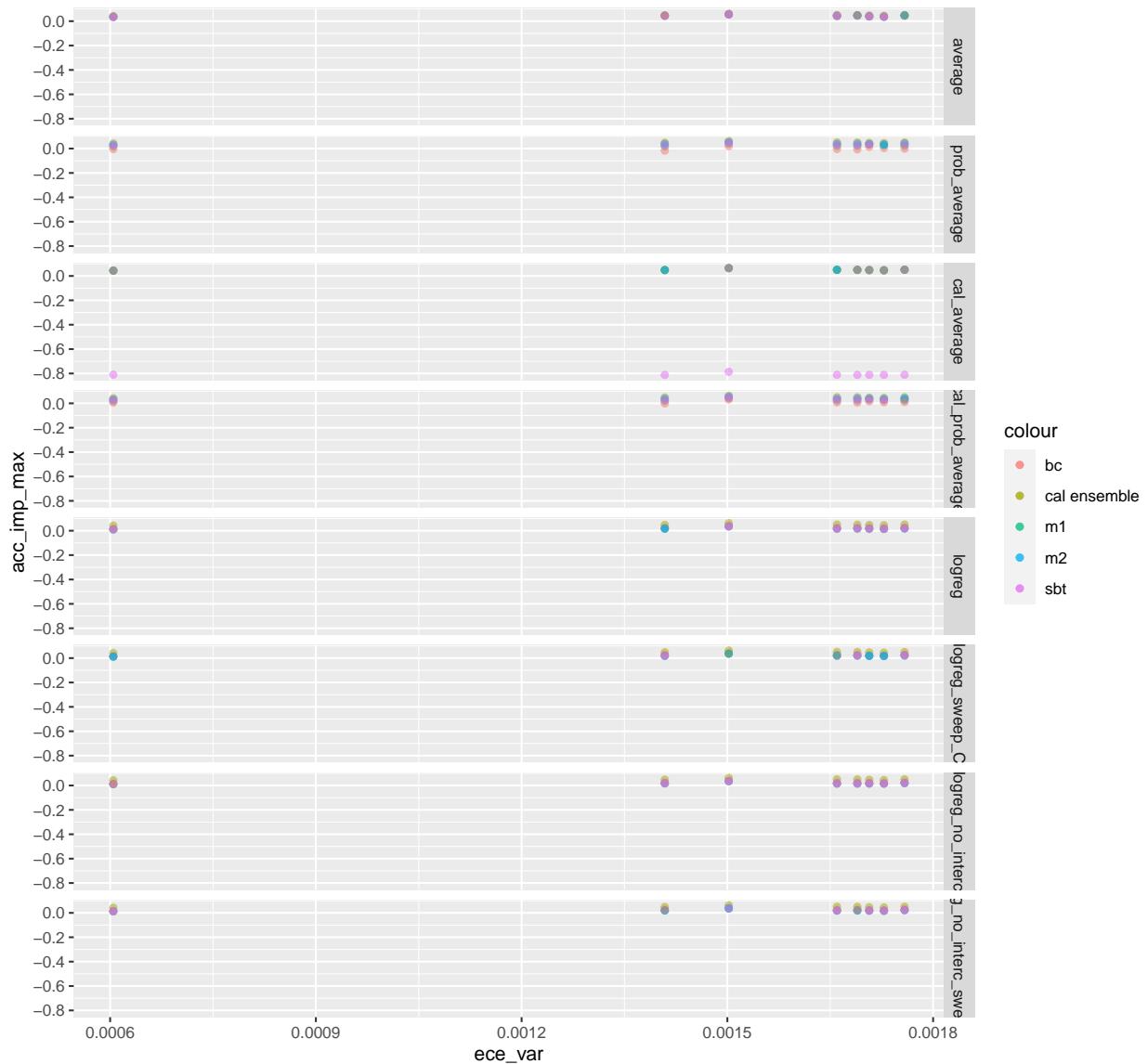
Accuracy improvement of ensemble over  
the best of networks vs acc\_avg.  
Ensemble size 7



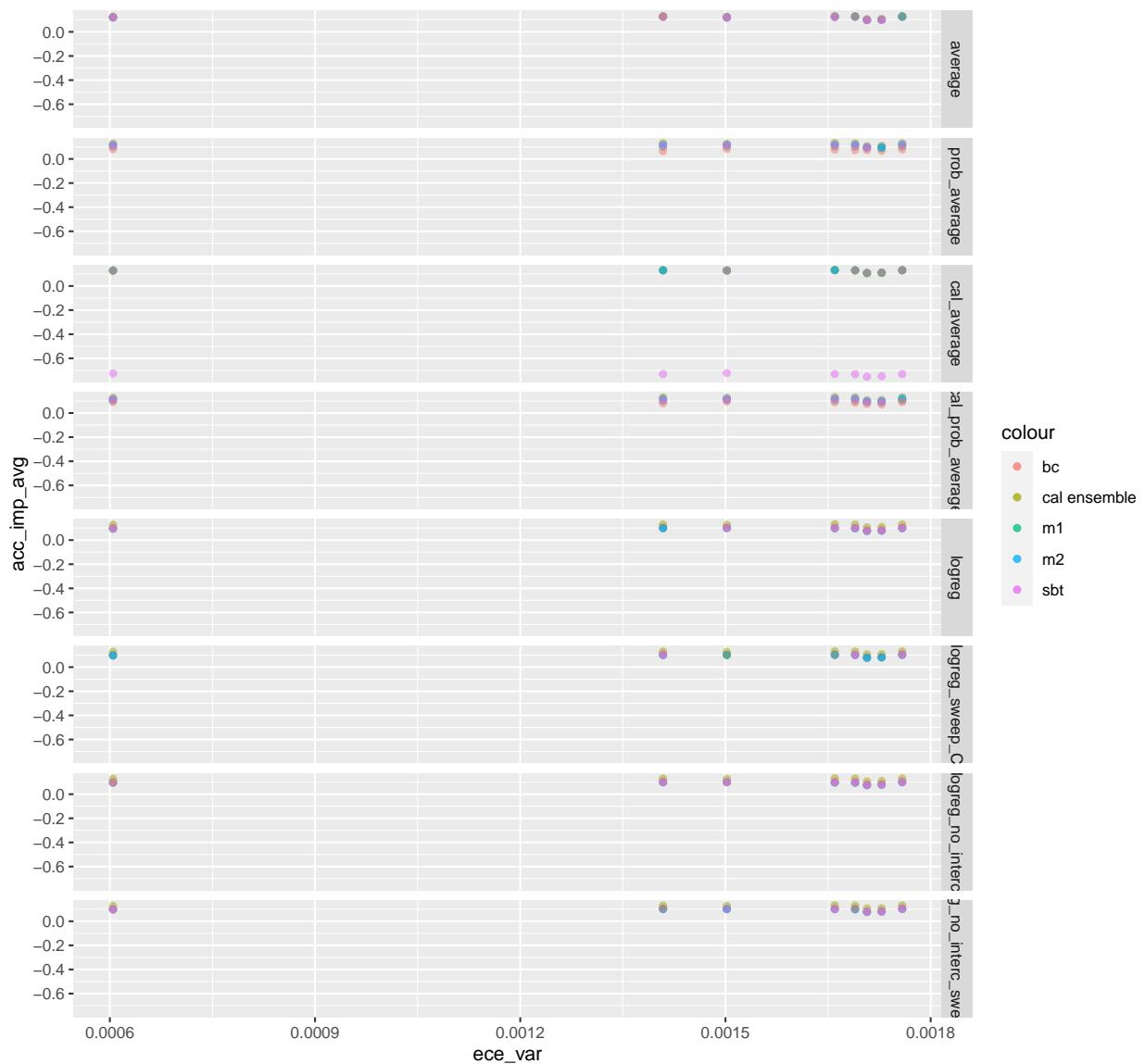
Accuracy improvement of ensemble over  
the average of networks vs acc\_avg.  
Ensemble size 7



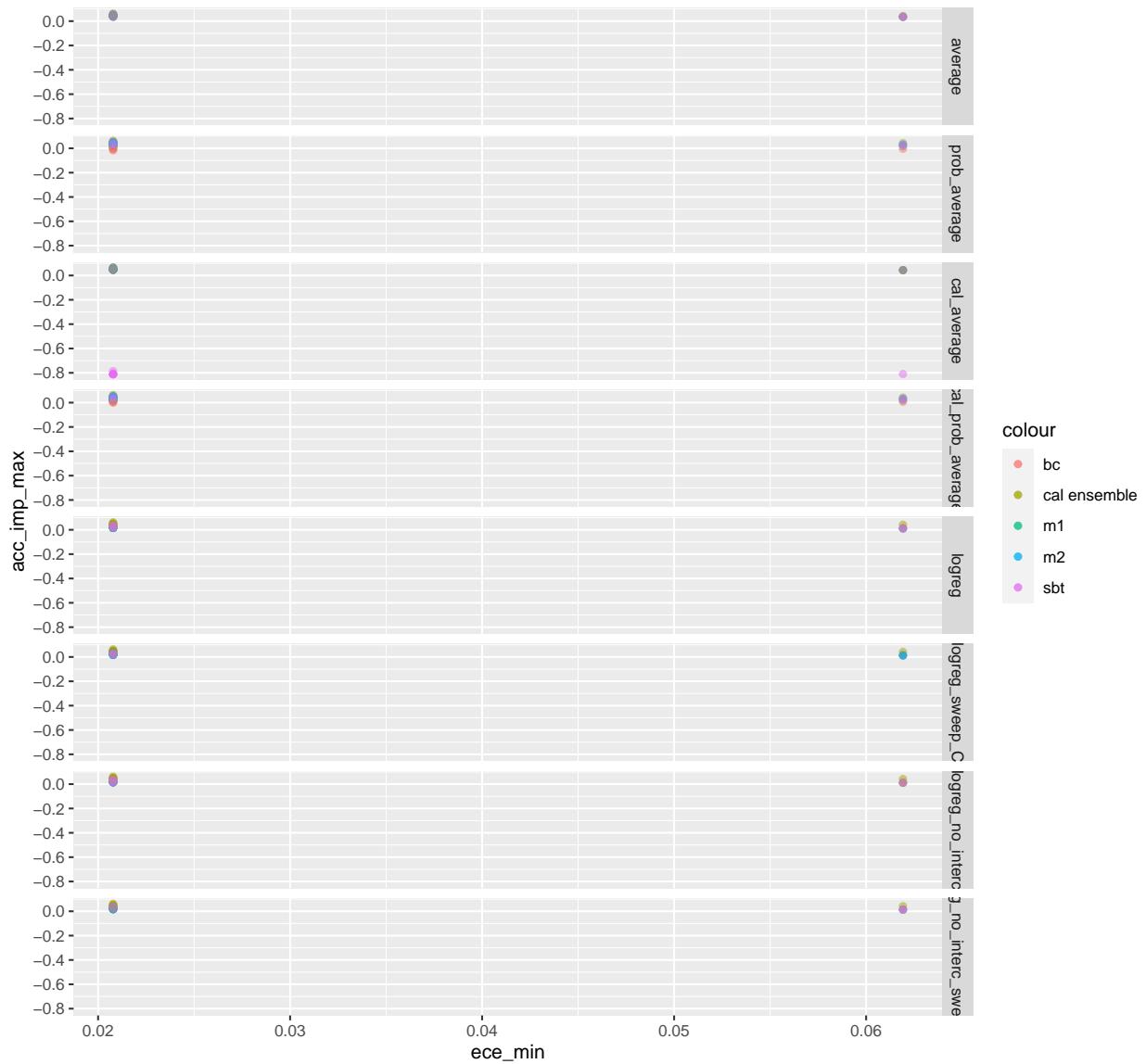
Accuracy improvement of ensemble over  
the best of networks vs ece\_var.  
Ensemble size 7



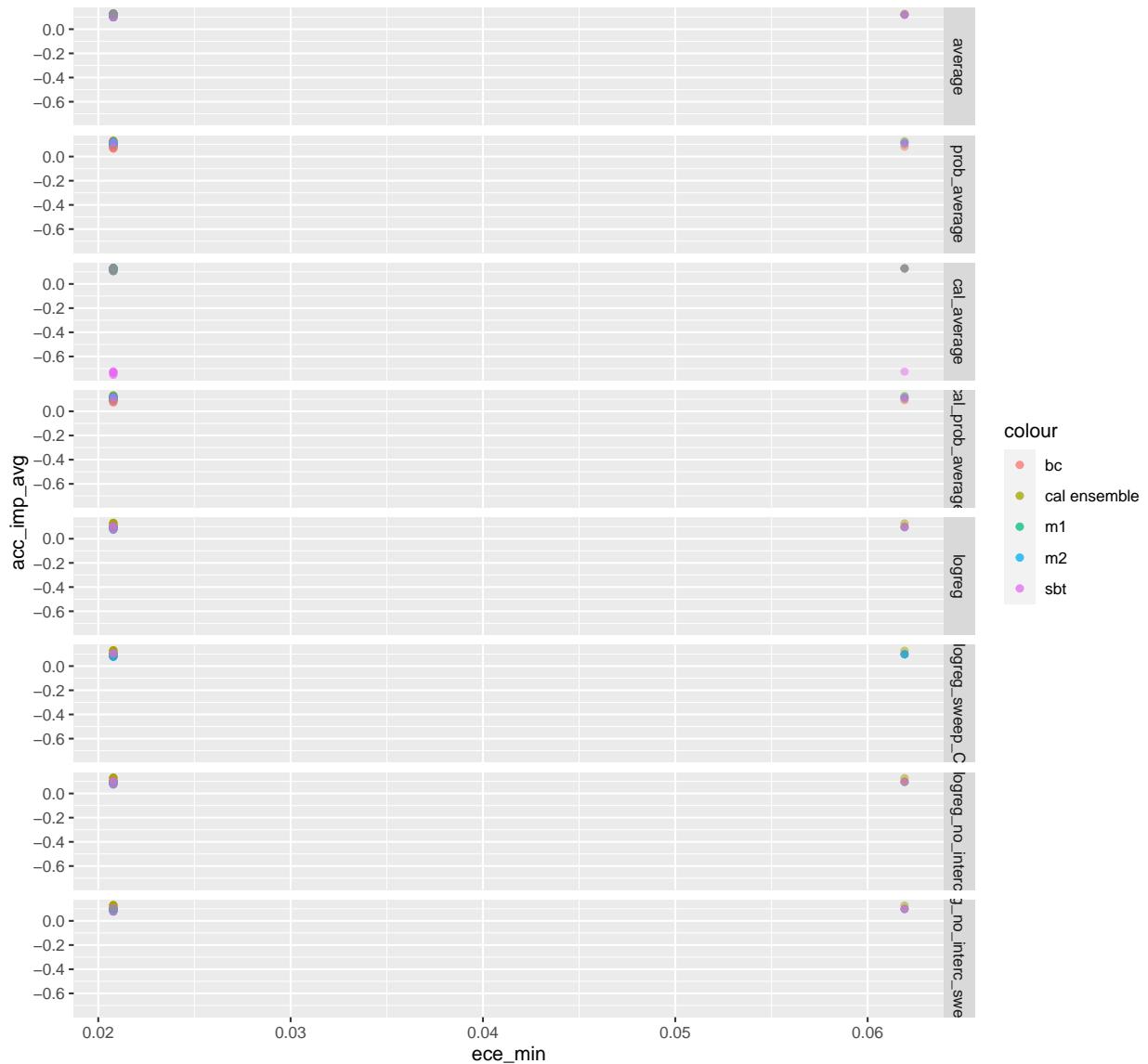
Accuracy improvement of ensemble over  
the average of networks vs ece\_var.  
Ensemble size 7



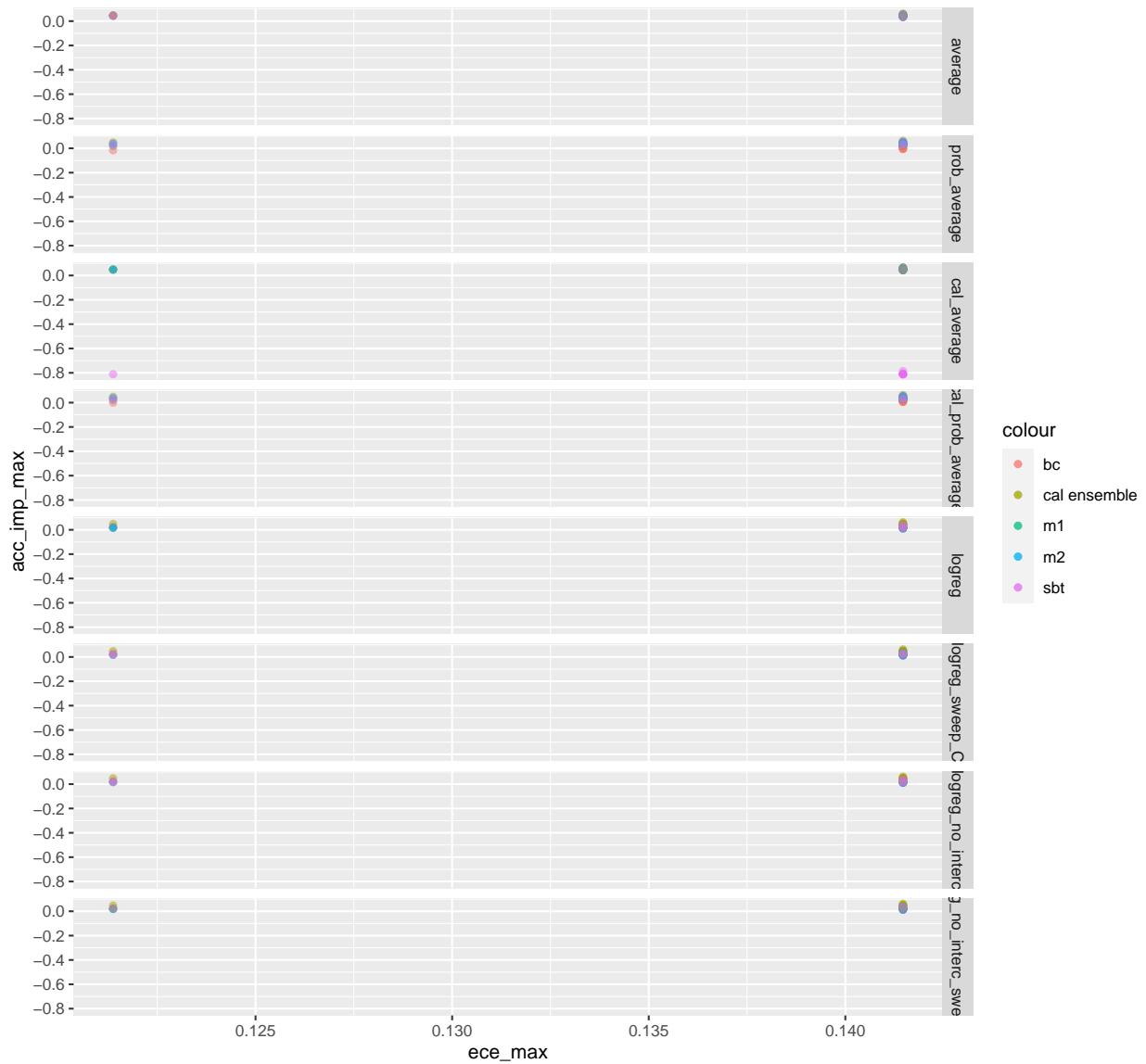
Accuracy improvement of ensemble over  
the best of networks vs ece\_min.  
Ensemble size 7



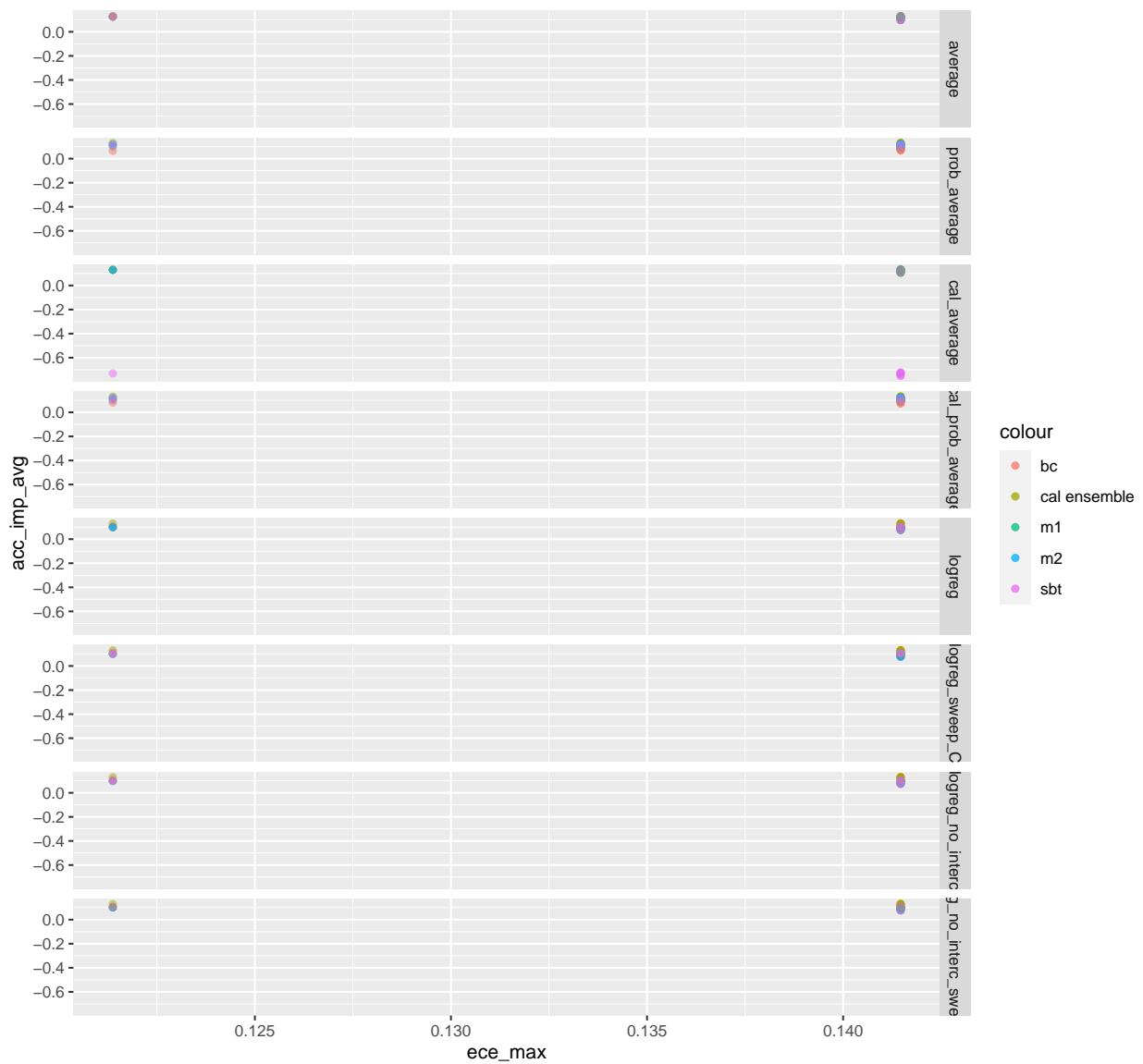
Accuracy improvement of ensemble over  
the average of networks vs ece\_min.  
Ensemble size 7



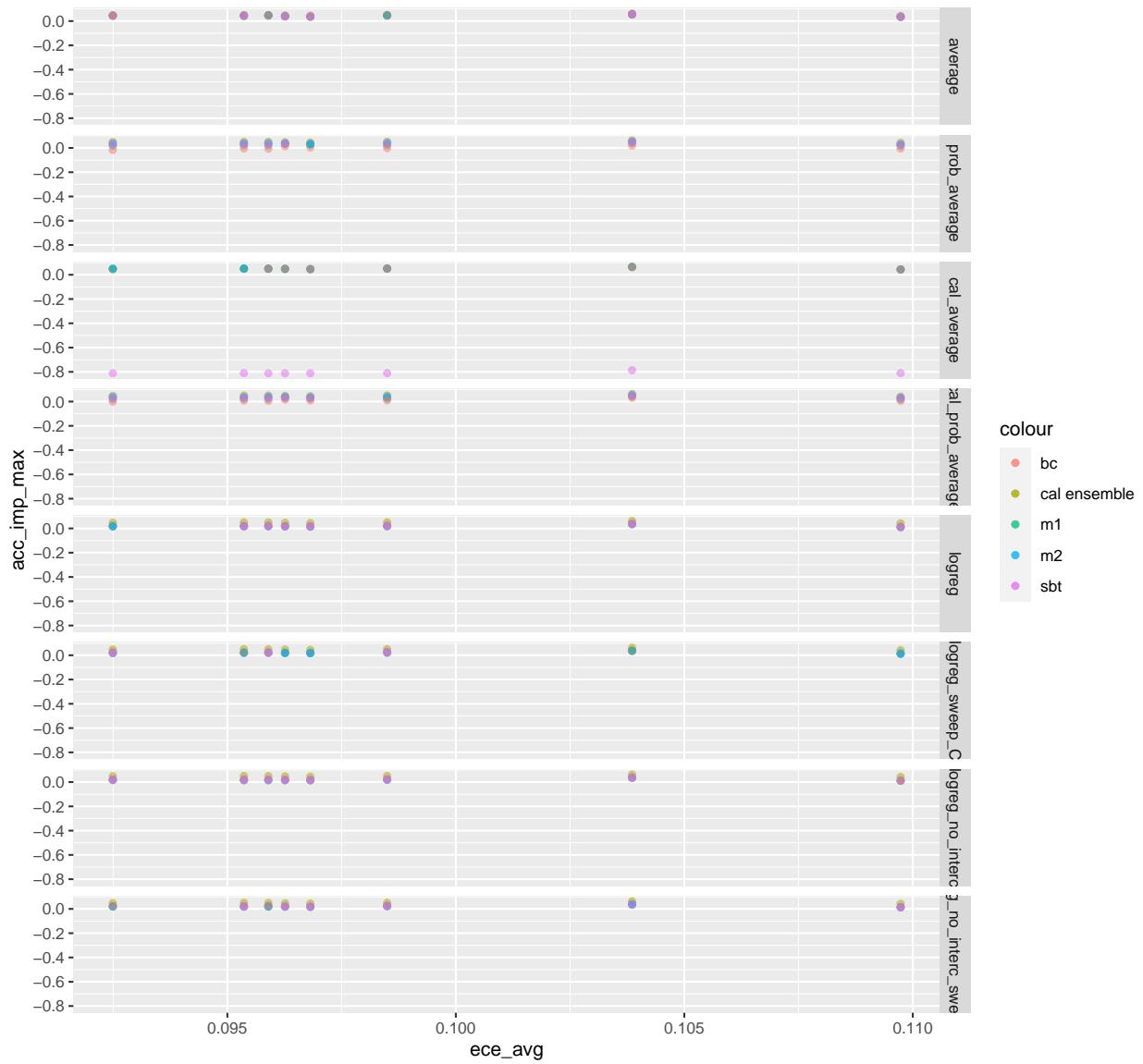
Accuracy improvement of ensemble over  
the best of networks vs ece\_max.  
Ensemble size 7



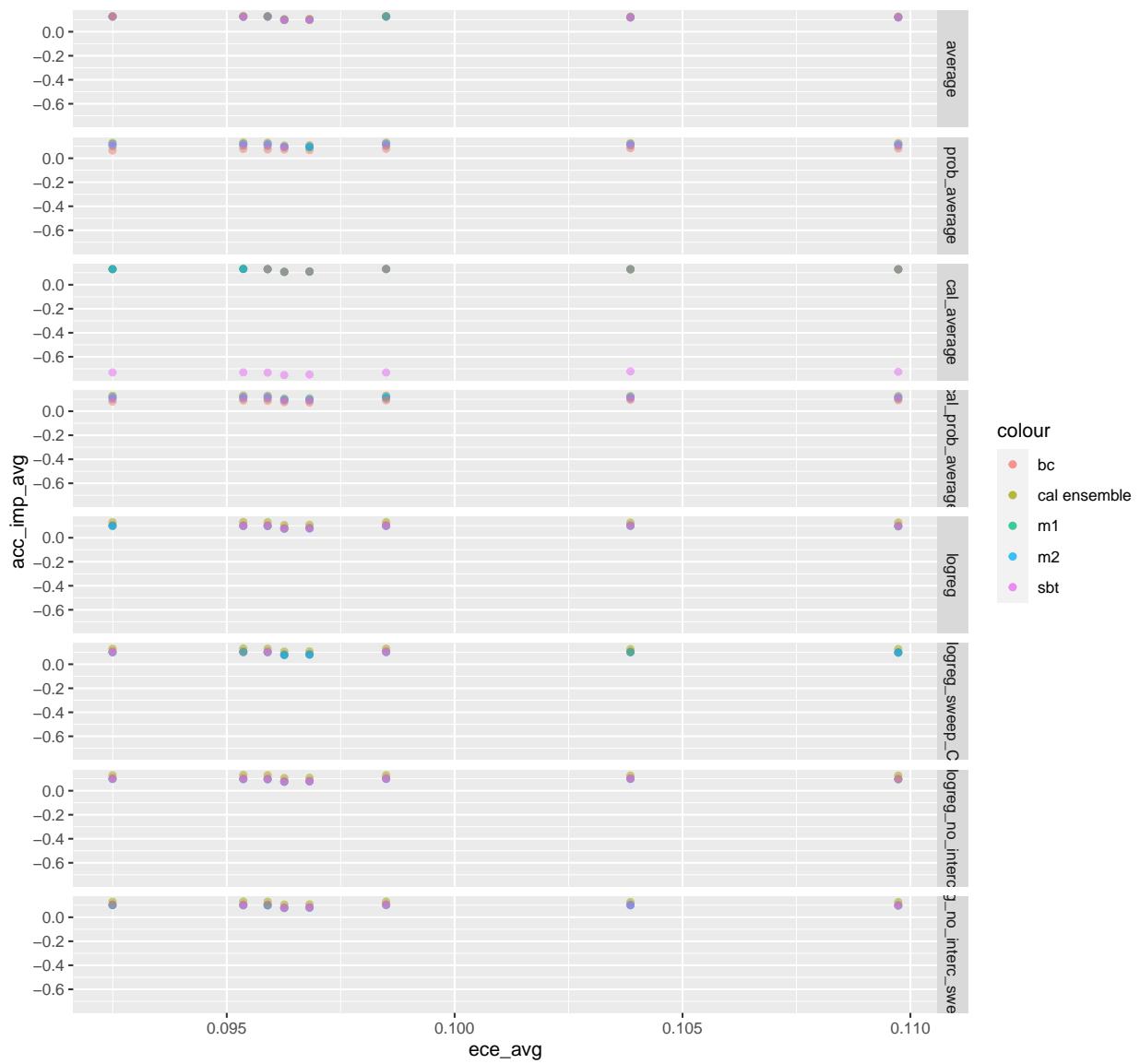
Accuracy improvement of ensemble over  
the average of networks vs ece\_max.  
Ensemble size 7



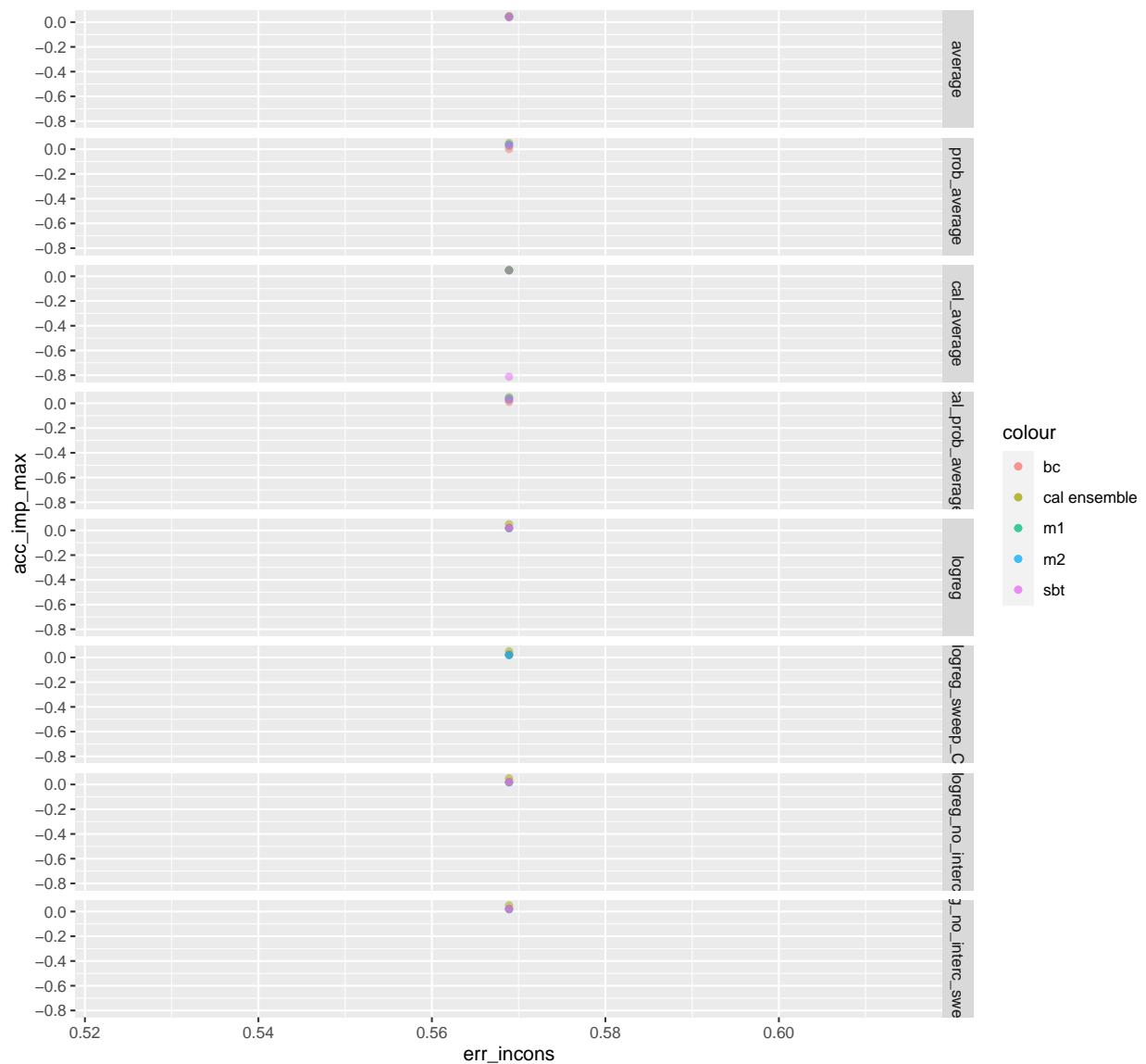
Accuracy improvement of ensemble over  
the best of networks vs ece\_avg.  
Ensemble size 7



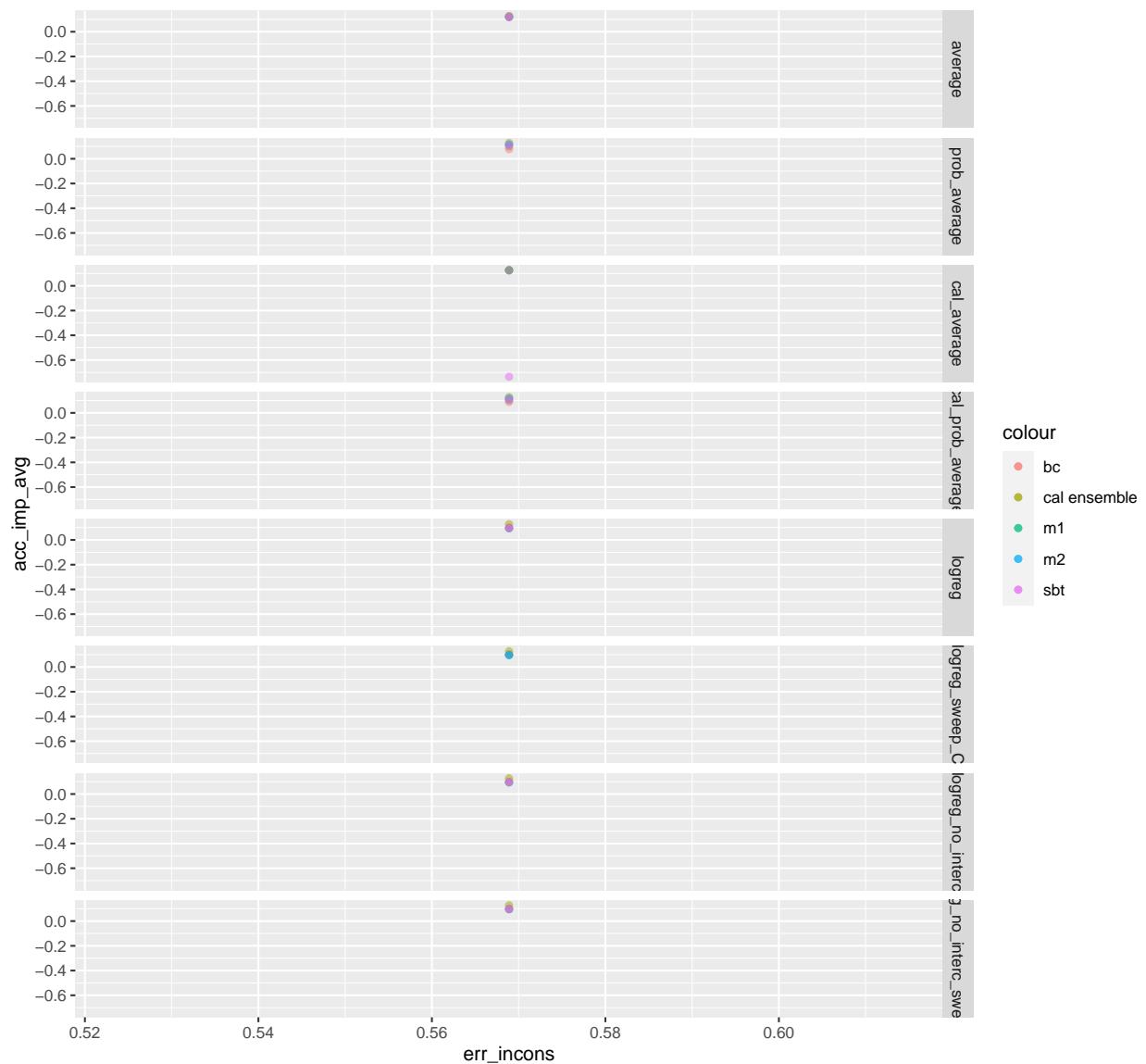
Accuracy improvement of ensemble over  
the average of networks vs ece\_avg.  
Ensemble size 7



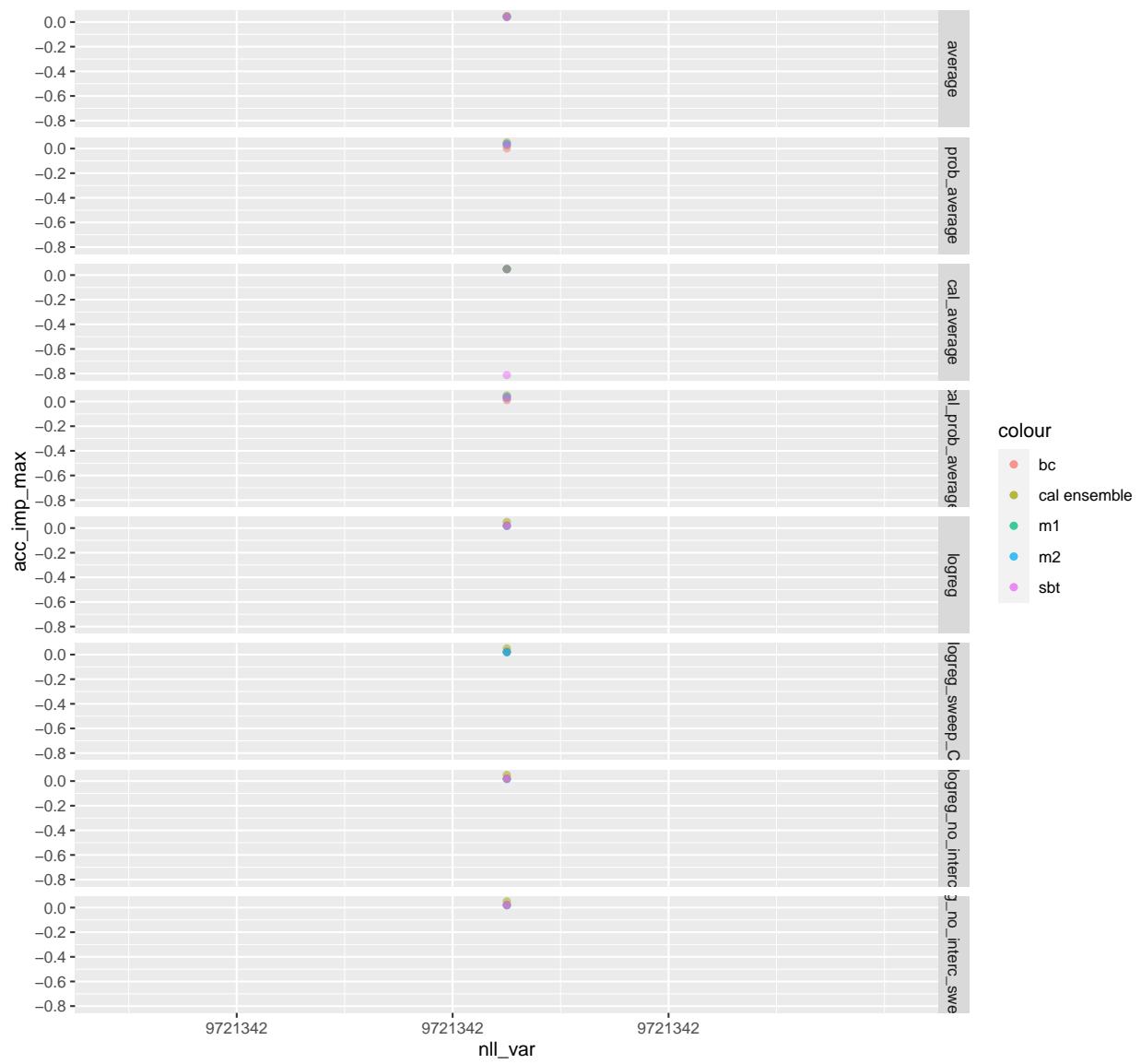
Accuracy improvement of ensemble over  
the best of networks vs err\_incons.  
Ensemble size 8



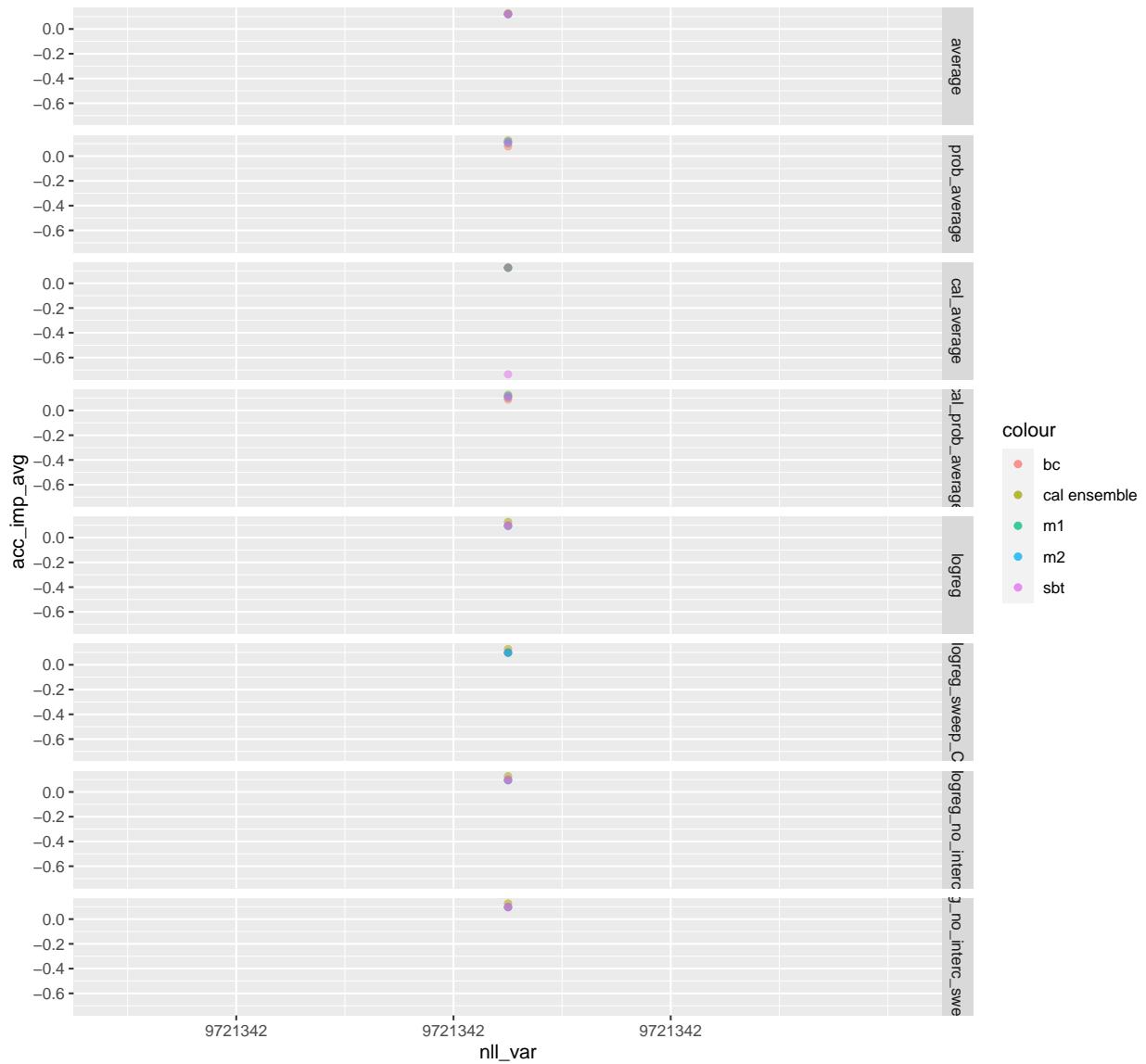
Accuracy improvement of ensemble over  
the average of networks vs err\_incons.  
Ensemble size 8



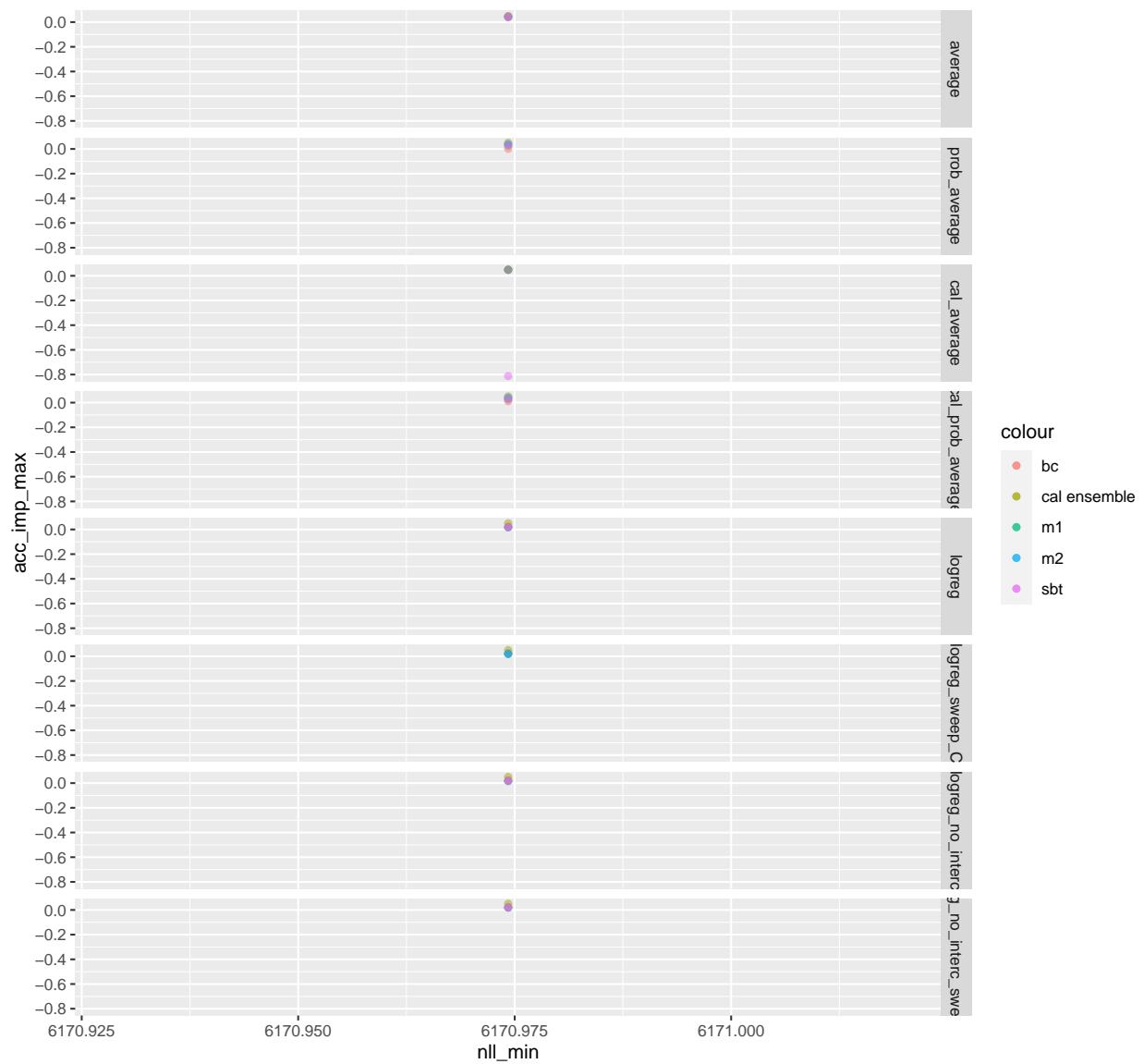
Accuracy improvement of ensemble over  
the best of networks vs nll\_var.  
Ensemble size 8



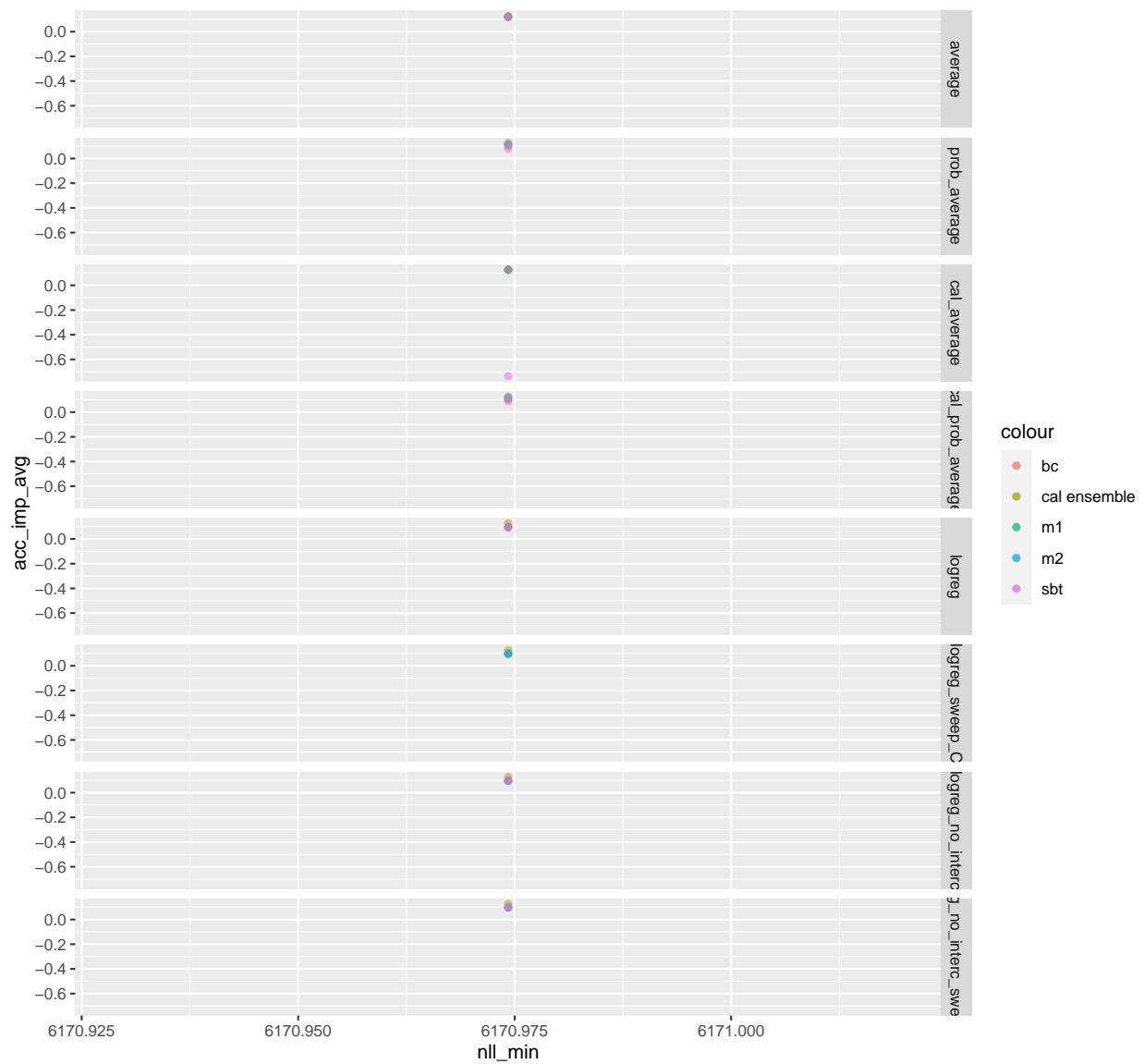
Accuracy improvement of ensemble over  
the average of networks vs nll\_var.  
Ensemble size 8



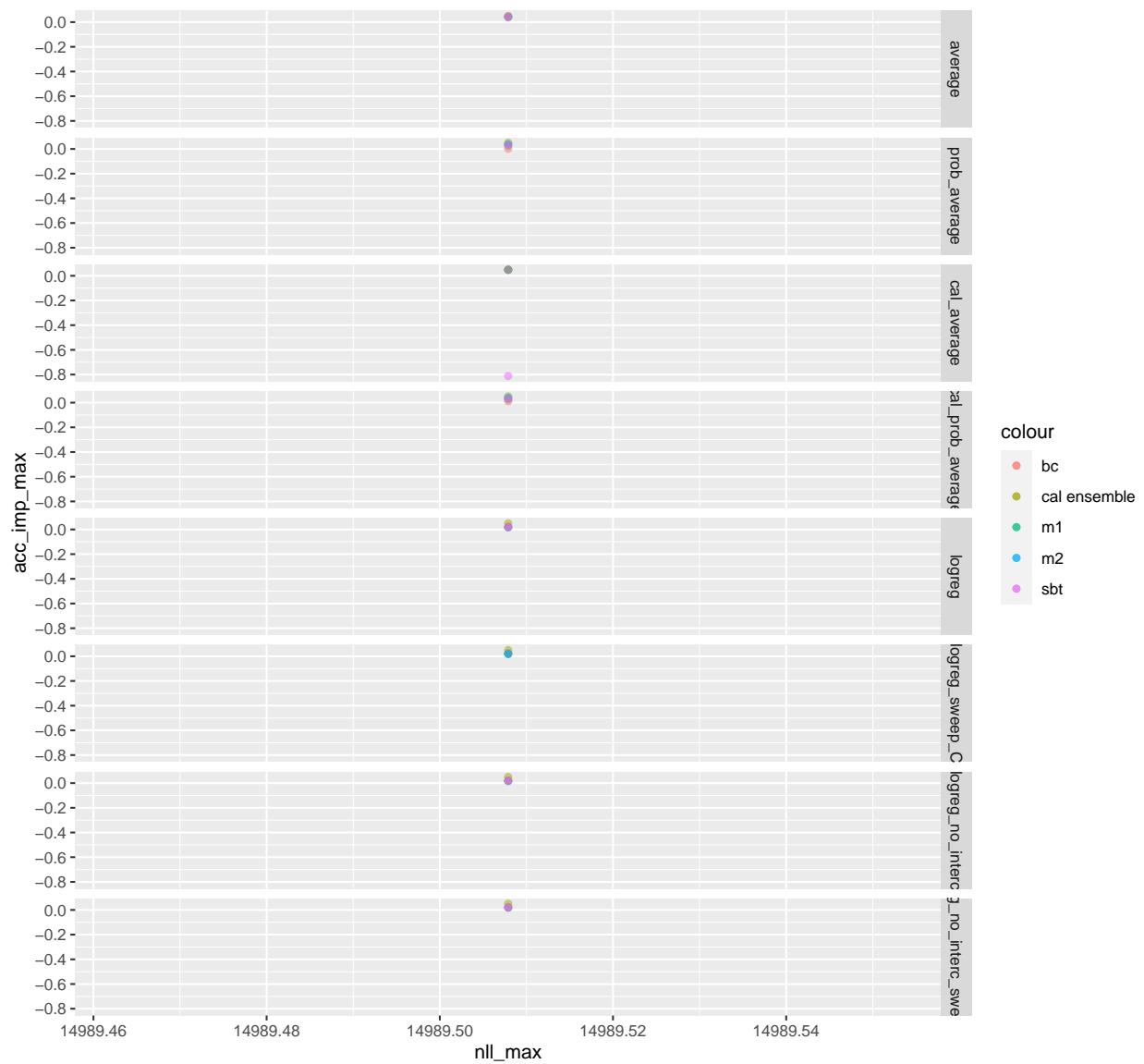
Accuracy improvement of ensemble over  
the best of networks vs nll\_min.  
Ensemble size 8



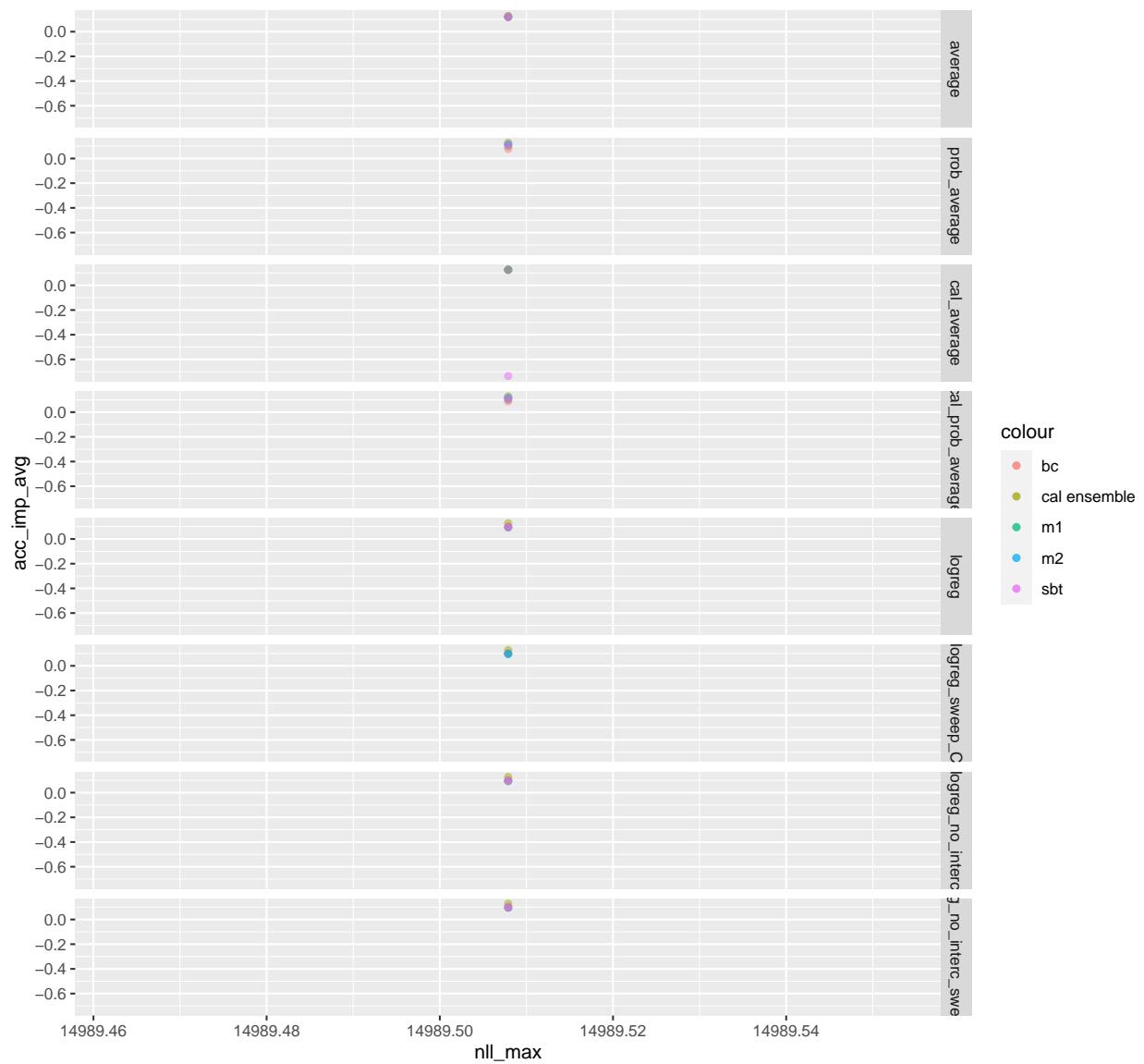
Accuracy improvement of ensemble over  
the average of networks vs nll\_min.  
Ensemble size 8



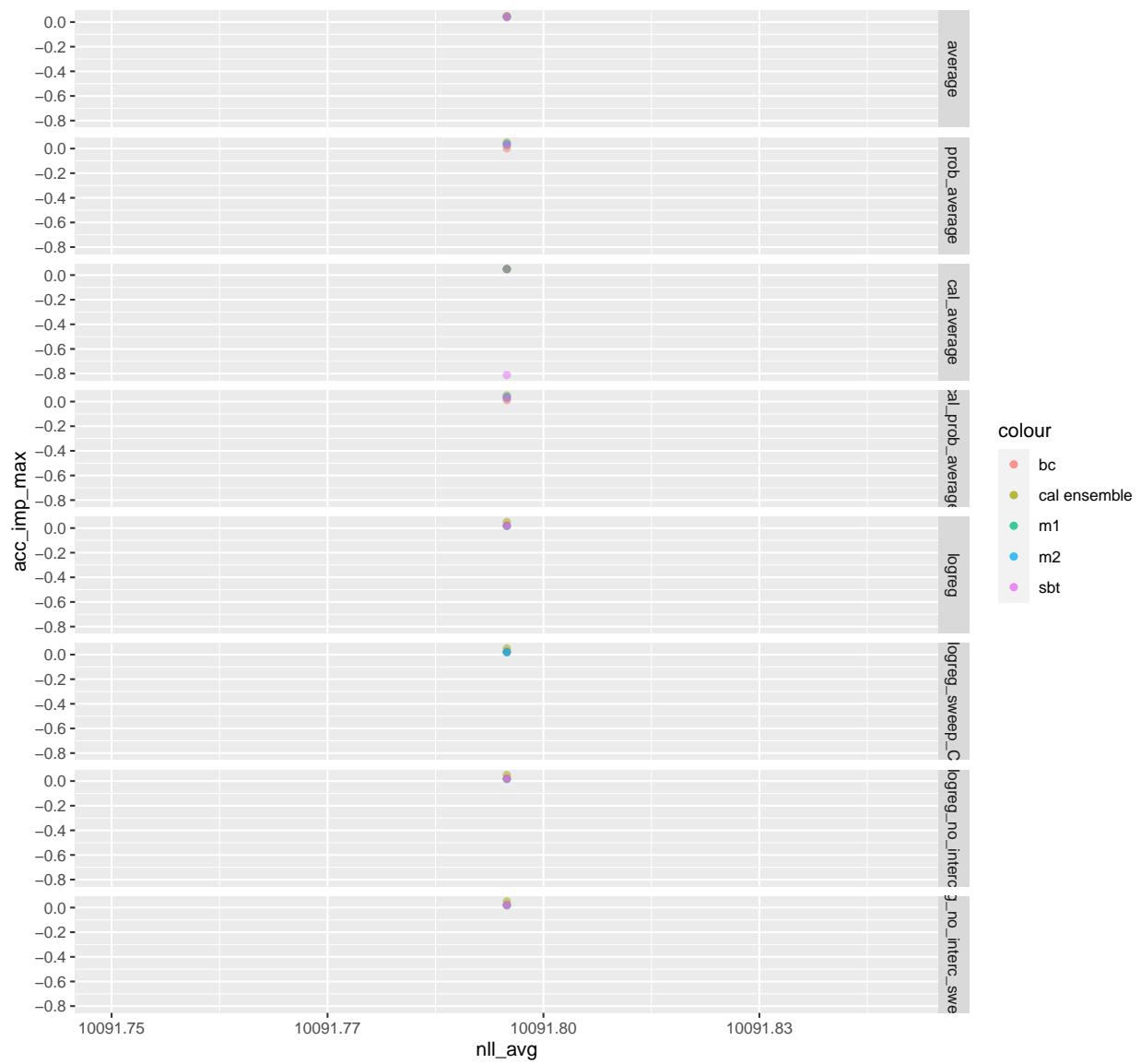
Accuracy improvement of ensemble over  
the best of networks vs nll\_max.  
Ensemble size 8



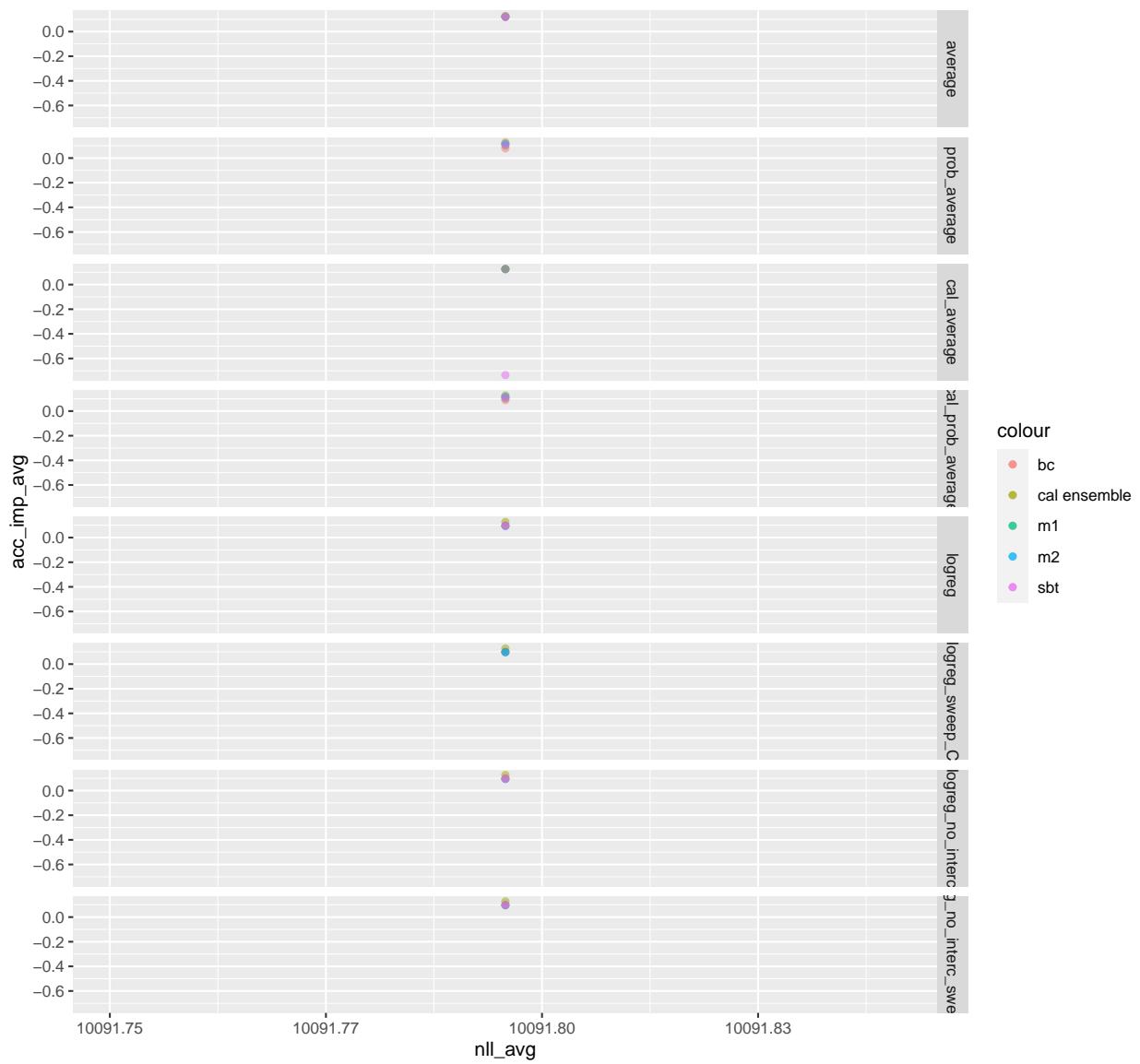
Accuracy improvement of ensemble over  
the average of networks vs nll\_max.  
Ensemble size 8



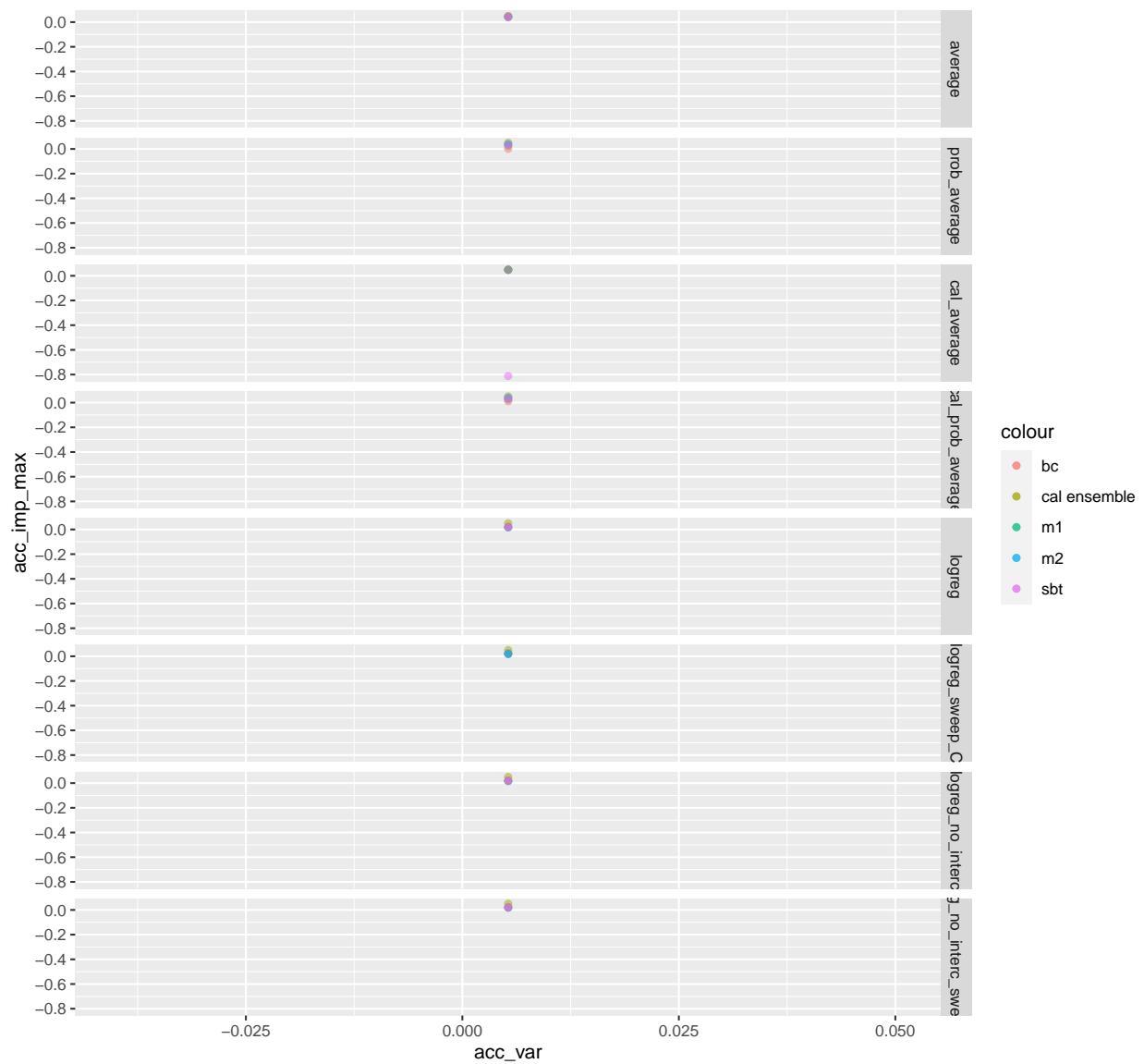
Accuracy improvement of ensemble over  
the best of networks vs nll\_avg.  
Ensemble size 8



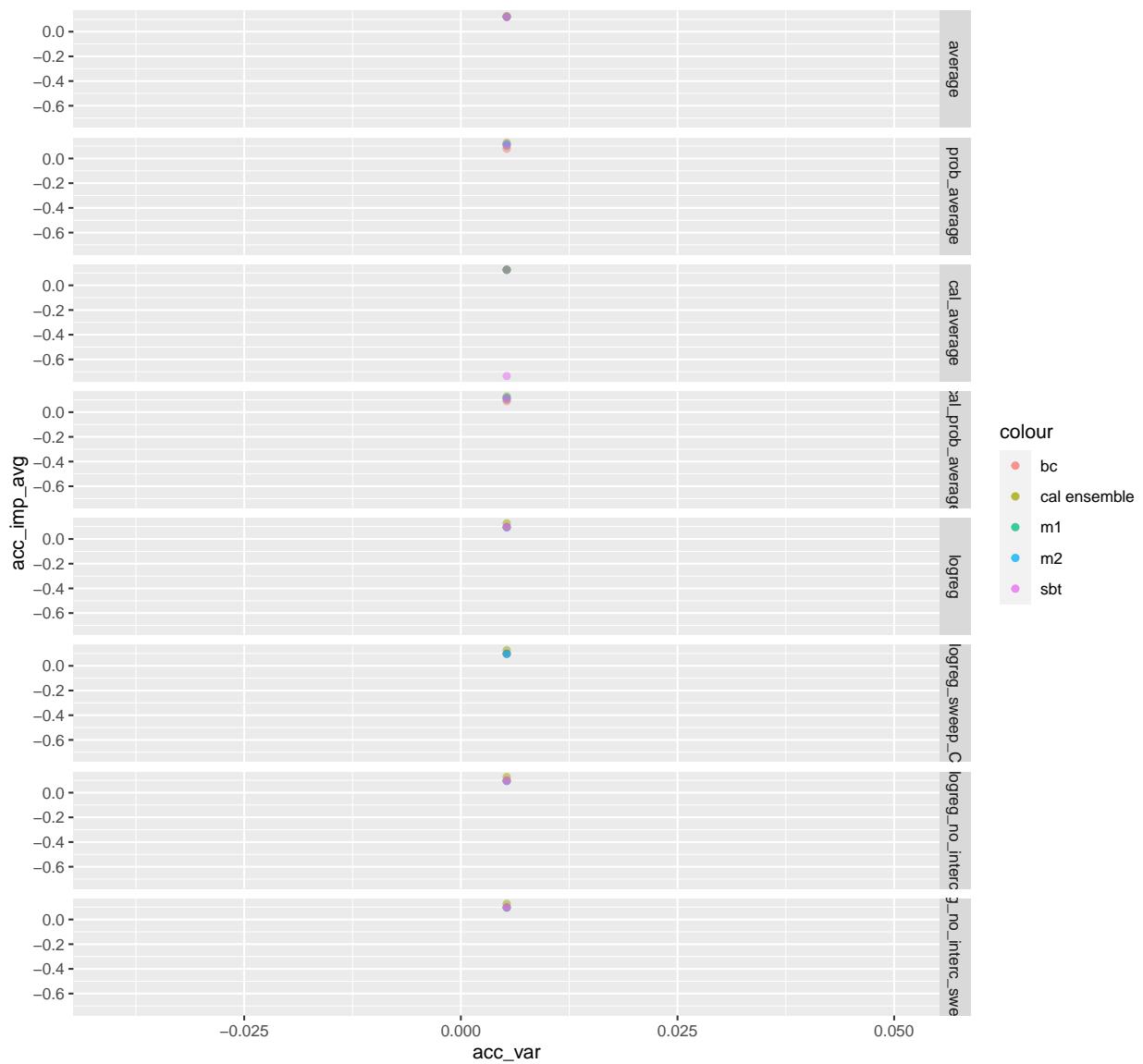
Accuracy improvement of ensemble over  
the average of networks vs nll\_avg.  
Ensemble size 8



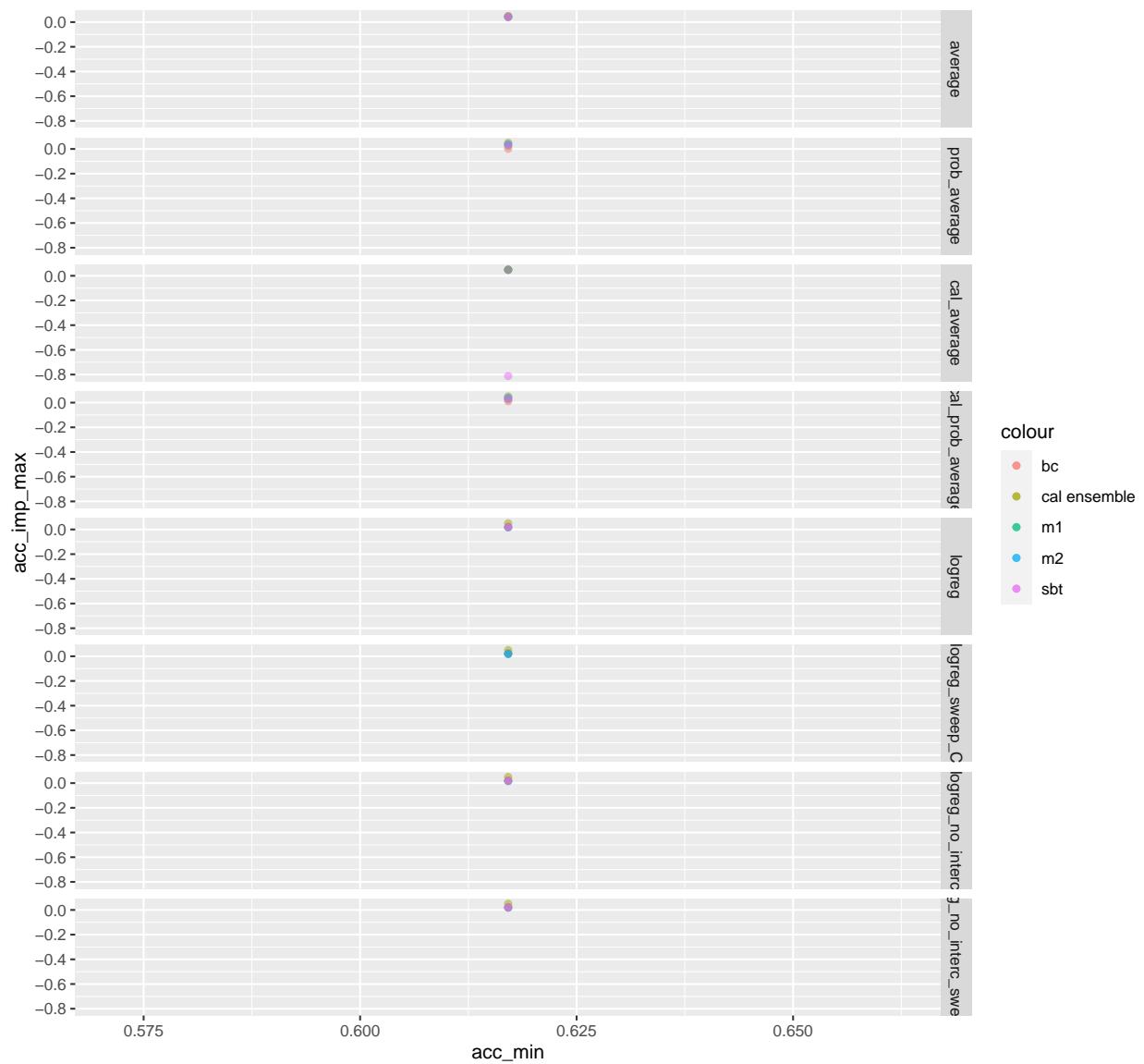
Accuracy improvement of ensemble over  
the best of networks vs acc\_var.  
Ensemble size 8



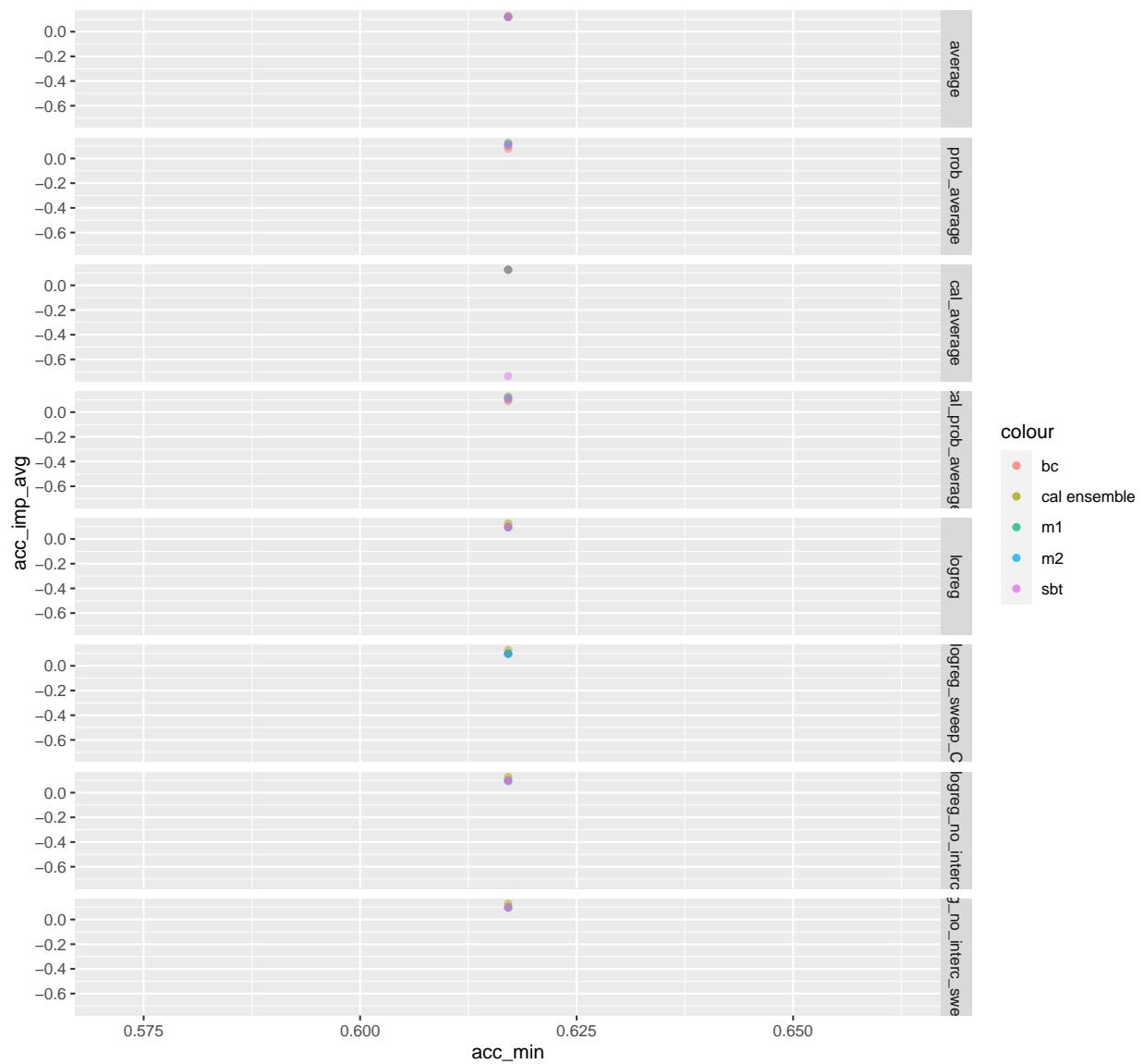
Accuracy improvement of ensemble over  
the average of networks vs acc\_var.  
Ensemble size 8



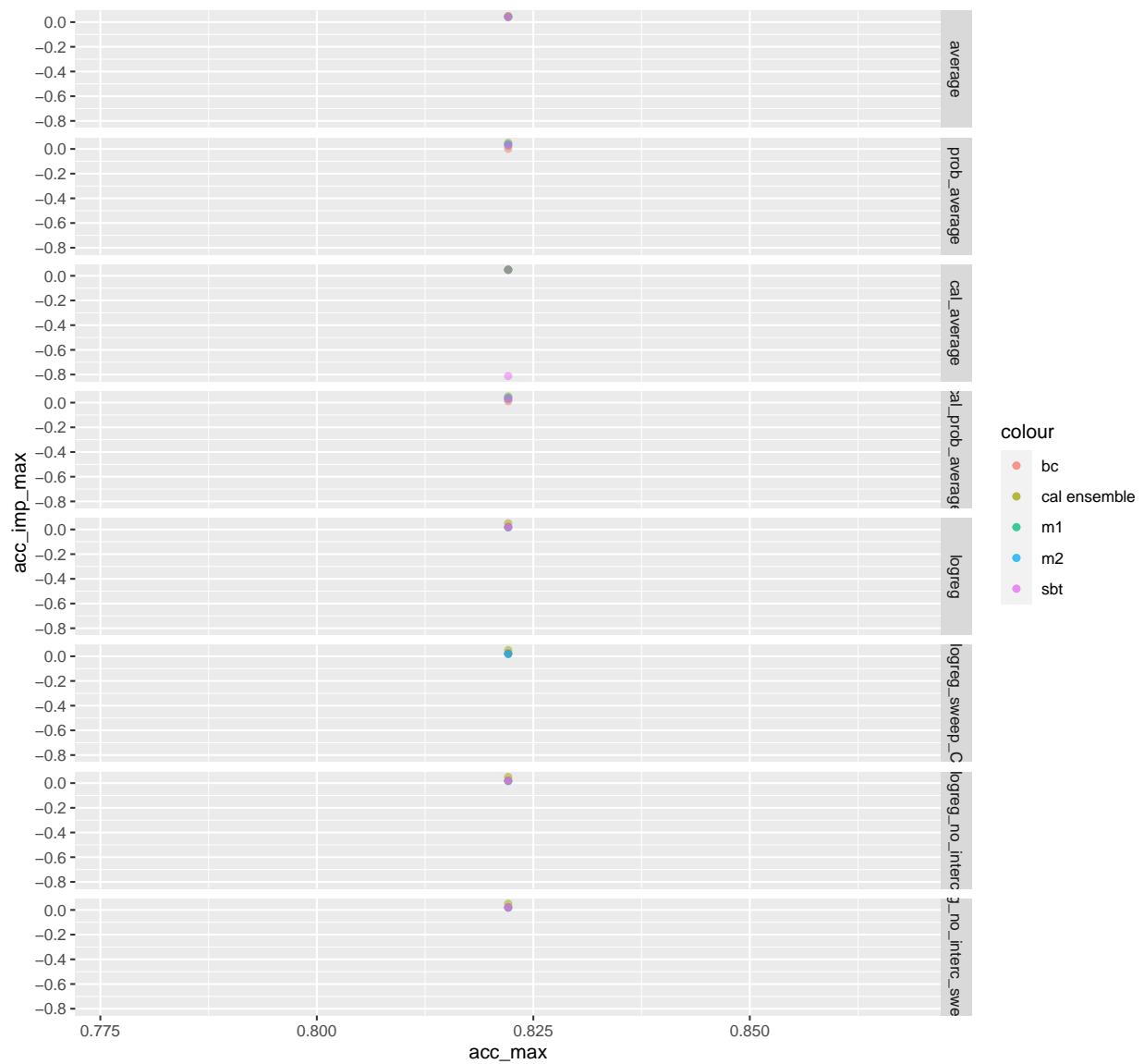
Accuracy improvement of ensemble over  
the best of networks vs acc\_min.  
Ensemble size 8



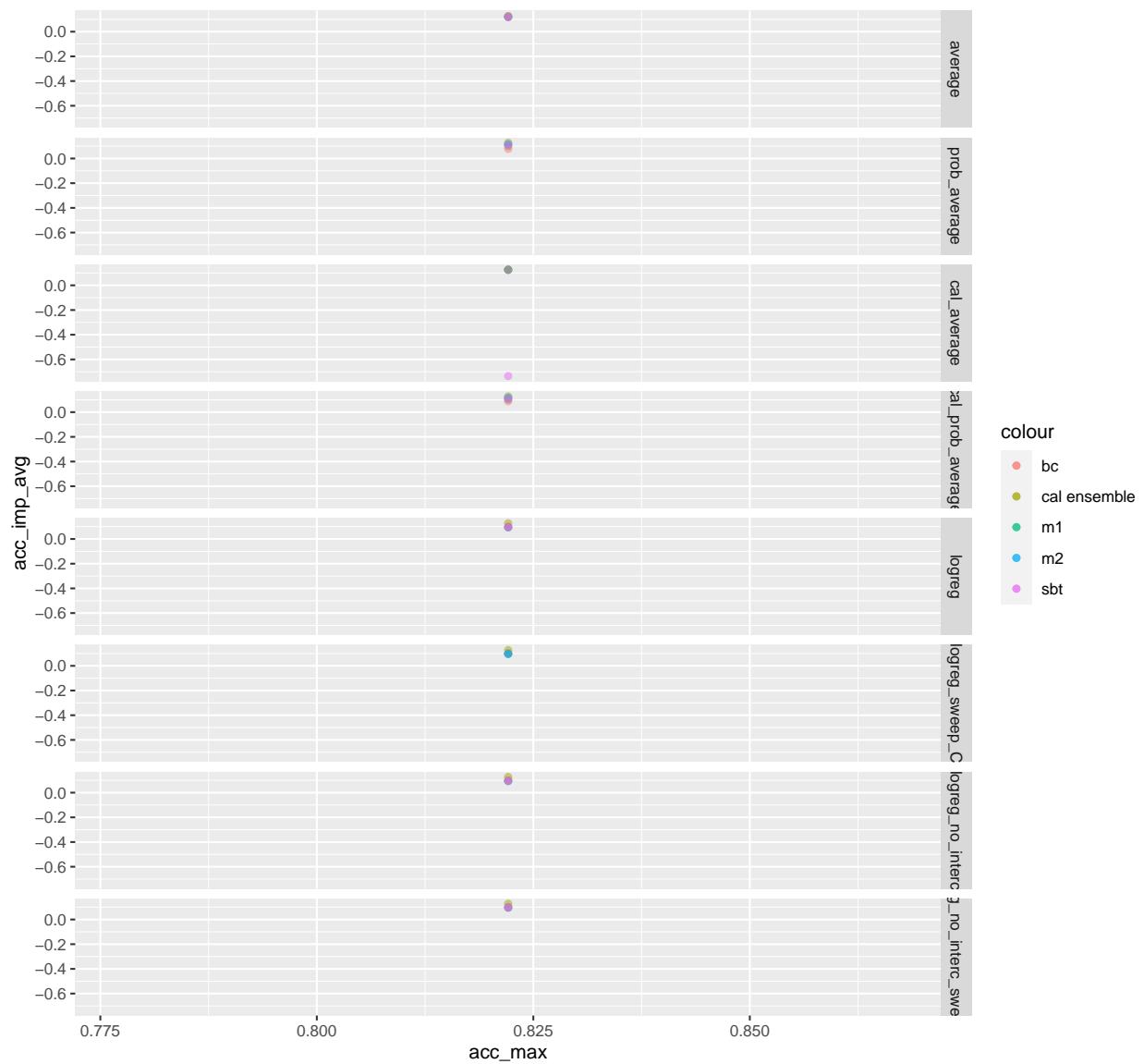
Accuracy improvement of ensemble over  
the average of networks vs acc\_min.  
Ensemble size 8



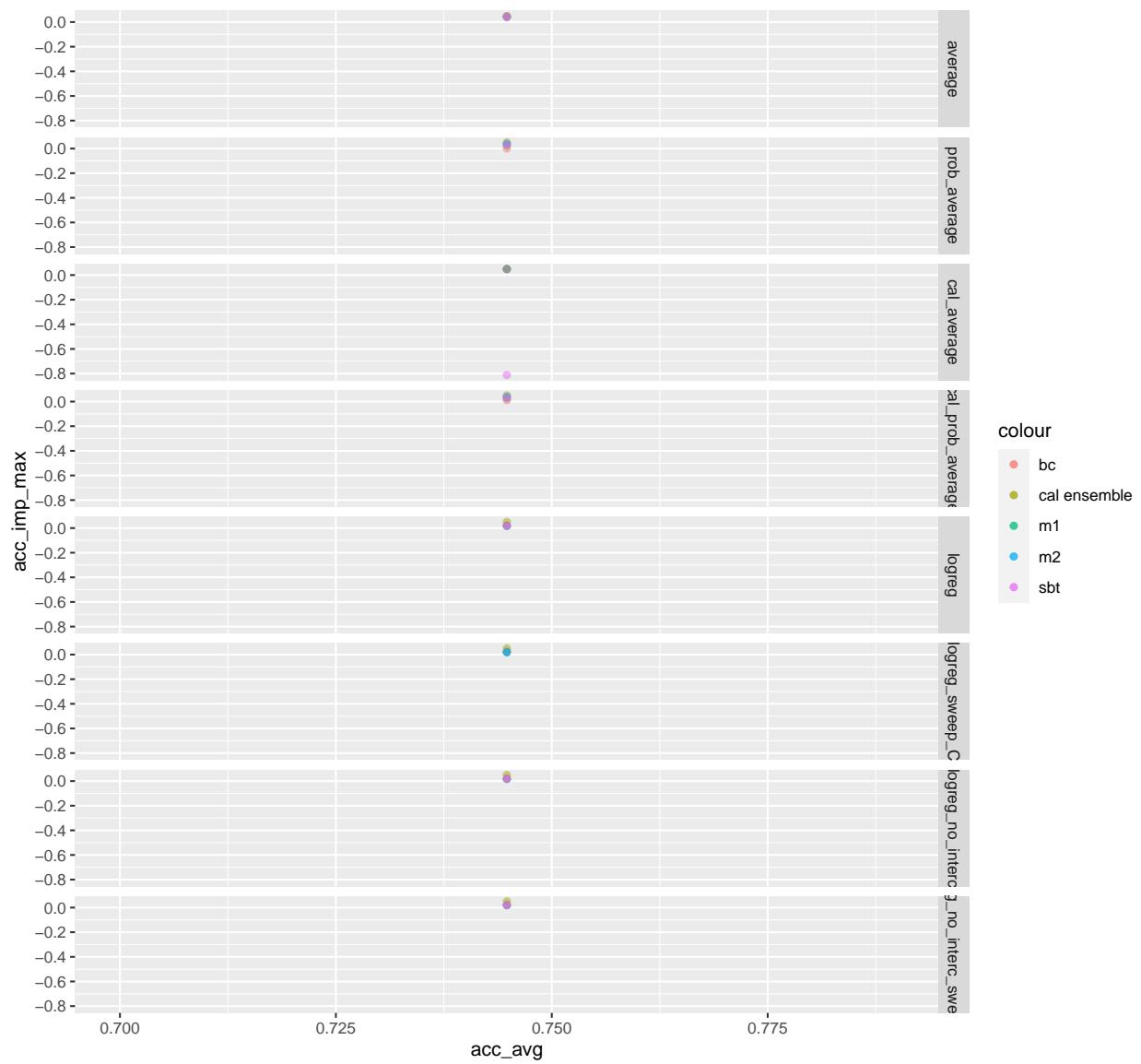
Accuracy improvement of ensemble over  
the best of networks vs acc\_max.  
Ensemble size 8



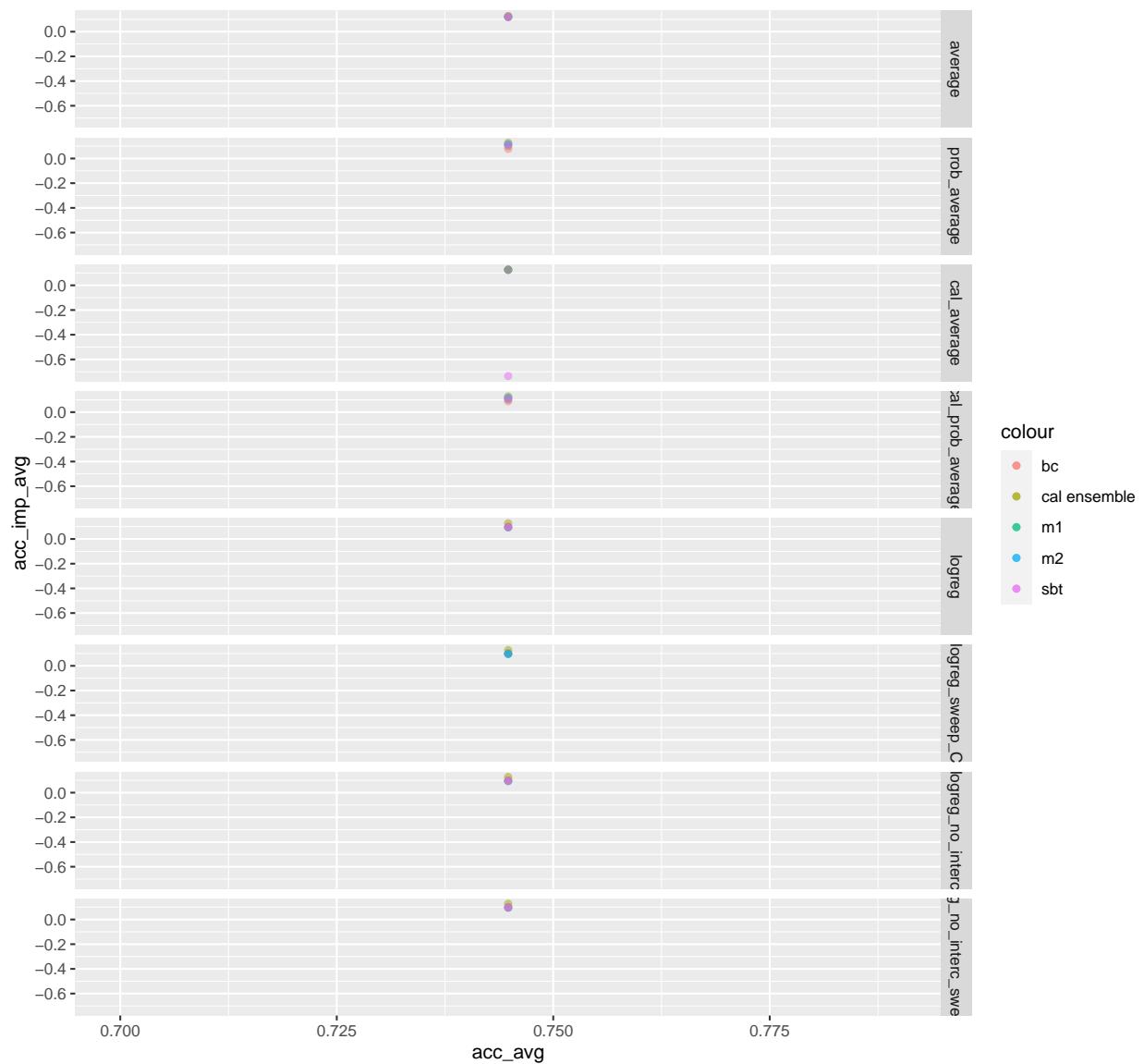
Accuracy improvement of ensemble over  
the average of networks vs acc\_max.  
Ensemble size 8



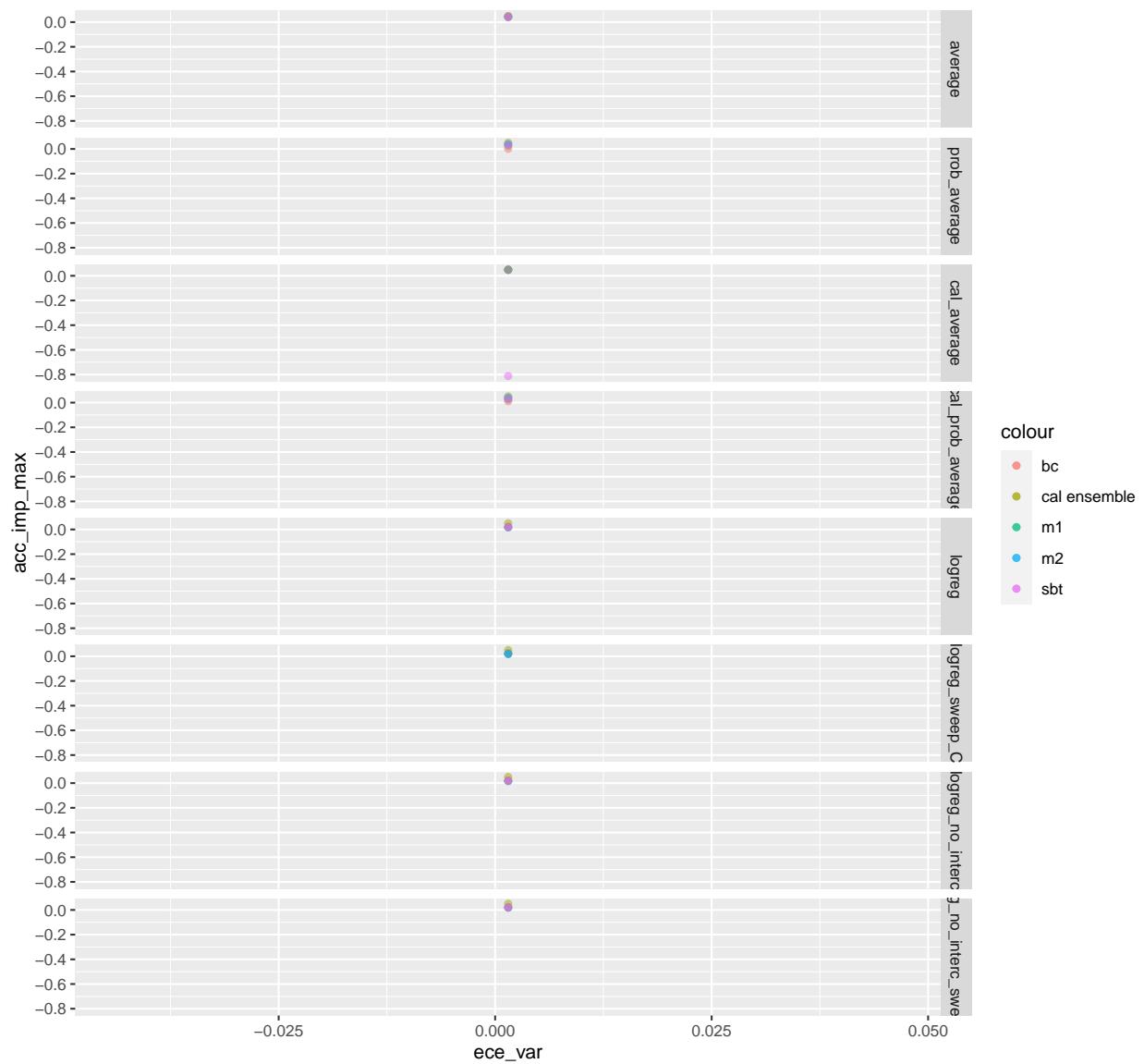
Accuracy improvement of ensemble over  
the best of networks vs acc\_avg.  
Ensemble size 8



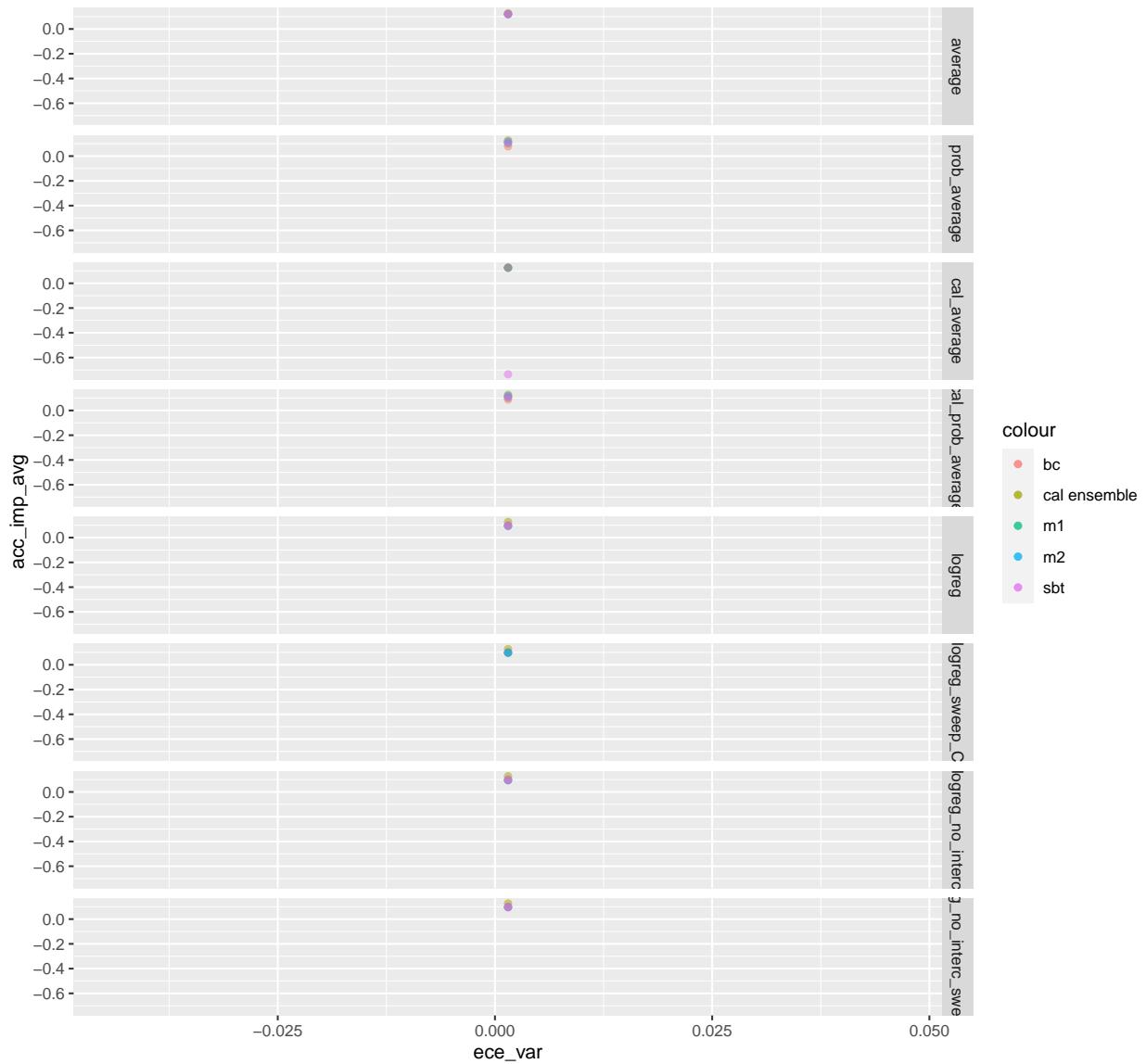
Accuracy improvement of ensemble over  
the average of networks vs acc\_avg.  
Ensemble size 8



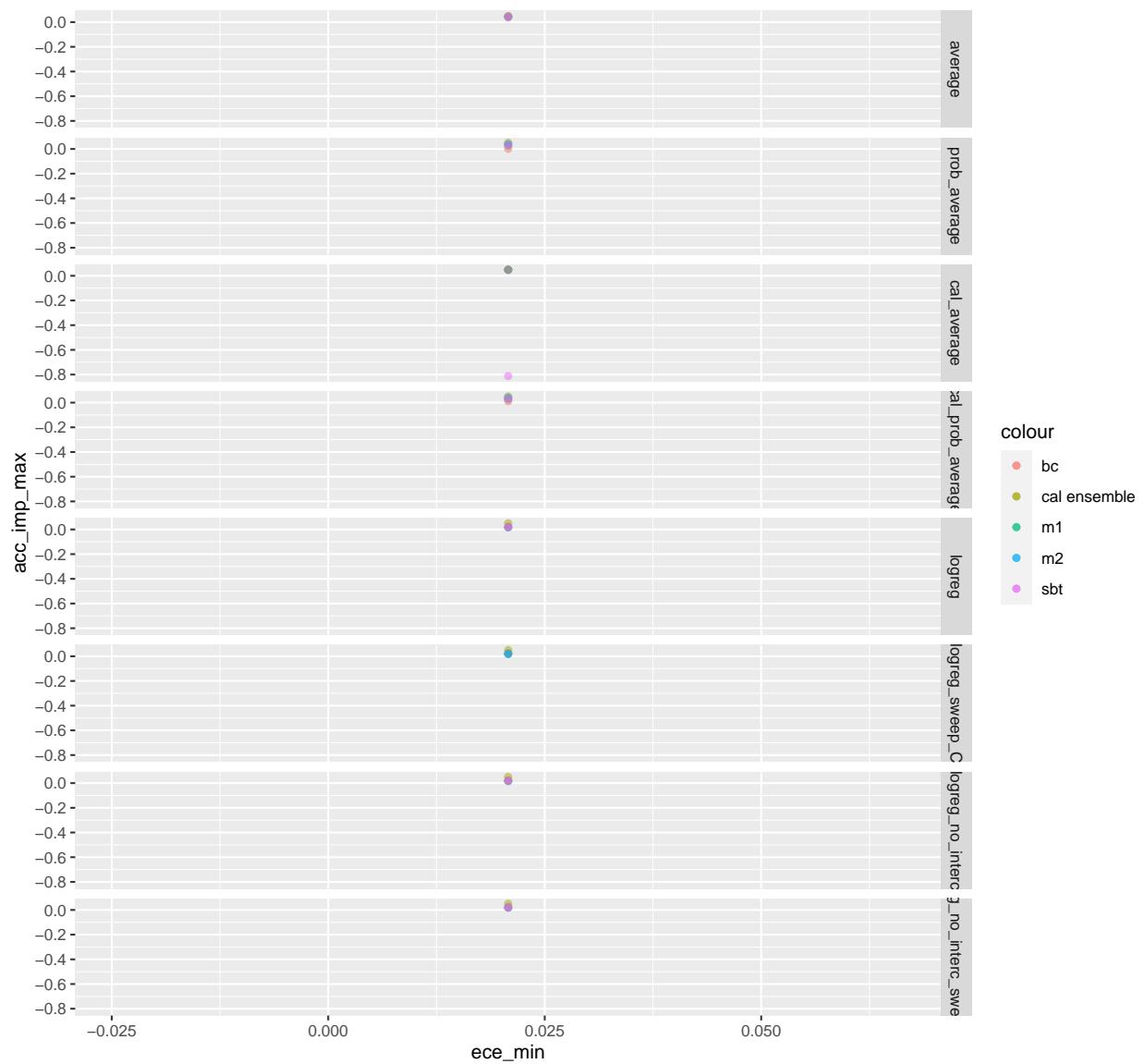
Accuracy improvement of ensemble over  
the best of networks vs ece\_var.  
Ensemble size 8



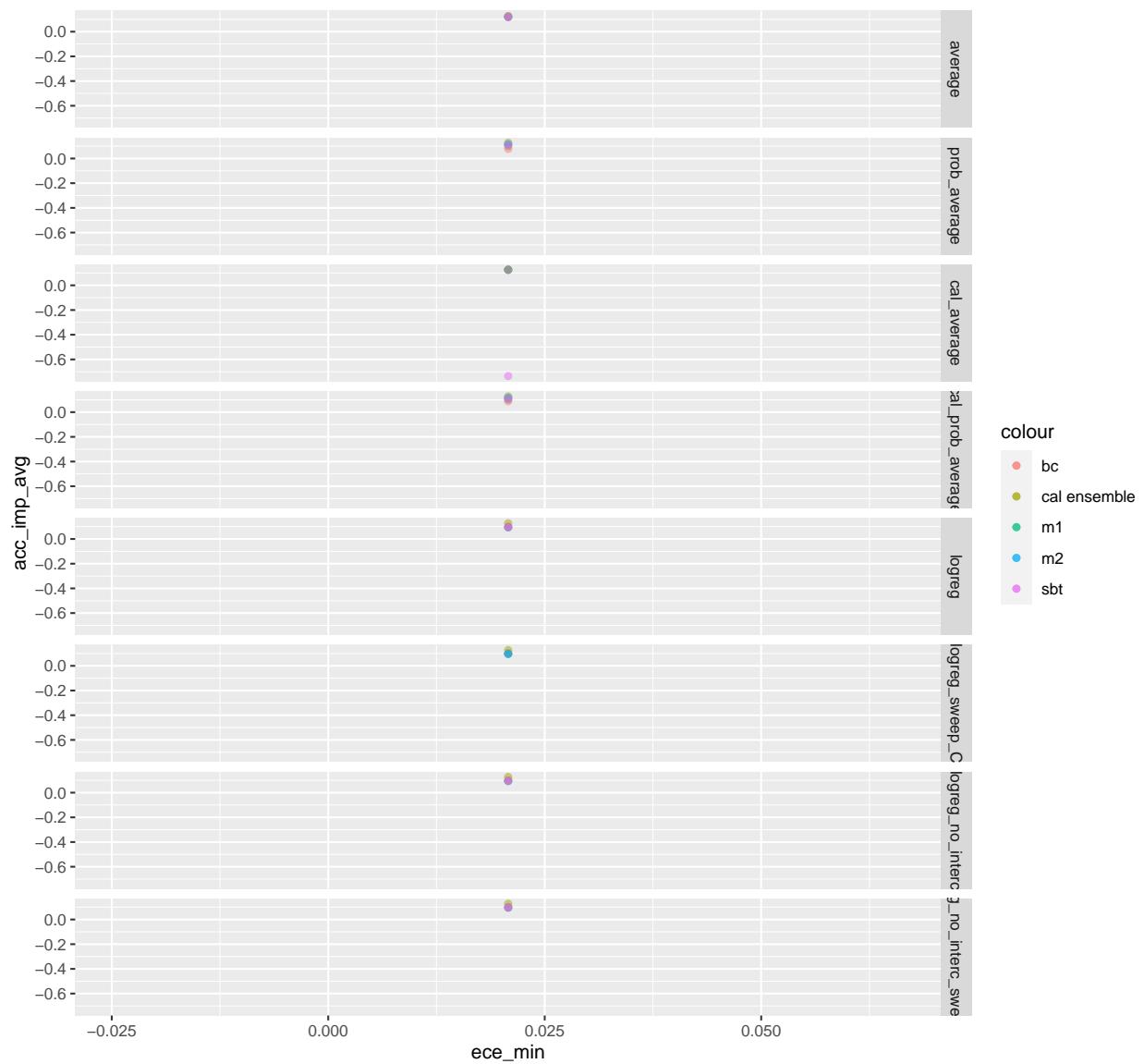
Accuracy improvement of ensemble over  
the average of networks vs ece\_var.  
Ensemble size 8



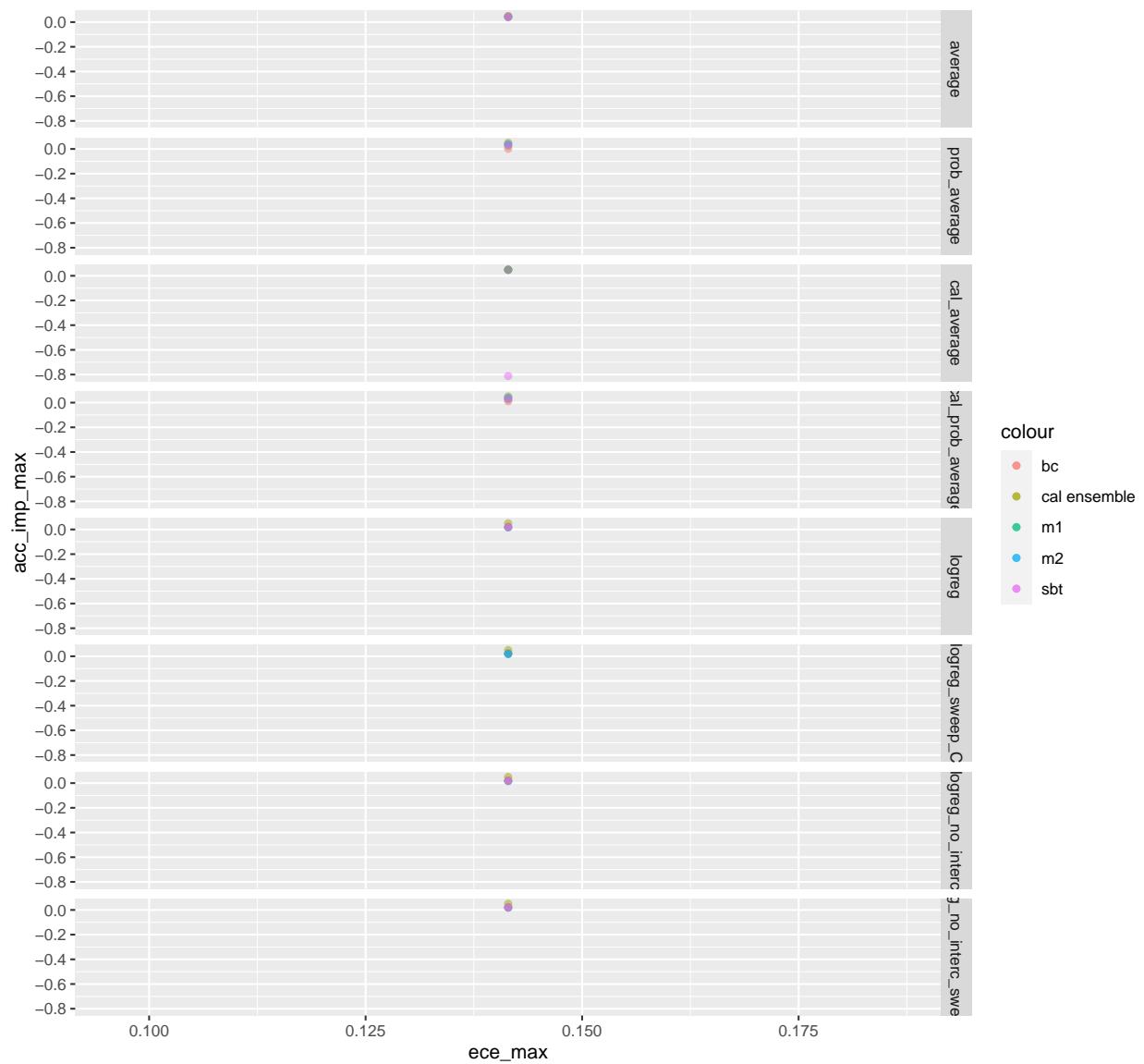
Accuracy improvement of ensemble over  
the best of networks vs ece\_min.  
Ensemble size 8



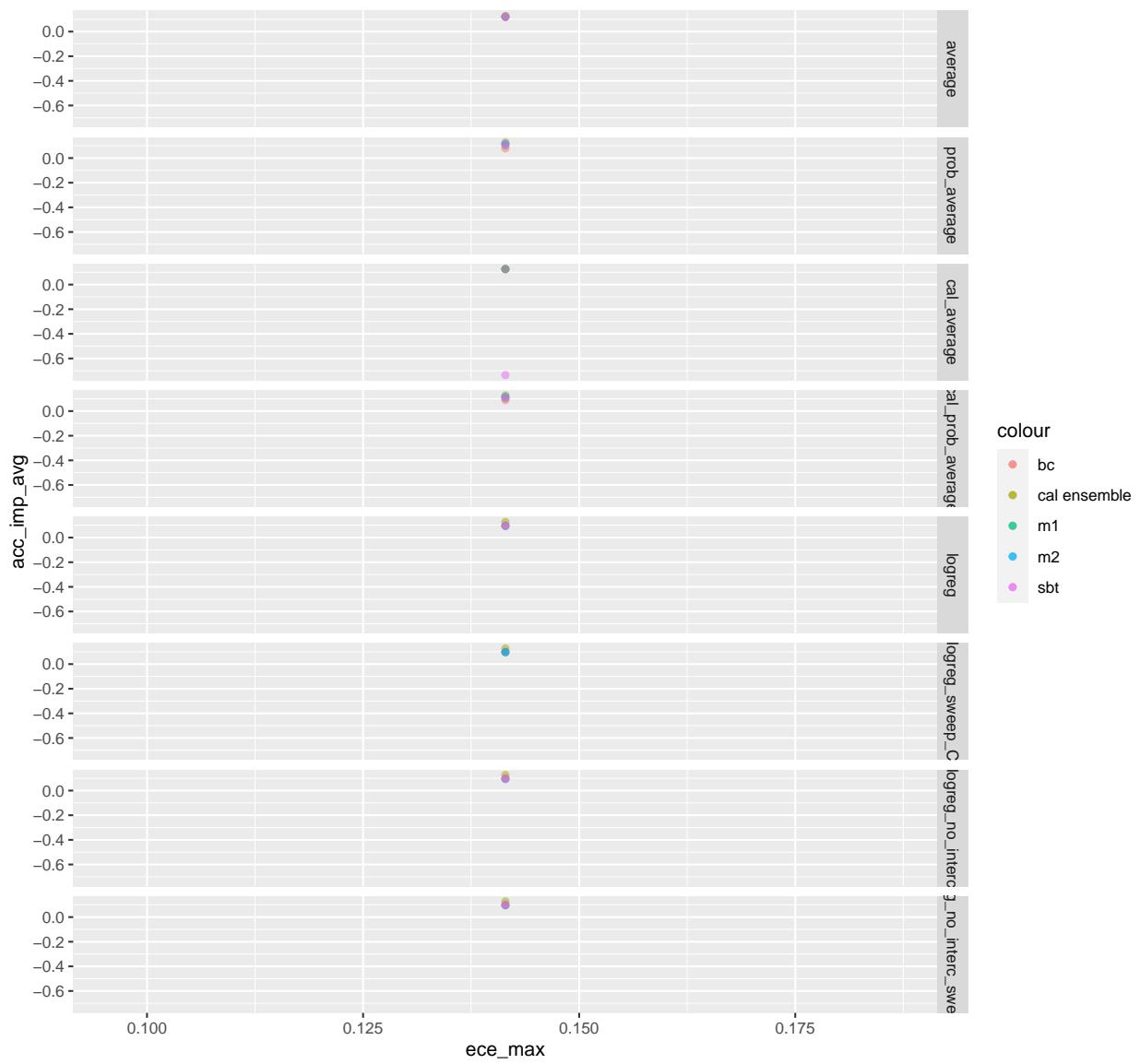
Accuracy improvement of ensemble over  
the average of networks vs ece\_min.  
Ensemble size 8



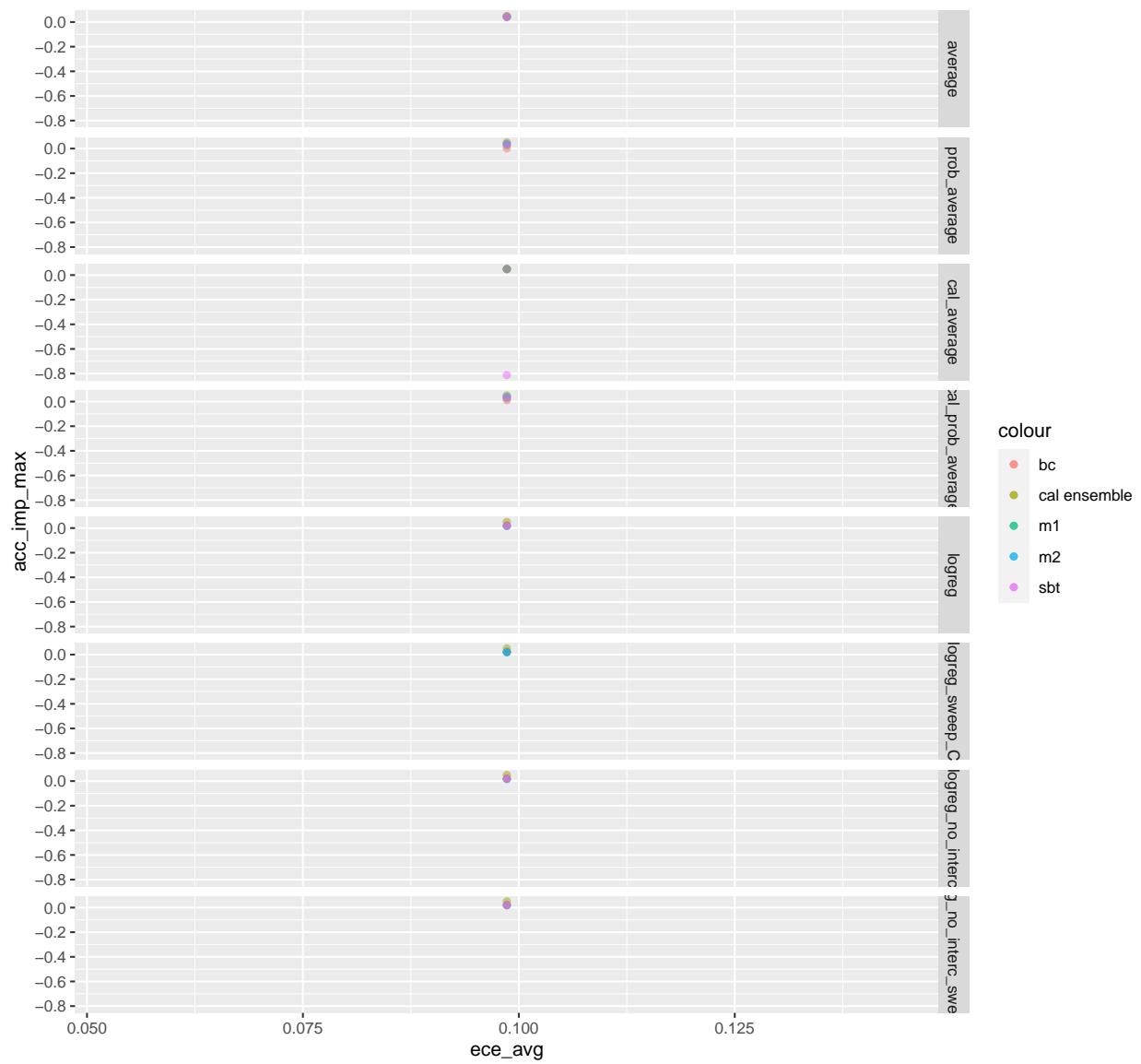
Accuracy improvement of ensemble over  
the best of networks vs ece\_max.  
Ensemble size 8



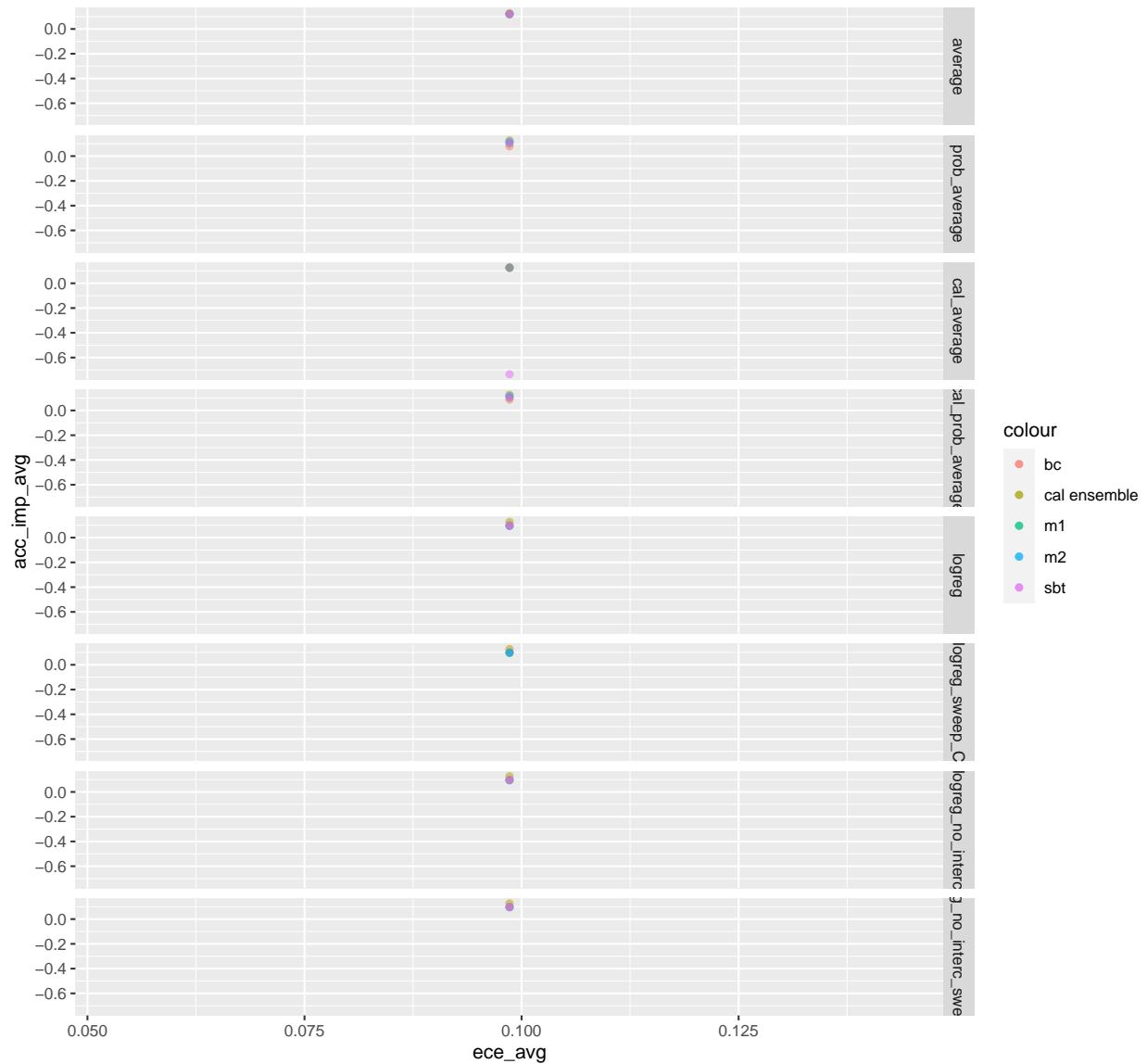
Accuracy improvement of ensemble over  
the average of networks vs ece\_max.  
Ensemble size 8



Accuracy improvement of ensemble over  
the best of networks vs ece\_avg.  
Ensemble size 8

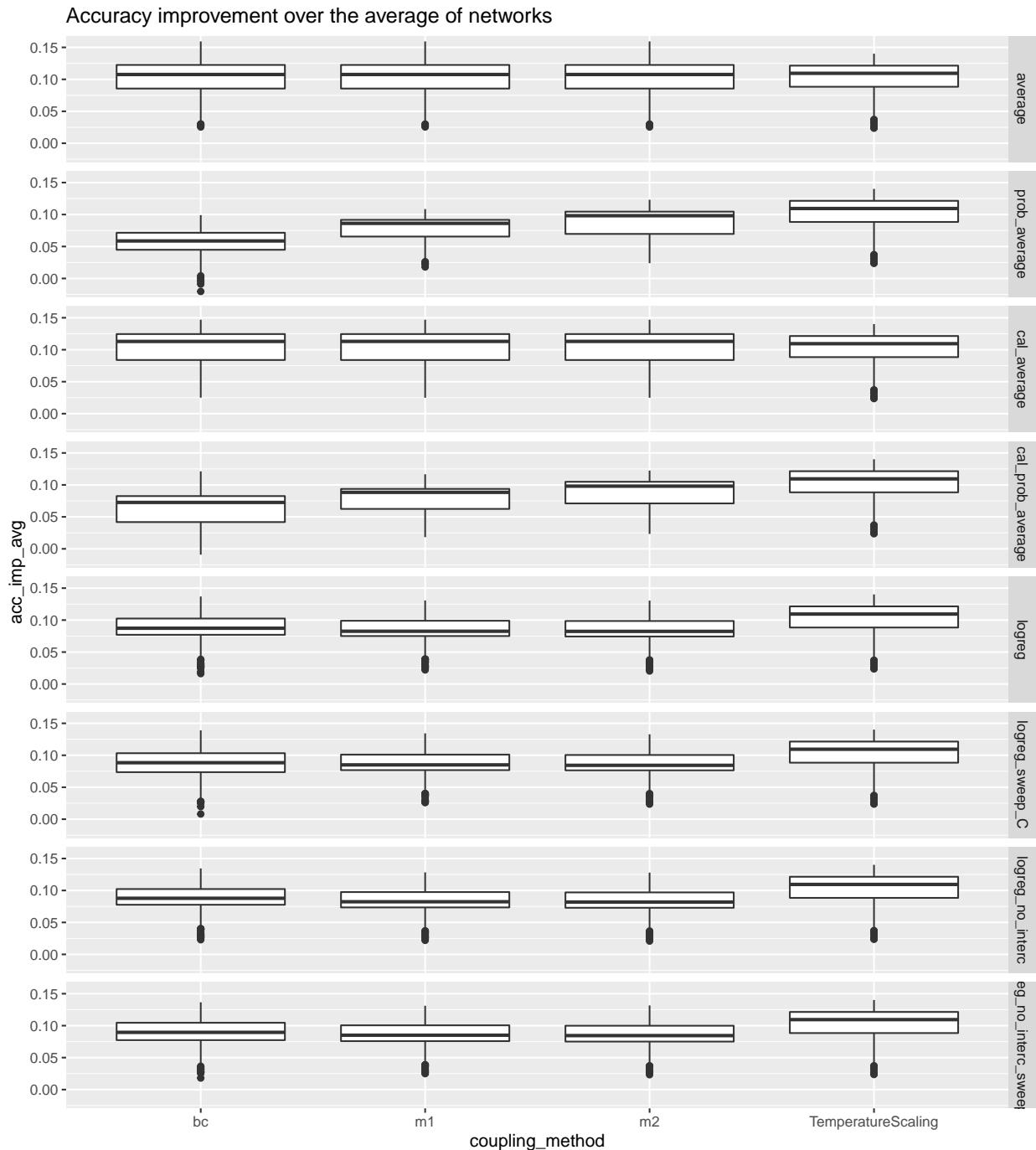


Accuracy improvement of ensemble over  
the average of networks vs ece\_avg.  
Ensemble size 8



```
imp_avg_plot <- ggplot() +
  geom_boxplot(data = ens_pwc_plt_df %>% filter(coupling_method != "sbt"), mapping = aes(x = coupling_m
  geom_boxplot(data = ens_cal_plt_df, mapping = aes(x = calibrating_method, y = acc_imp_avg)) +
  facet_grid(rows = vars(combining_method)) +
  ggtitle("Accuracy improvement over the average of networks")

print(imp_avg_plot)
```

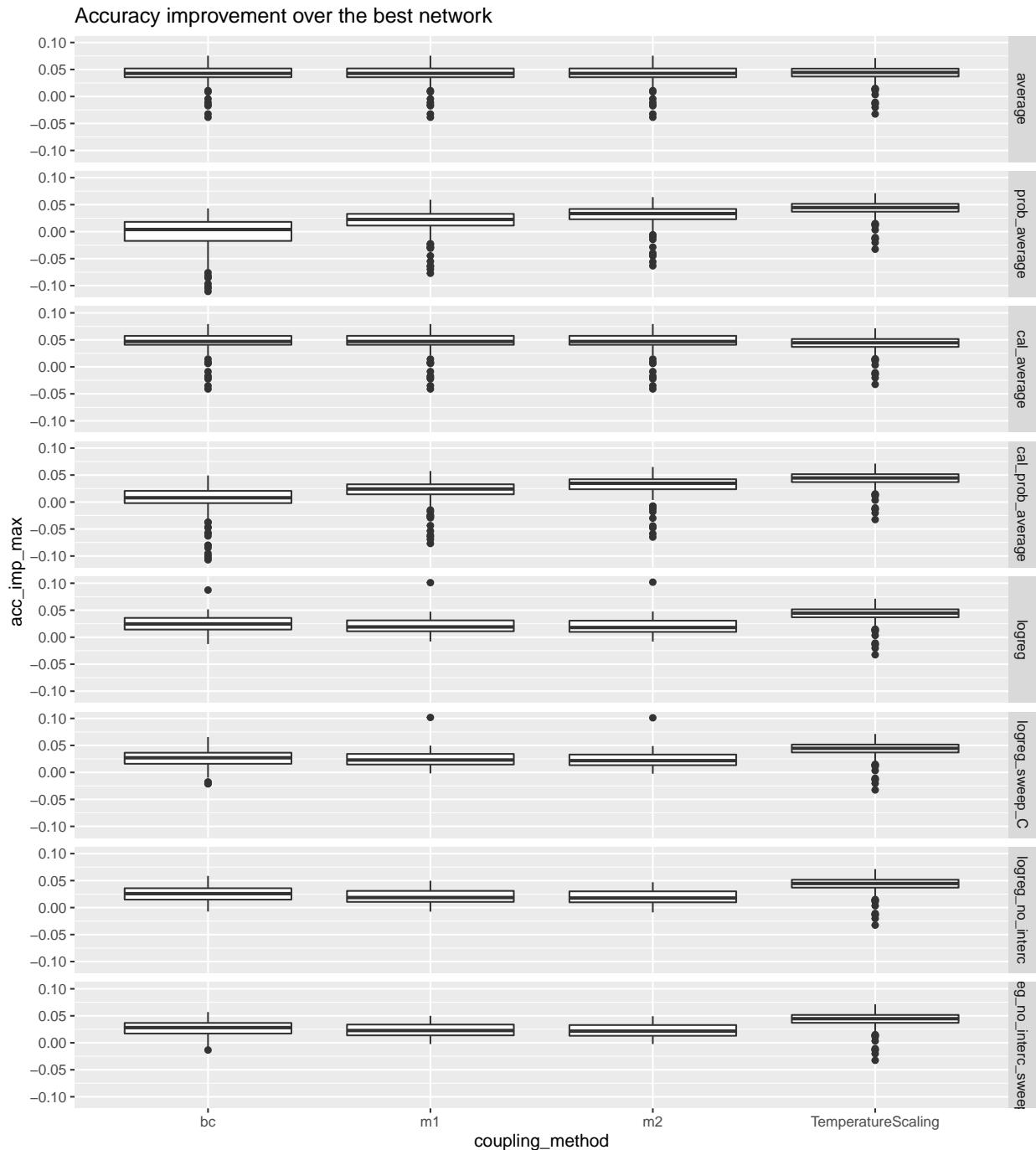


```

imp_max_plot <- ggplot() +
  geom_boxplot(data = ens_pwc_plt_df %>% filter(coupling_method != "sbt"), mapping = aes(x = coupling_m
  geom_boxplot(data = ens_cal_plt_df, mapping = aes(x = calibrating_method, y = acc_imp_max)) +
  facet_grid(rows = vars(combining_method)) +
  ggtitle("Accuracy improvement over the best network")

print(imp_max_plot)

```



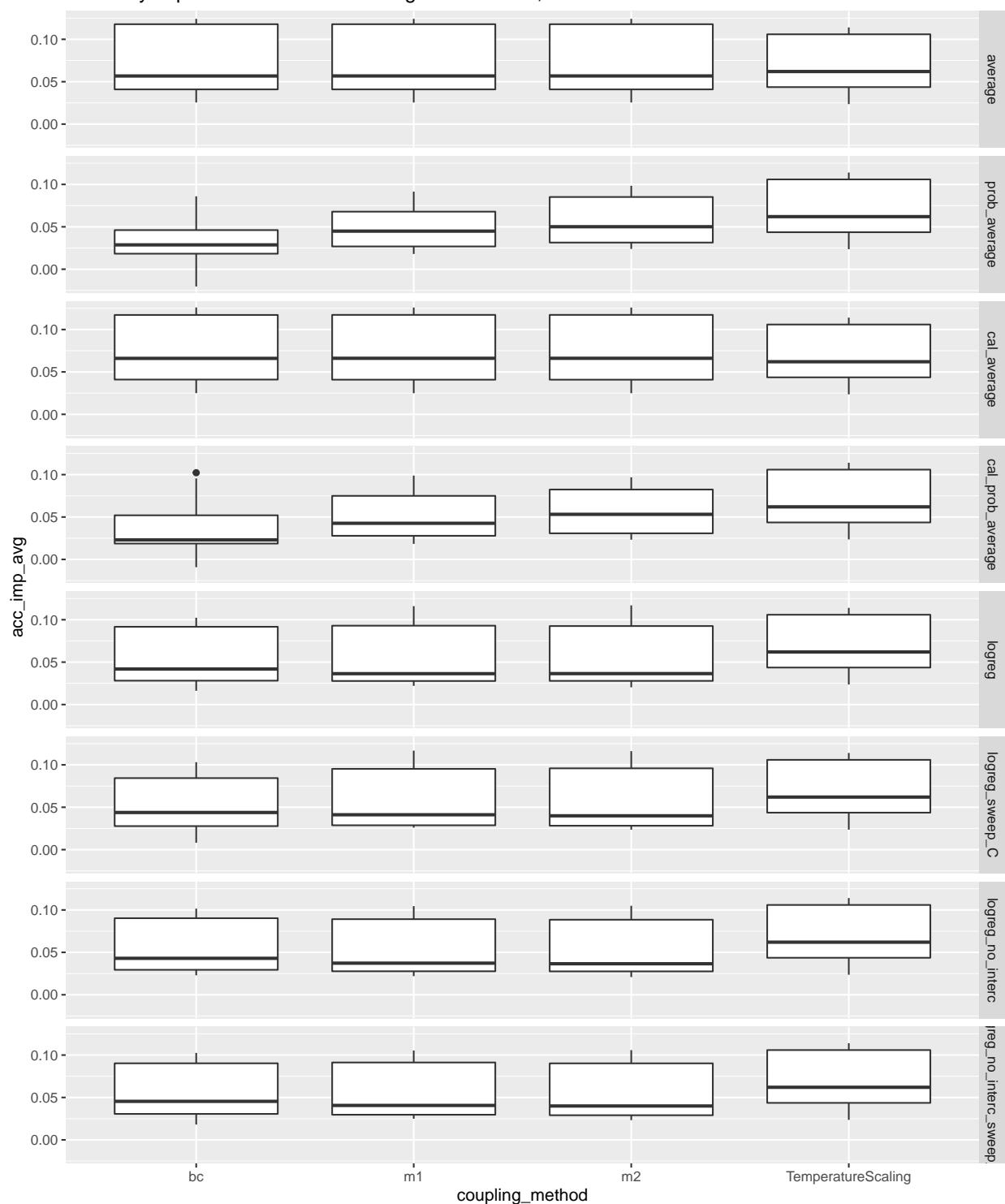
```

for (sss in unique(ens_cal_plt_df$combination_size))
{
  imp_avg_plot <- ggplot() +
    geom_boxplot(data = ens_pwc_plt_df %>% filter(coupling_method != "sbt" & combination_size == sss),
    geom_boxplot(data = ens_cal_plt_df %>% filter(combination_size == sss), mapping = aes(x = calibration,
    facet_grid(rows = vars(combining_method)) +
    ggttitle(sprintf("Accuracy improvement over the average of networks, ensemble size %s", sss))

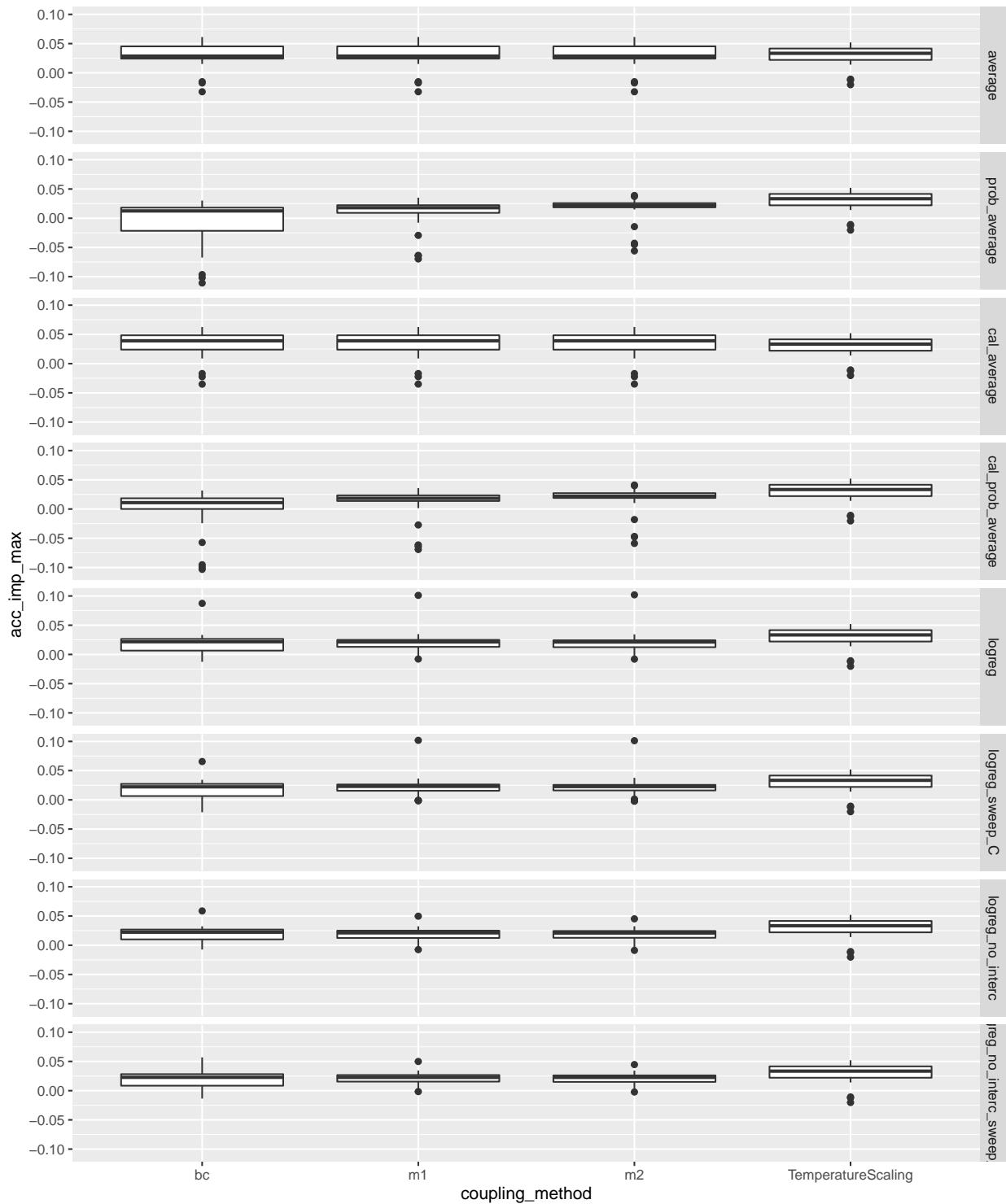
  print(imp_avg_plot)
}
  
```

```
imp_max_plot <- ggplot() +  
  geom_boxplot(data = ens_pwc_plt_df %>% filter(coupling_method != "sbt" & combination_size == sss),  
  geom_boxplot(data = ens_cal_plt_df %>% filter(combination_size == sss), mapping = aes(x = calibration_size,  
  facet_grid(rows = vars(combining_method)) +  
  ggtitle(sprintf("Accuracy improvement over the best of networks, ensemble size %s", sss))  
  
  print(imp_max_plot)  
}
```

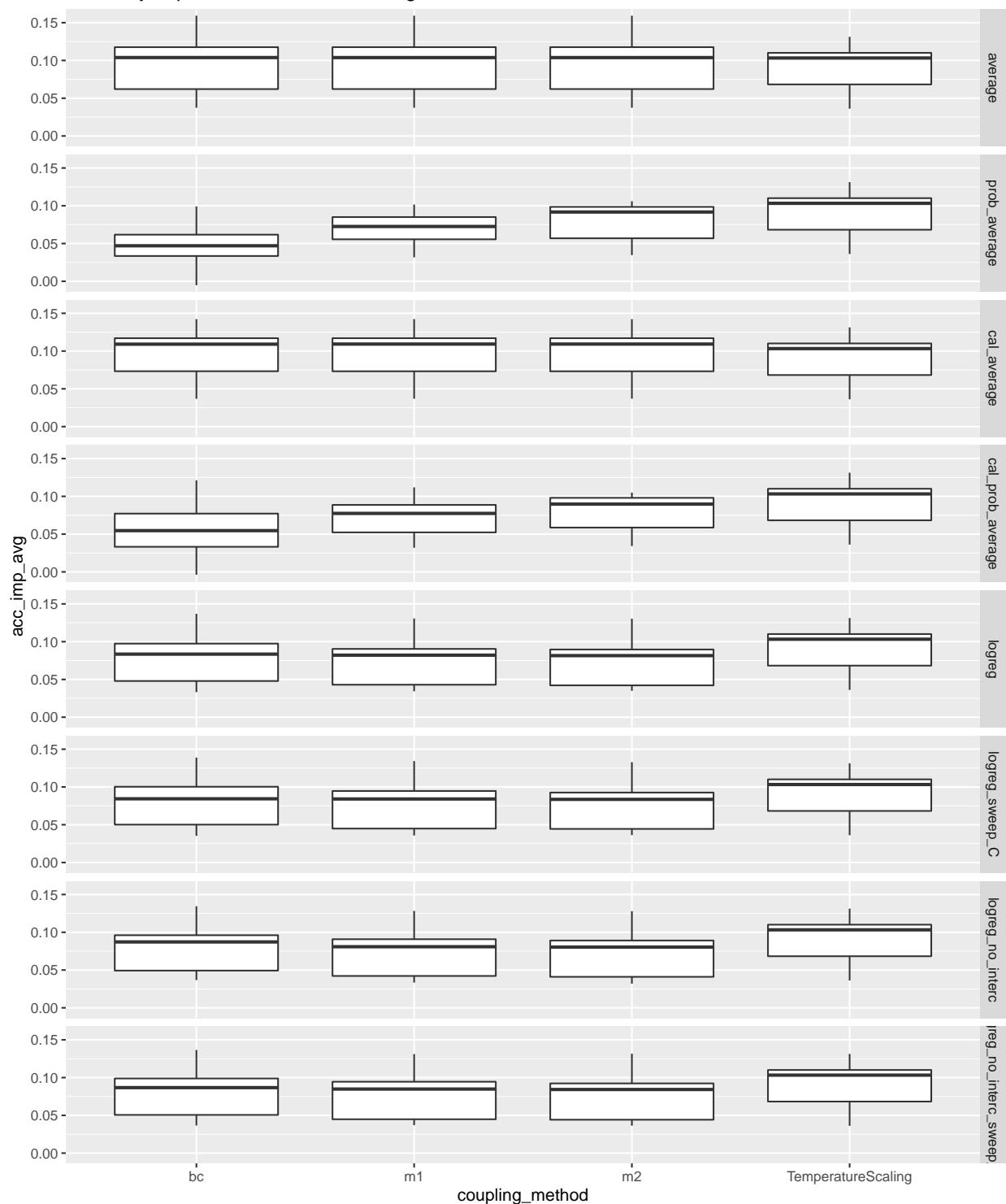
Accuracy improvement over the average of networks, ensemble size 2



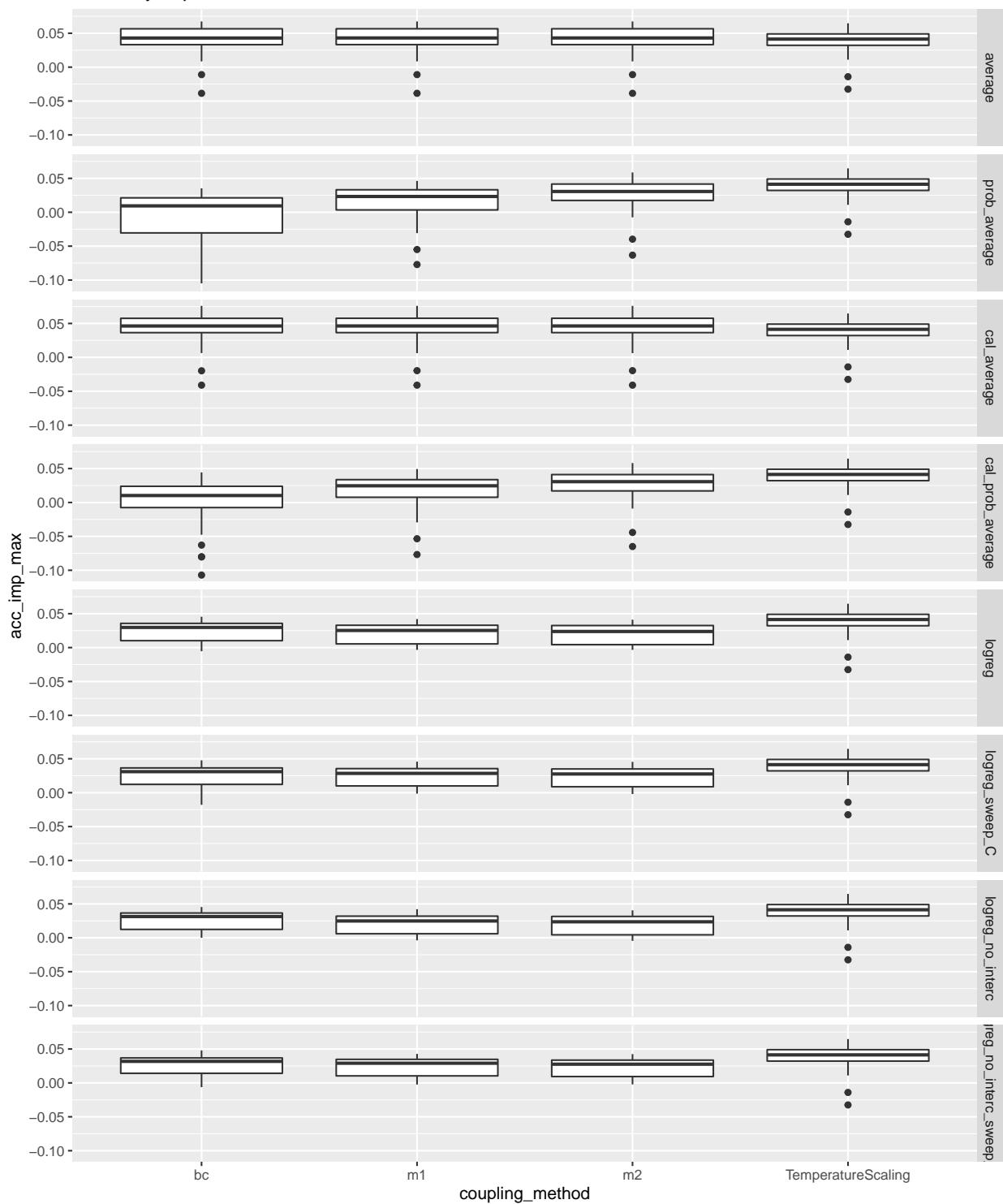
Accuracy improvement over the best of networks, ensemble size 2

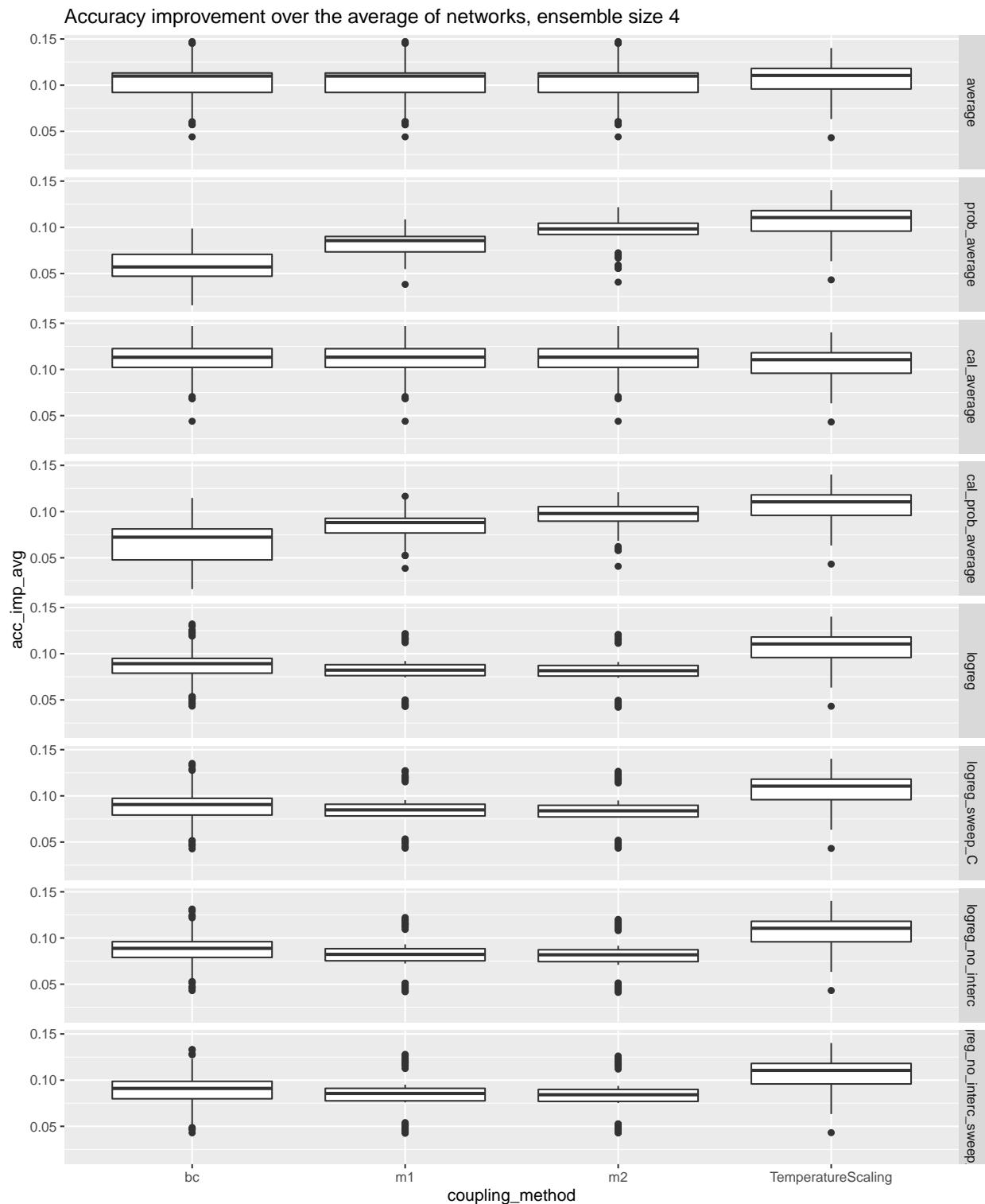


Accuracy improvement over the average of networks, ensemble size 3

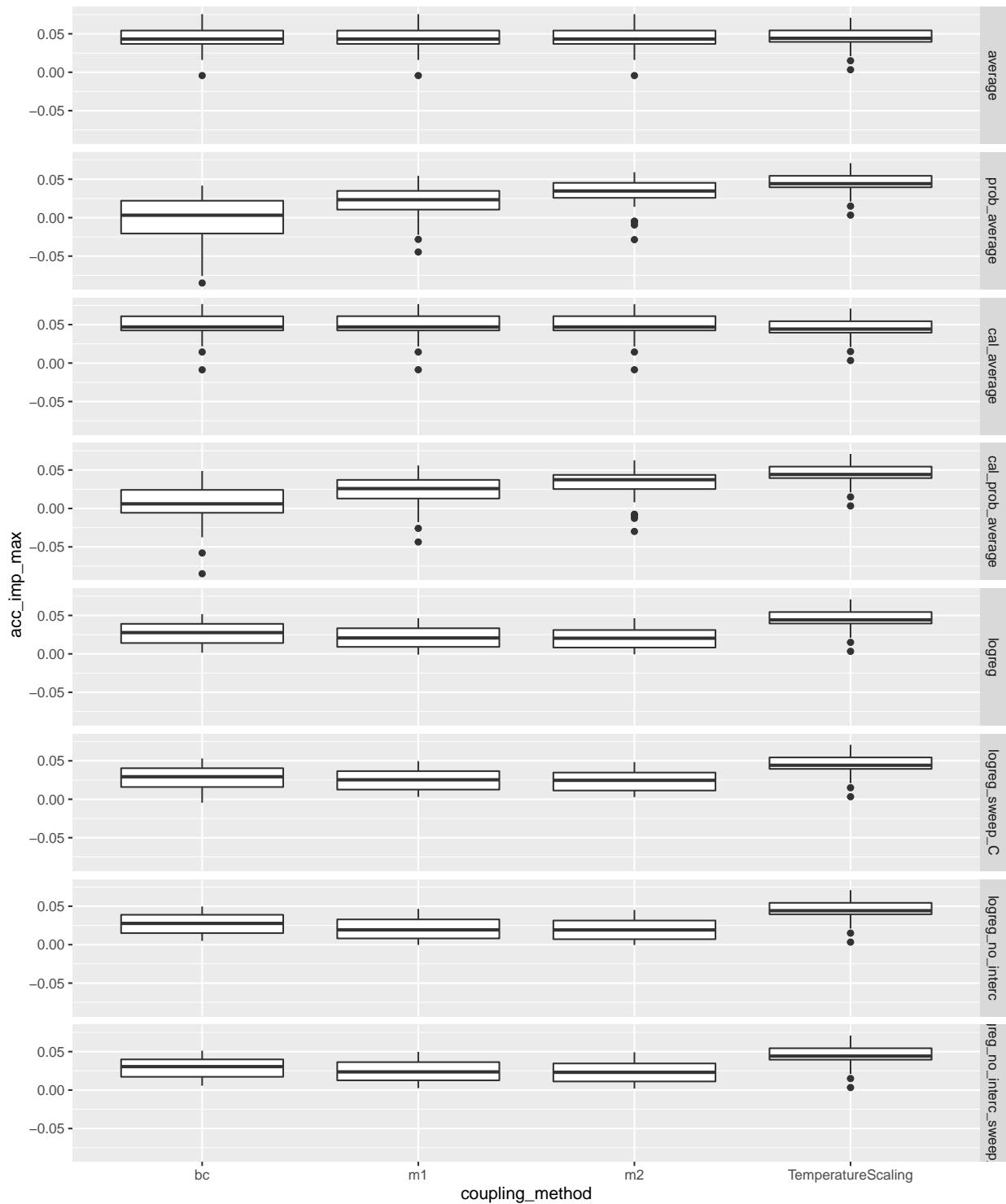


Accuracy improvement over the best of networks, ensemble size 3

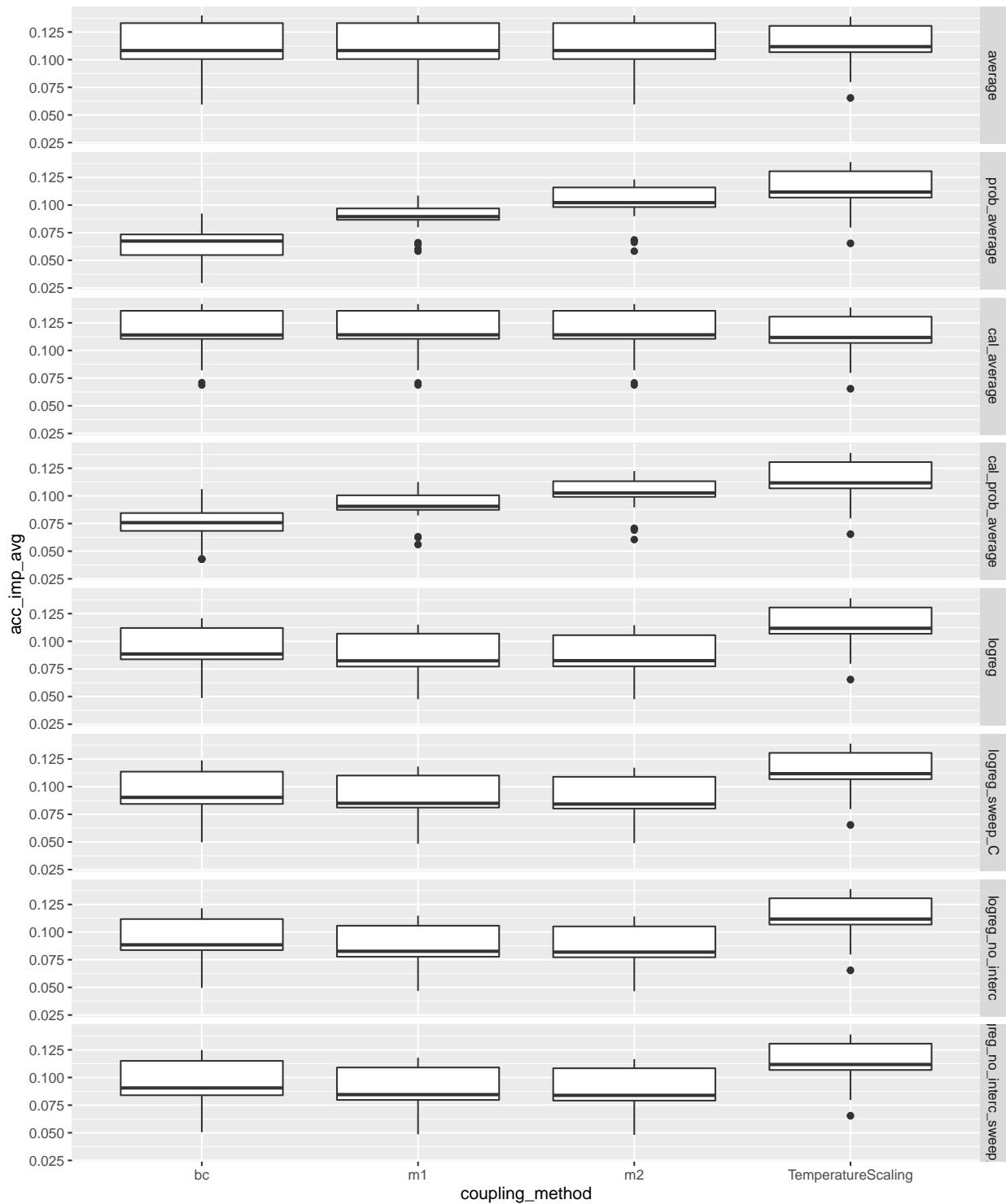




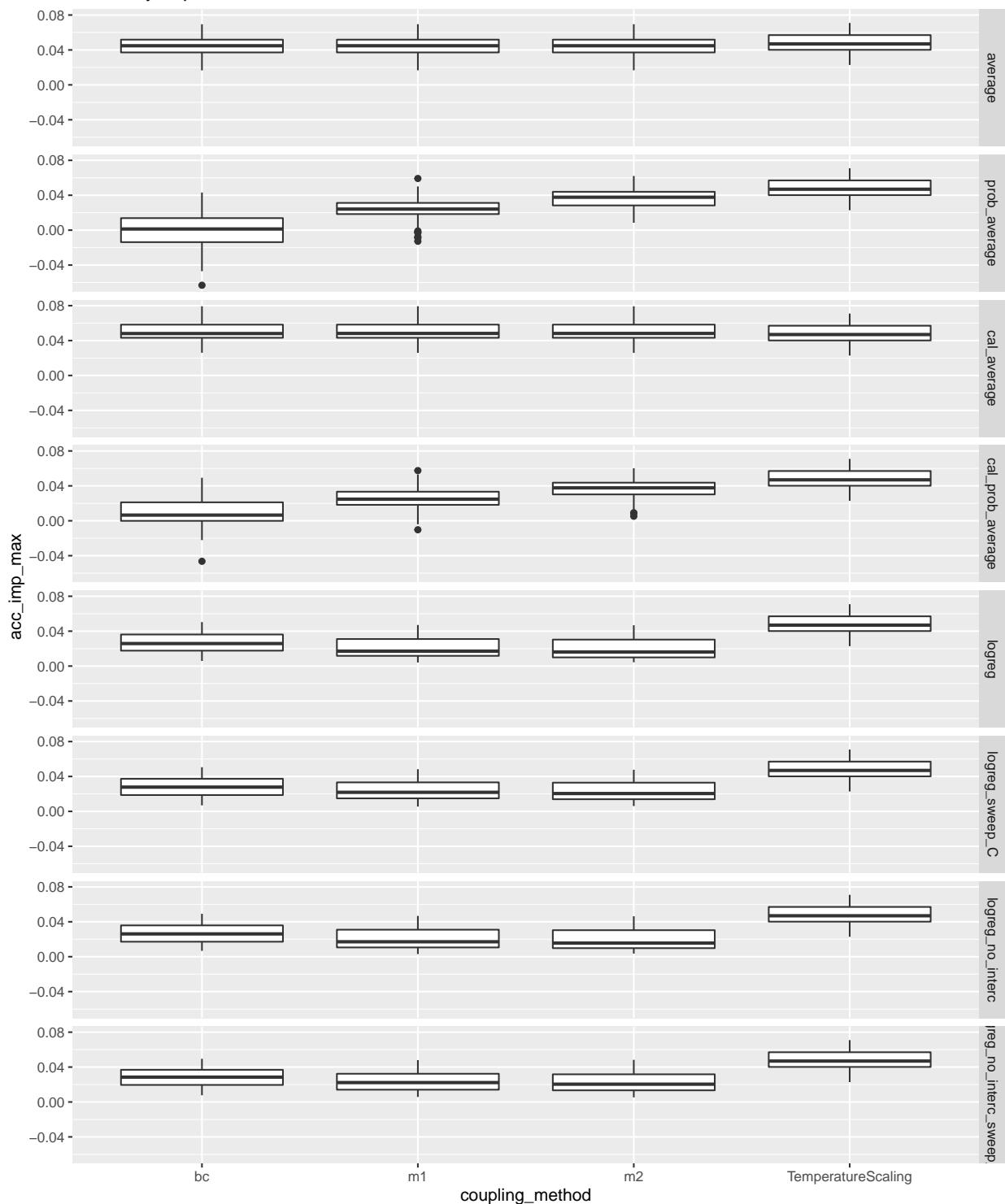
Accuracy improvement over the best of networks, ensemble size 4

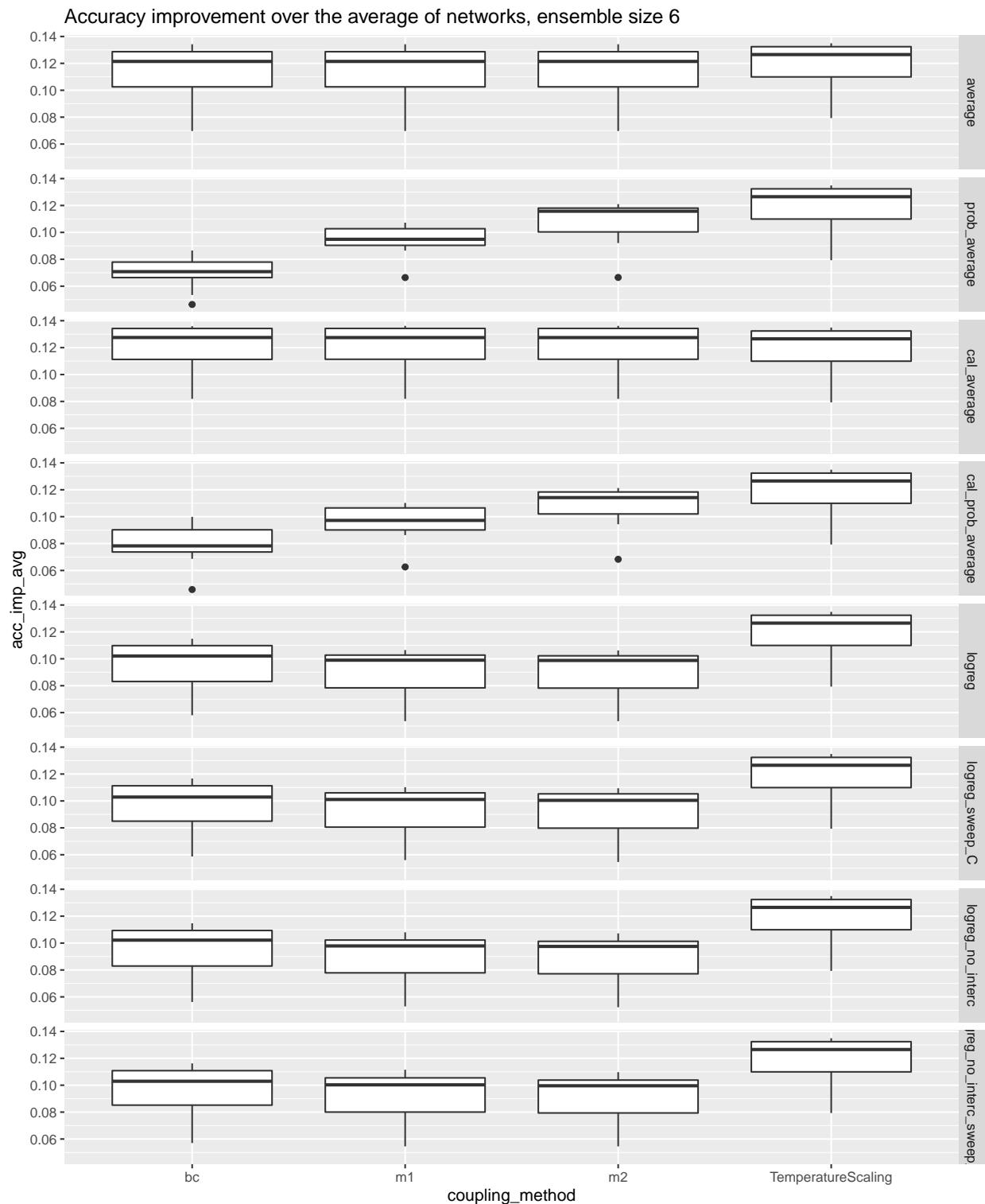


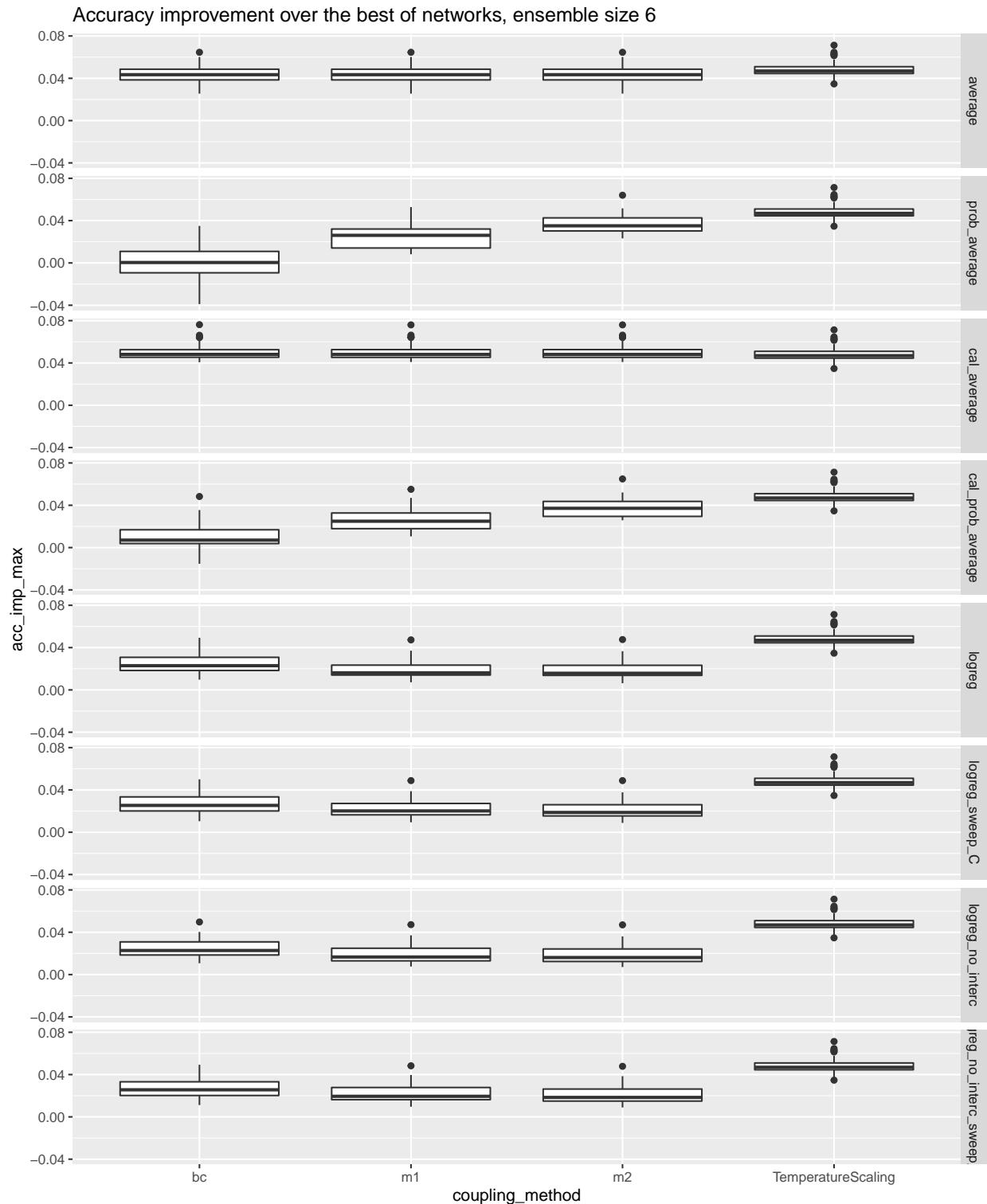
Accuracy improvement over the average of networks, ensemble size 5



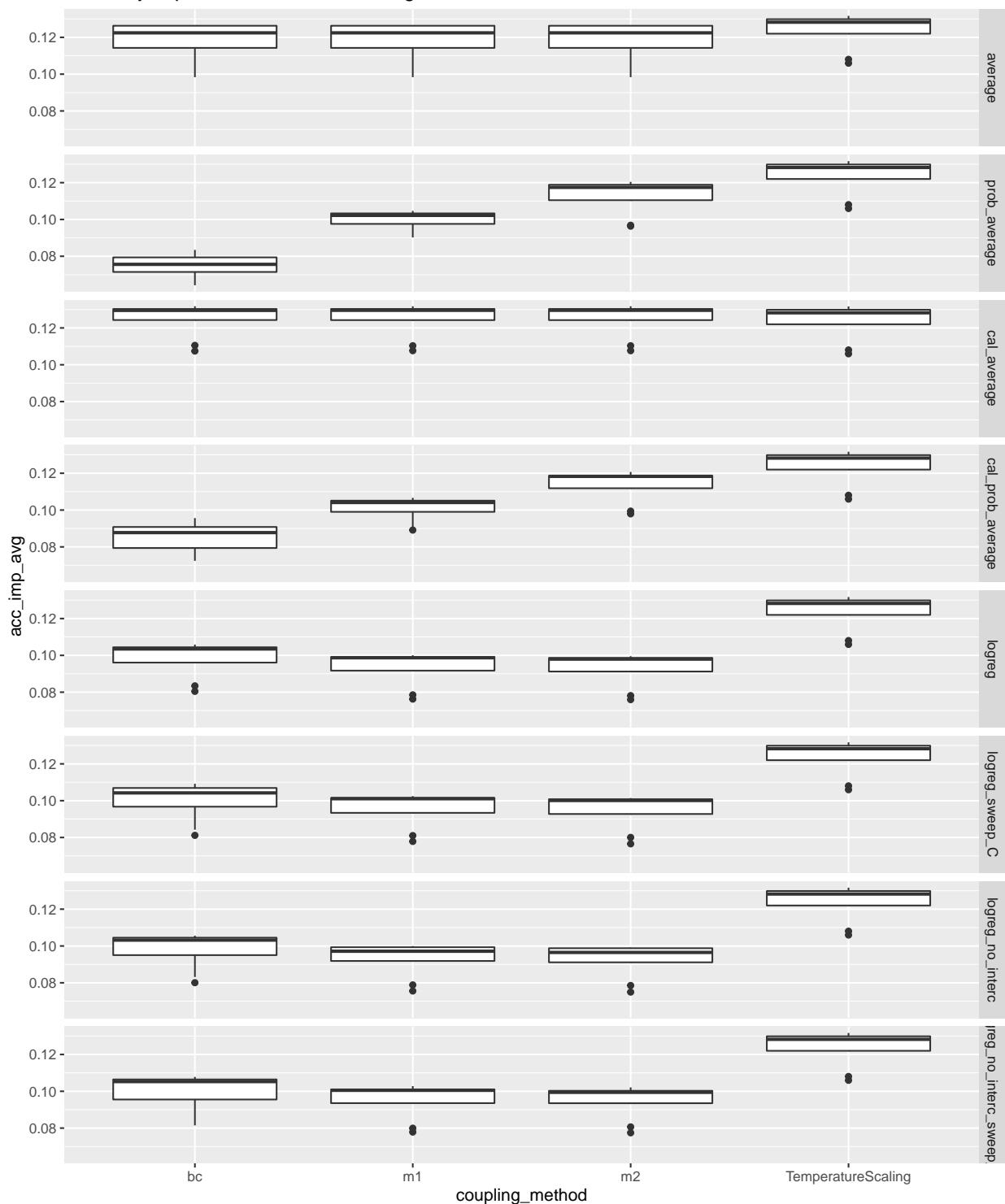
Accuracy improvement over the best of networks, ensemble size 5



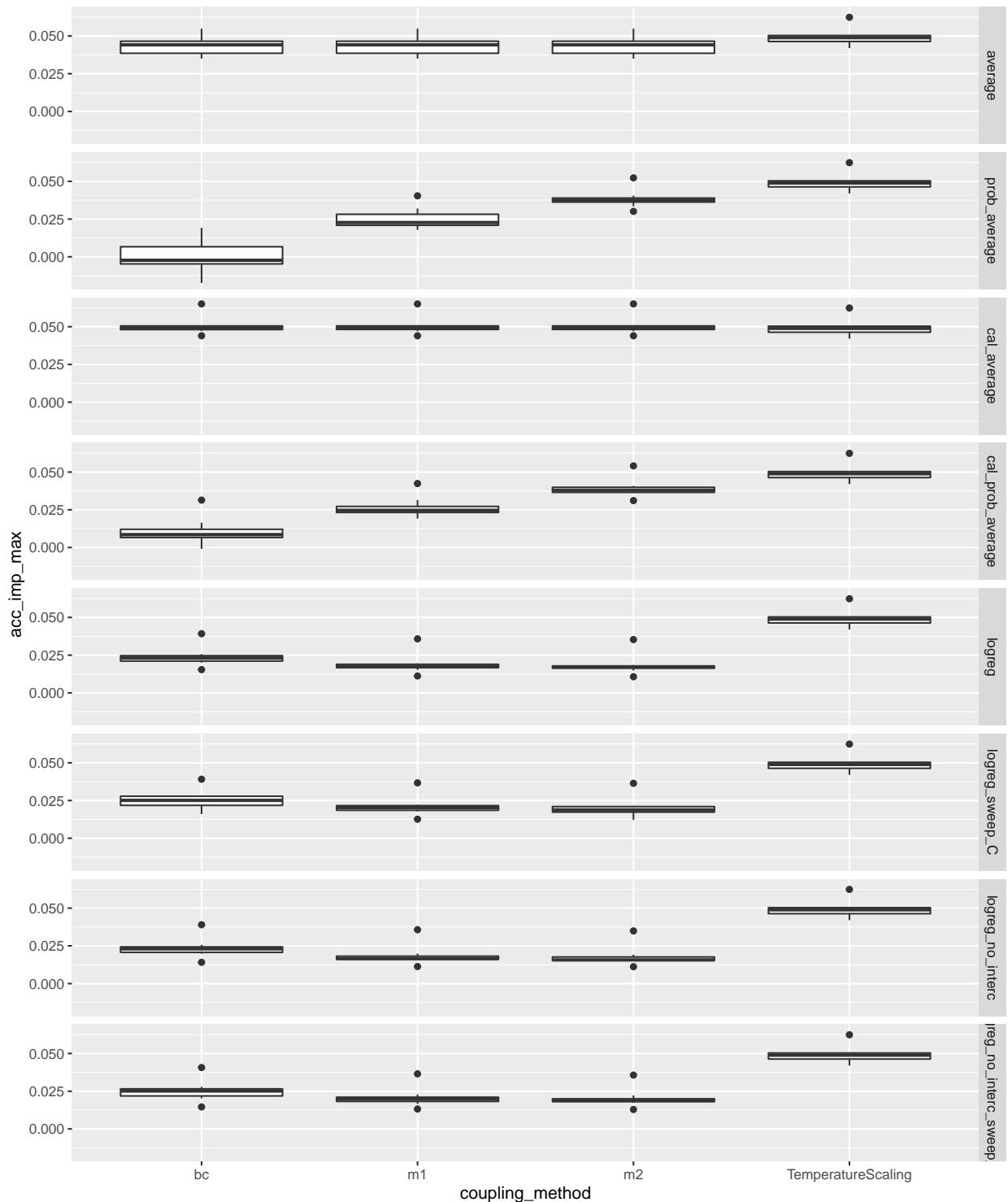




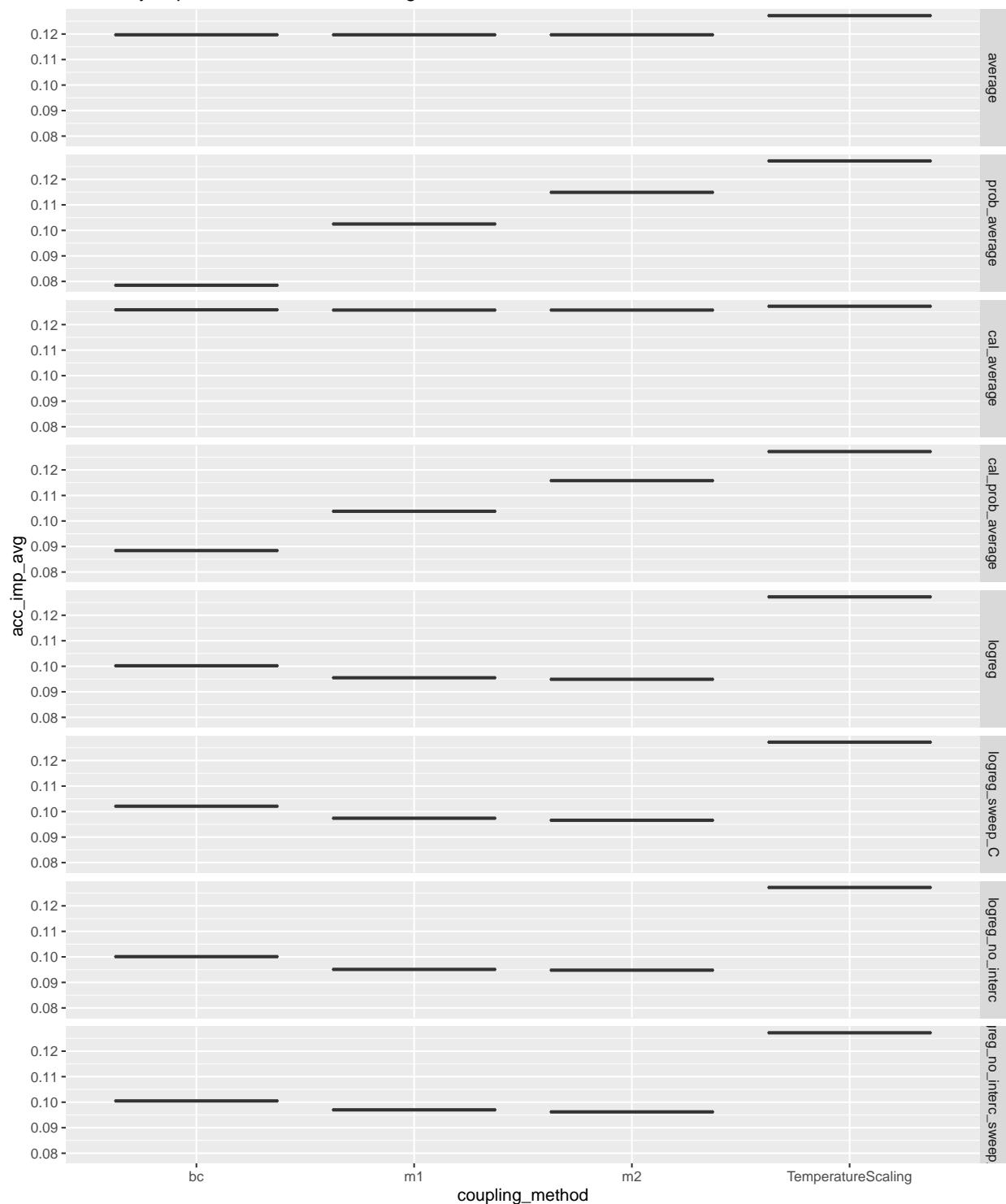
Accuracy improvement over the average of networks, ensemble size 7



Accuracy improvement over the best of networks, ensemble size 7



Accuracy improvement over the average of networks, ensemble size 8



Accuracy improvement over the best of networks, ensemble size 8

