

WeightedLinearEnsemble evaluation new

```
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

library(xtable)
```

```

my_plot_hook <- function(x, options)
  paste("\n", knitr::hook_plot_tex(x, options), "\n")
knitr::knit_hooks$set(plot = my_plot_hook)

base_dir <- "D:/skola/1/weighted_ensembles/tests/test_cifar_2021/data/data_tv_5000_c100/0/new"
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)
  )
}
}

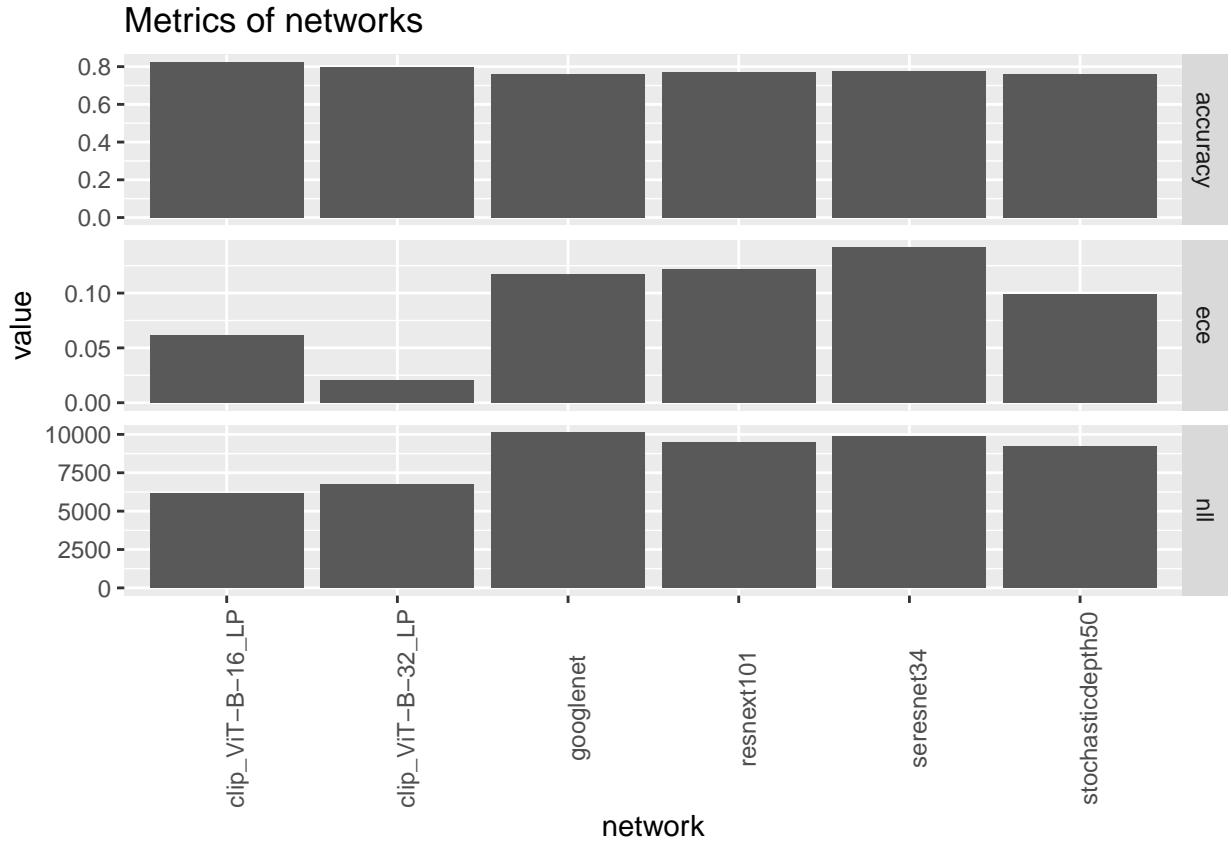
ens_df_cal <- merge(ens_df_cal, comb_stats_df)
ens_df_cal$acc_imp_avg <- ens_df_cal$accuracy - ens_df_cal$acc_avg
ens_df_cal$acc_imp_max <- ens_df_cal$accuracy - ens_df_cal$acc_max

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

comb_methods <- c(
  "average", "prob_average",
  "cal_average", "cal_prob_average",
  "logreg", "logreg_sweep_C",
  "logreg_no_interc", "logreg_no_interc_sweep_C",
  "grad_m1", "grad_m2", "grad_bc"
)
comb_methods <- c(sapply(X=comb_methods, FUN=function(cm) c(cm, paste(cm, "uncert", sep="."))))

ens_pwc_plt_df$combining_method <- factor(ens_pwc_plt_df$combining_method,
  levels = comb_methods)

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_hpline(
        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_hpline(
        data = cur_ens_pwc,

```

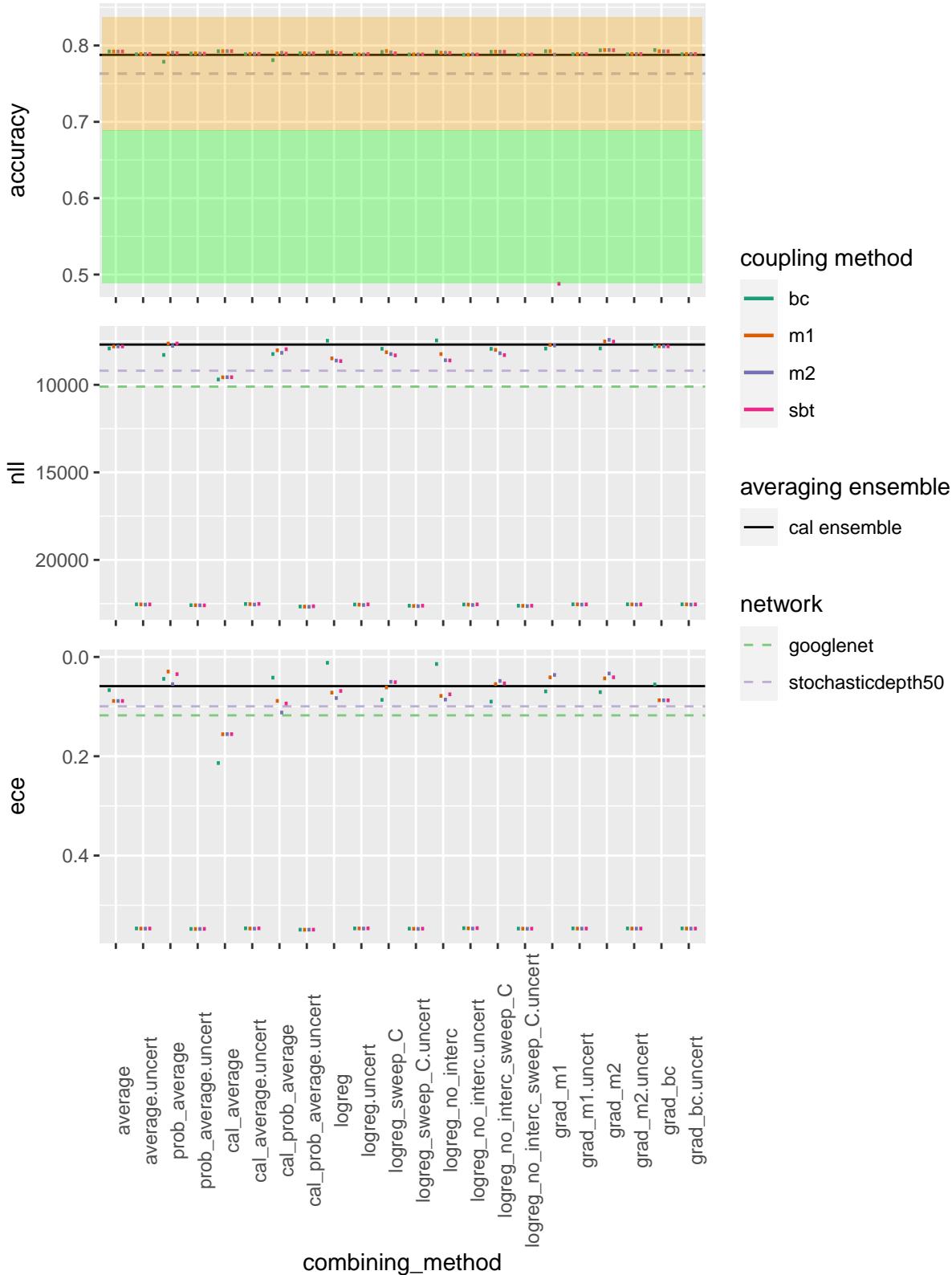
```

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))

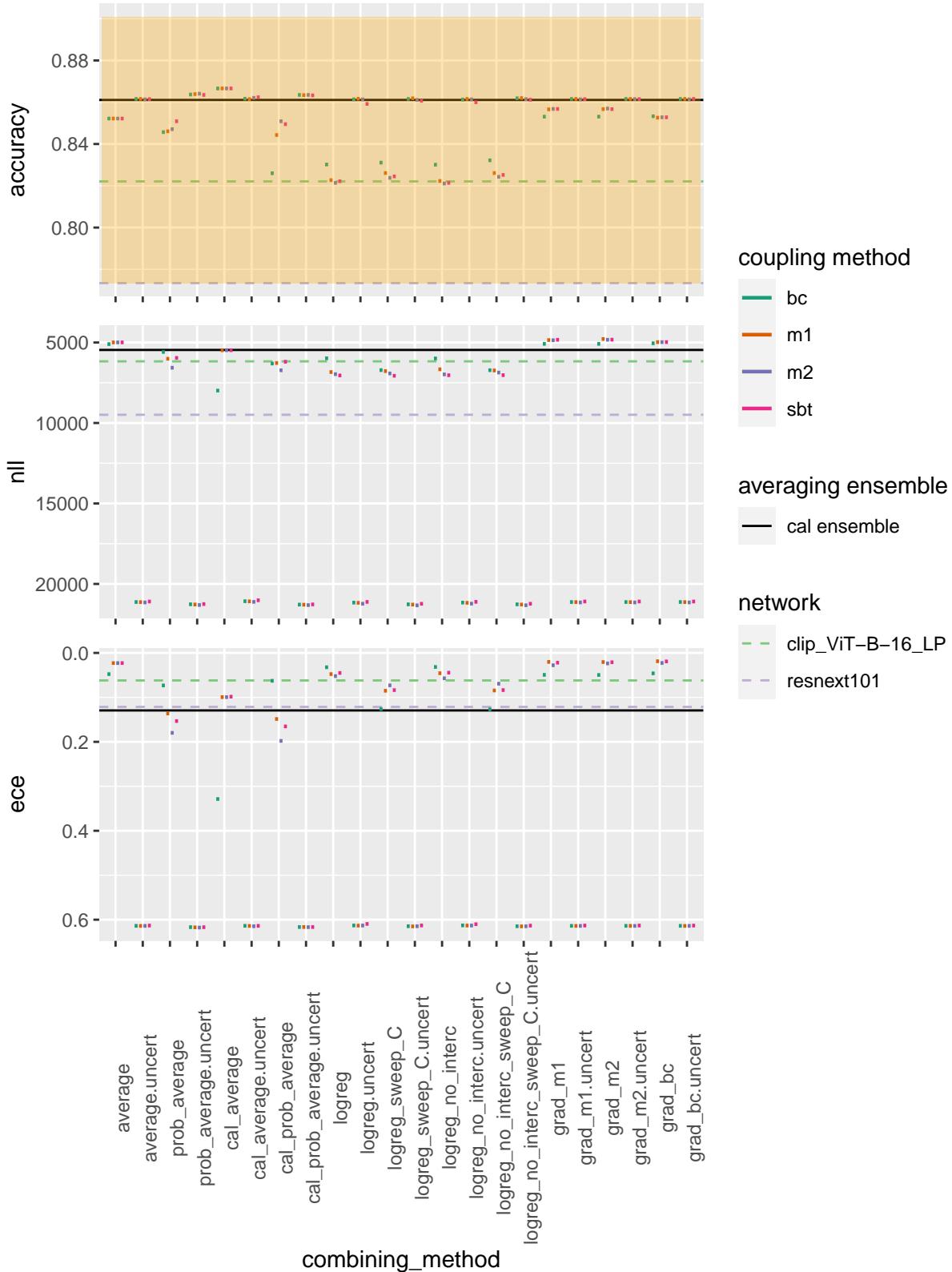
print(acc_plot / nll_plot / ece_plot + plot_layout(guides = "collect") +
  plot_annotation(title = paste(
    "Ensemble metrics",
    paste(
      c("Error inconsistency", cur_ens_cal$err_incons[[1]]),
      collapse = " "
    ),
    sep = "\n"
  )))
}
}

```

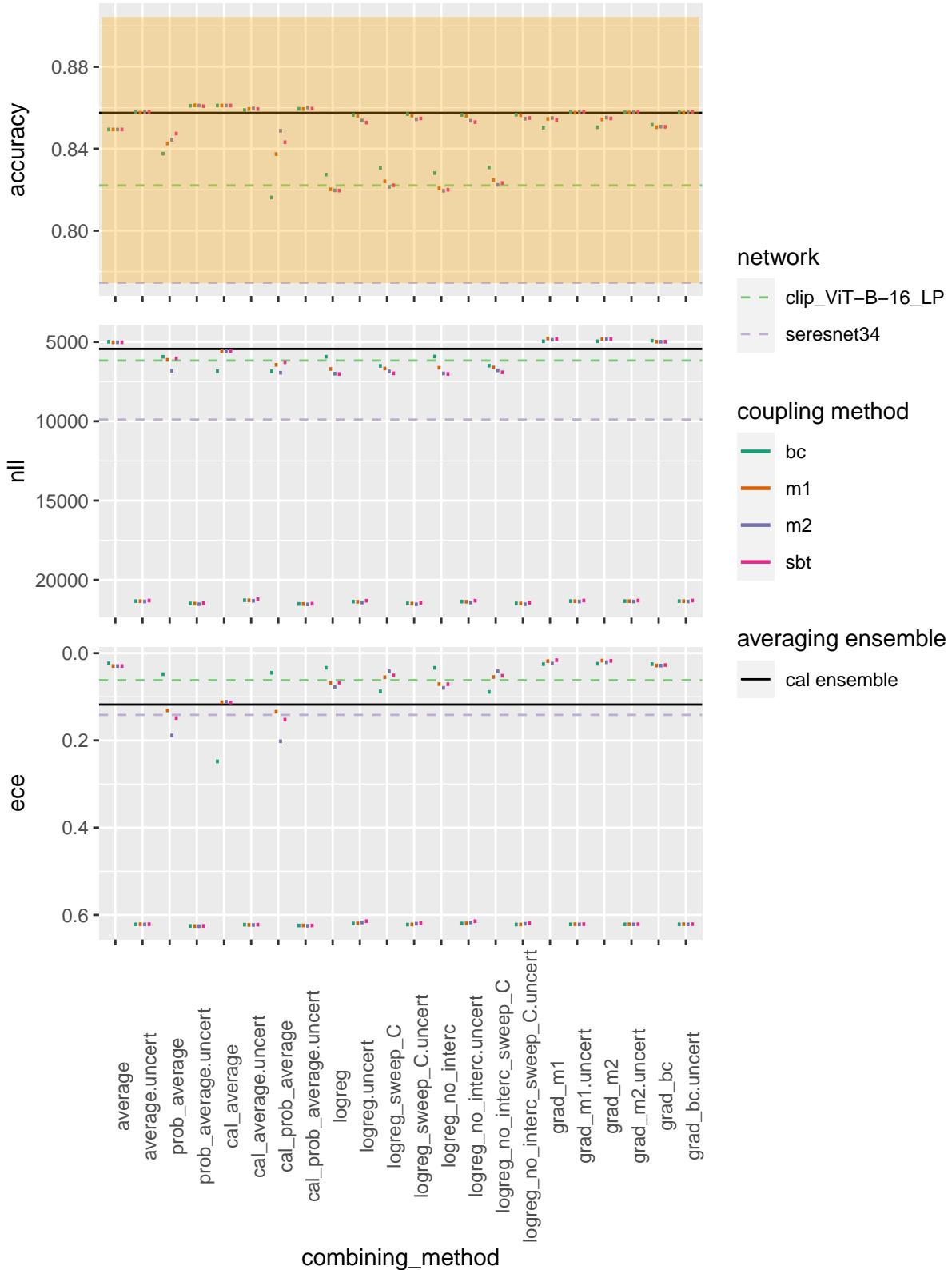
Ensemble metrics
Error inconsistency 0.148800000548363



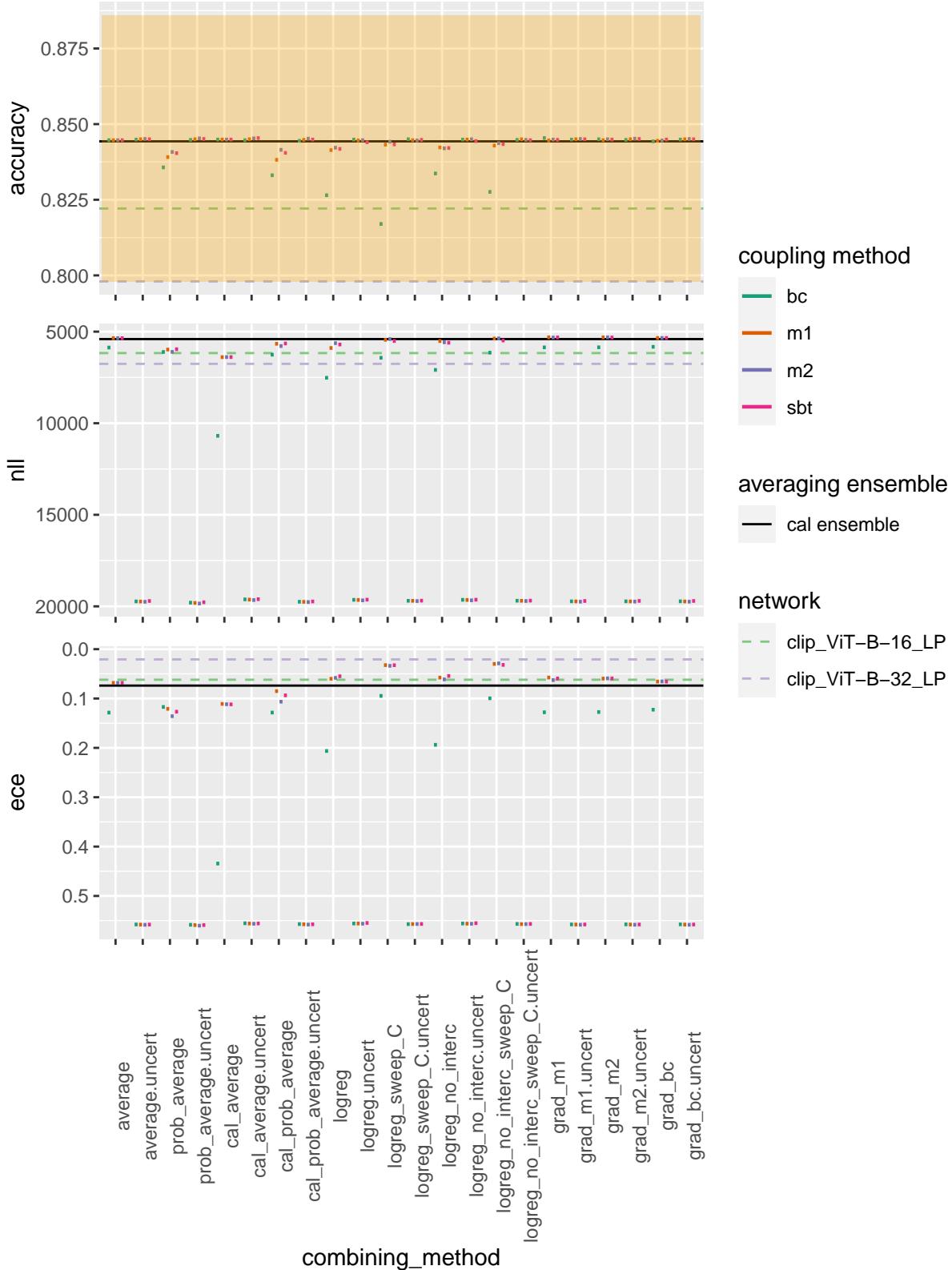
Ensemble metrics
Error inconsistency 0.206299990415573



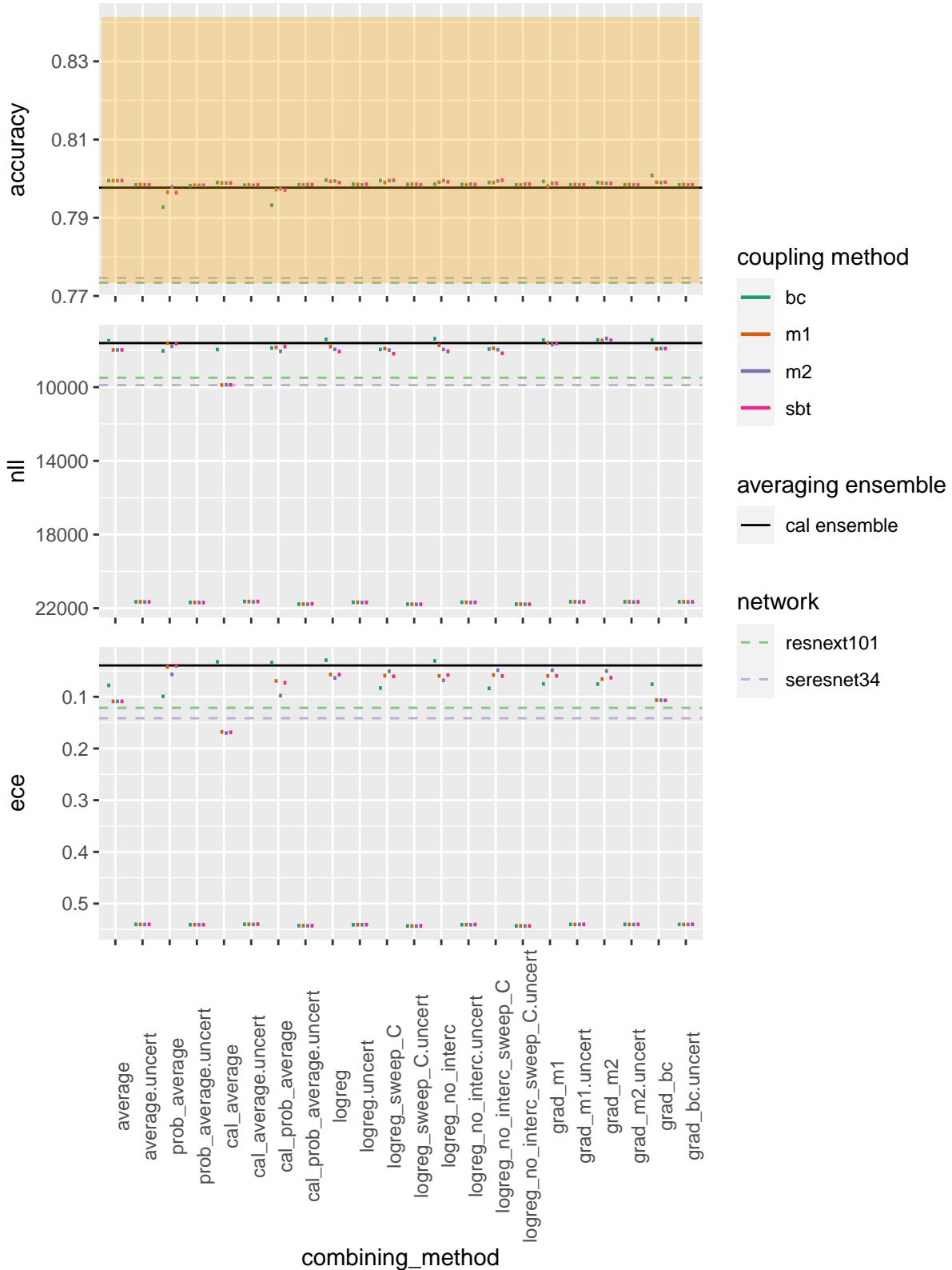
Ensemble metrics
Error inconsistency 0.212099999189377



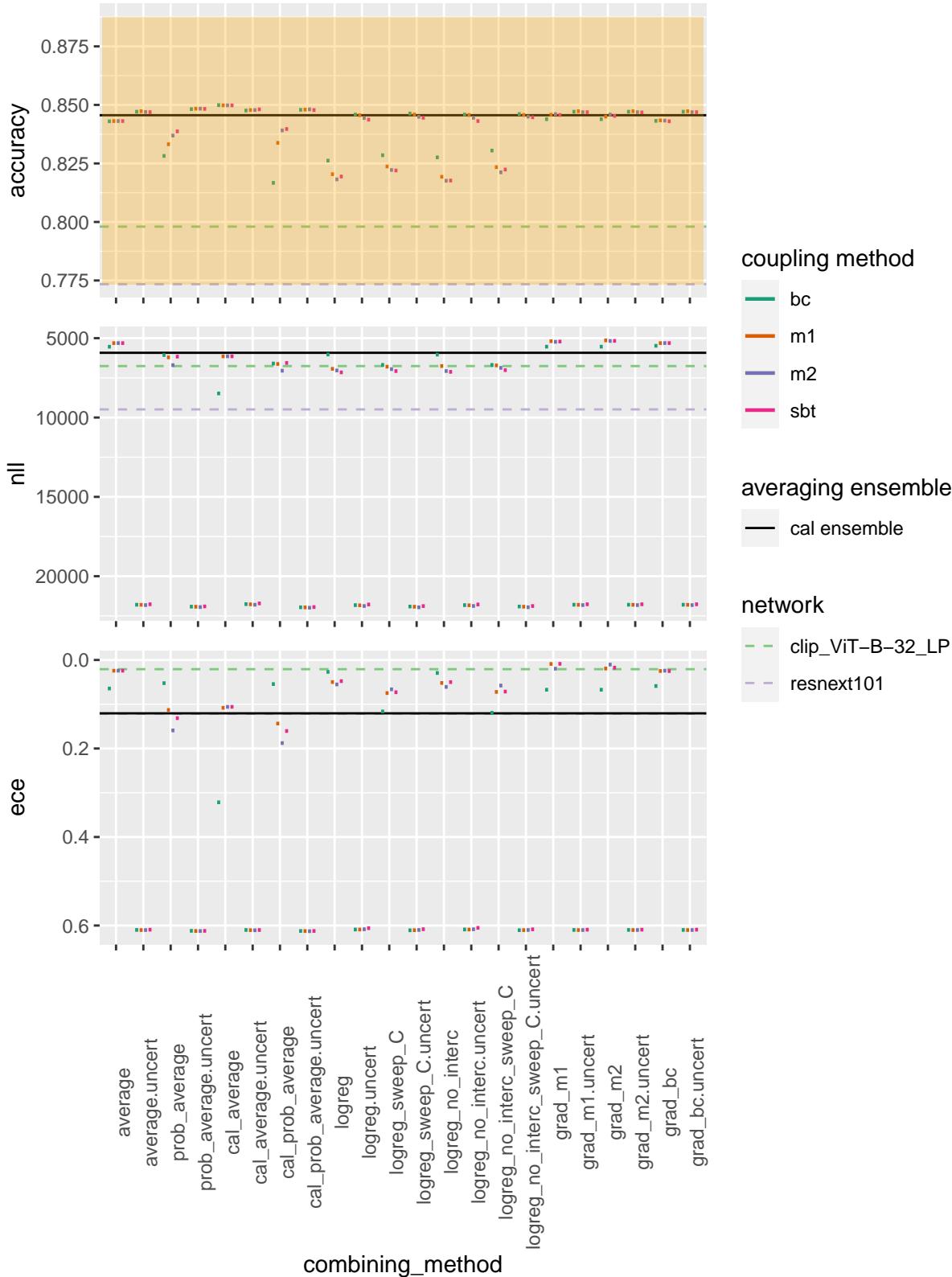
Ensemble metrics
Error inconsistency 0.151899993419647



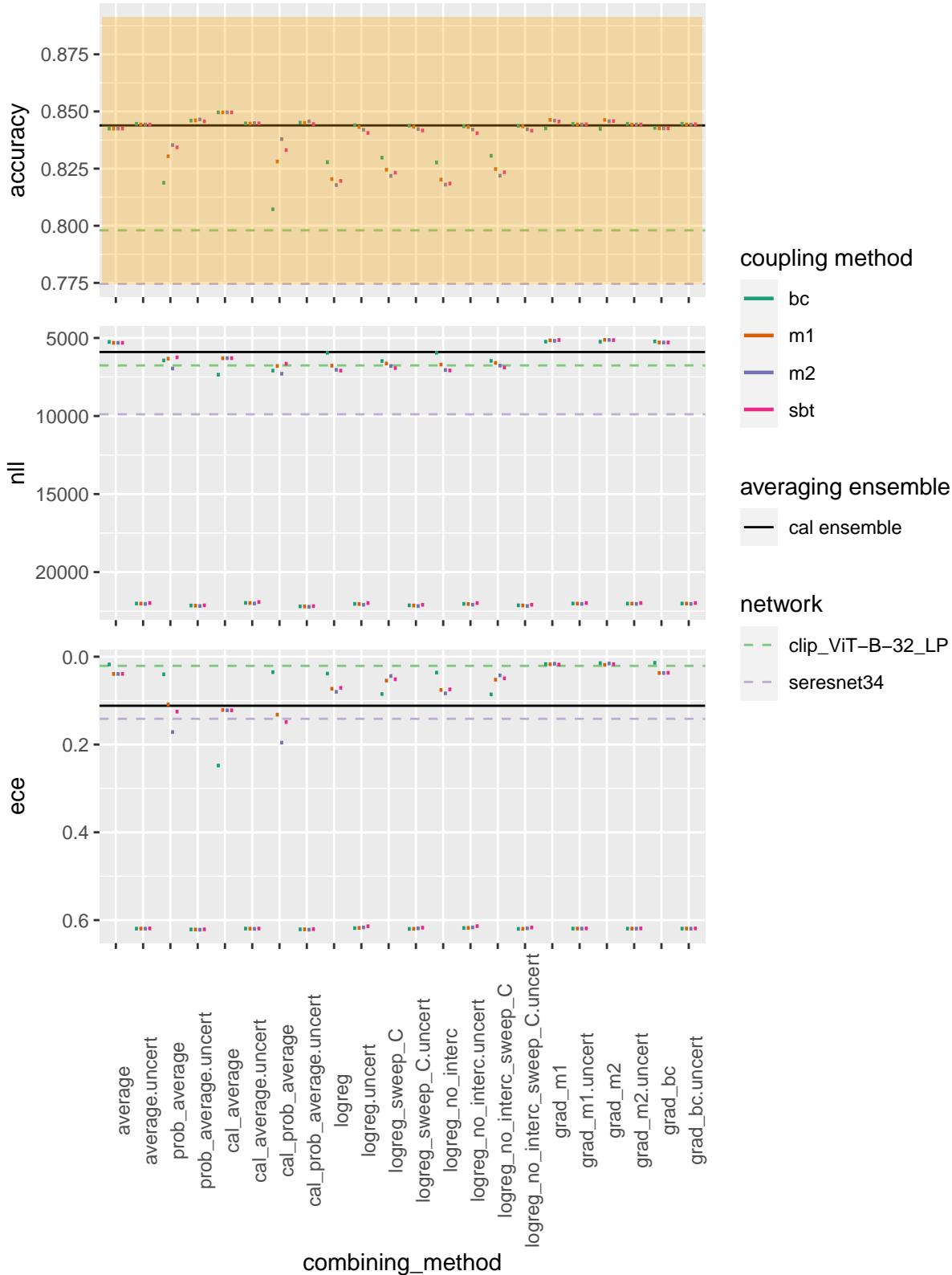
Ensemble metrics
Error inconsistency 0.134800001978874



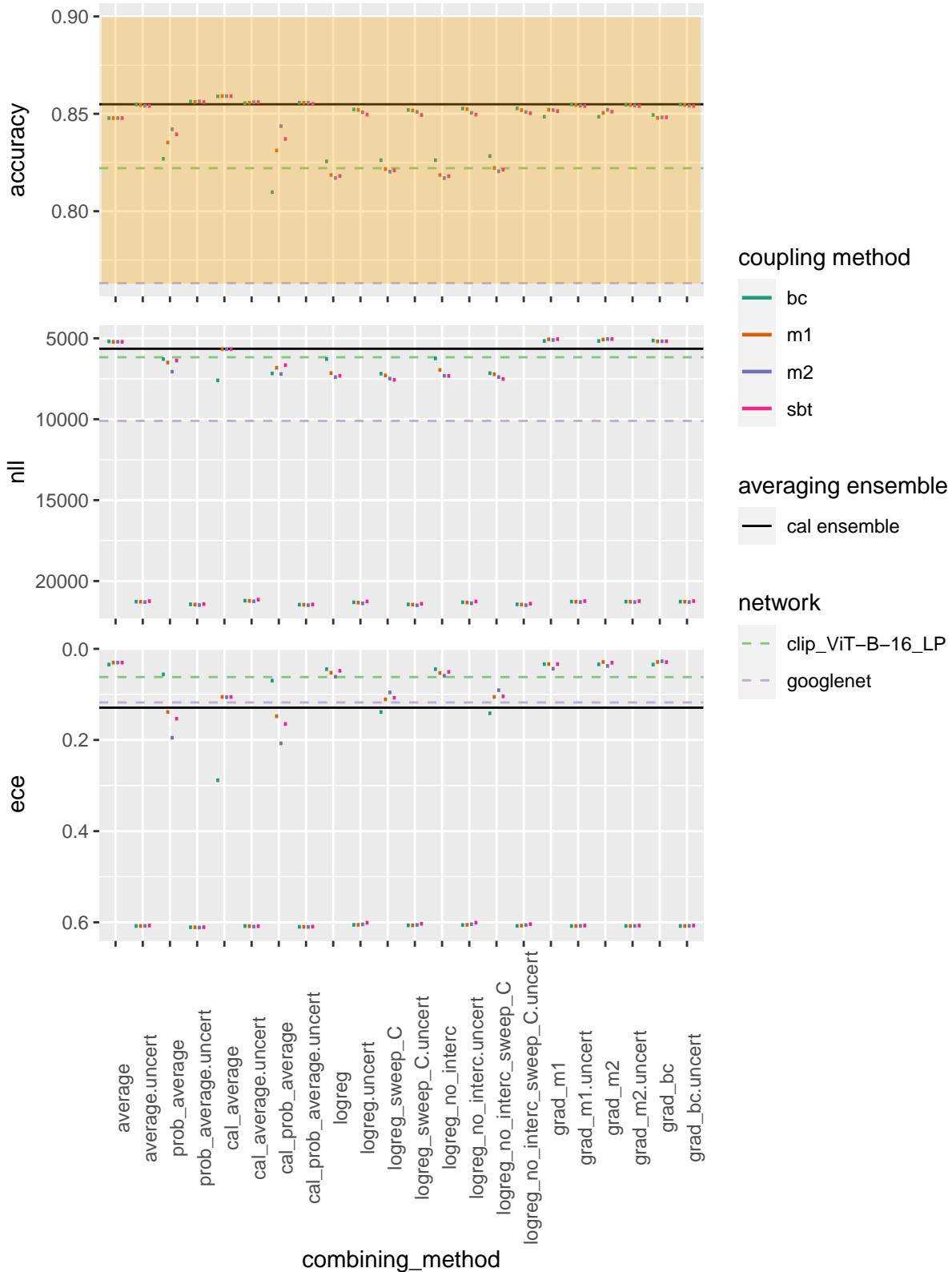
Ensemble metrics
Error inconsistency 0.203799992799759



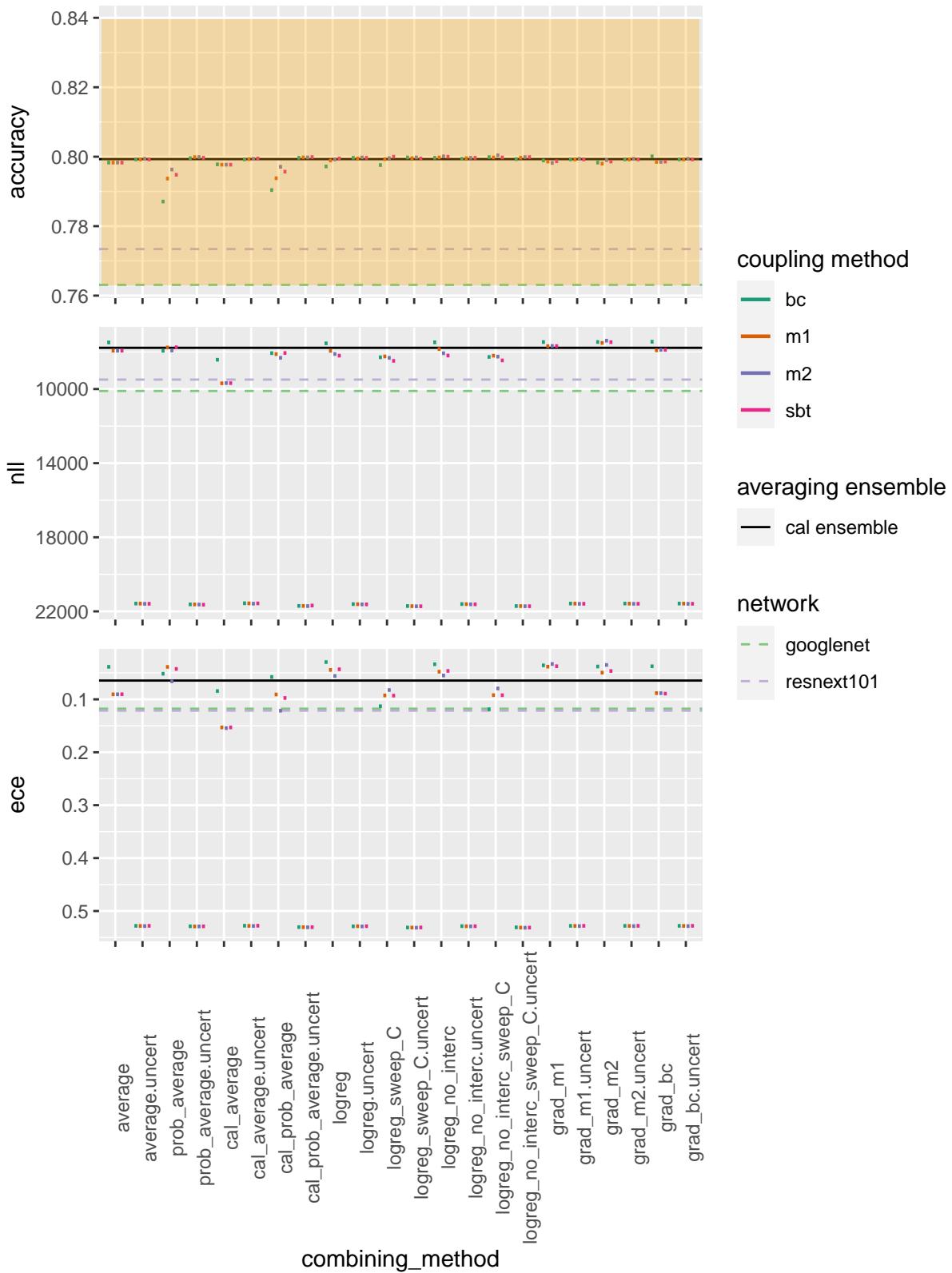
Ensemble metrics
Error inconsistency 0.210199996829033



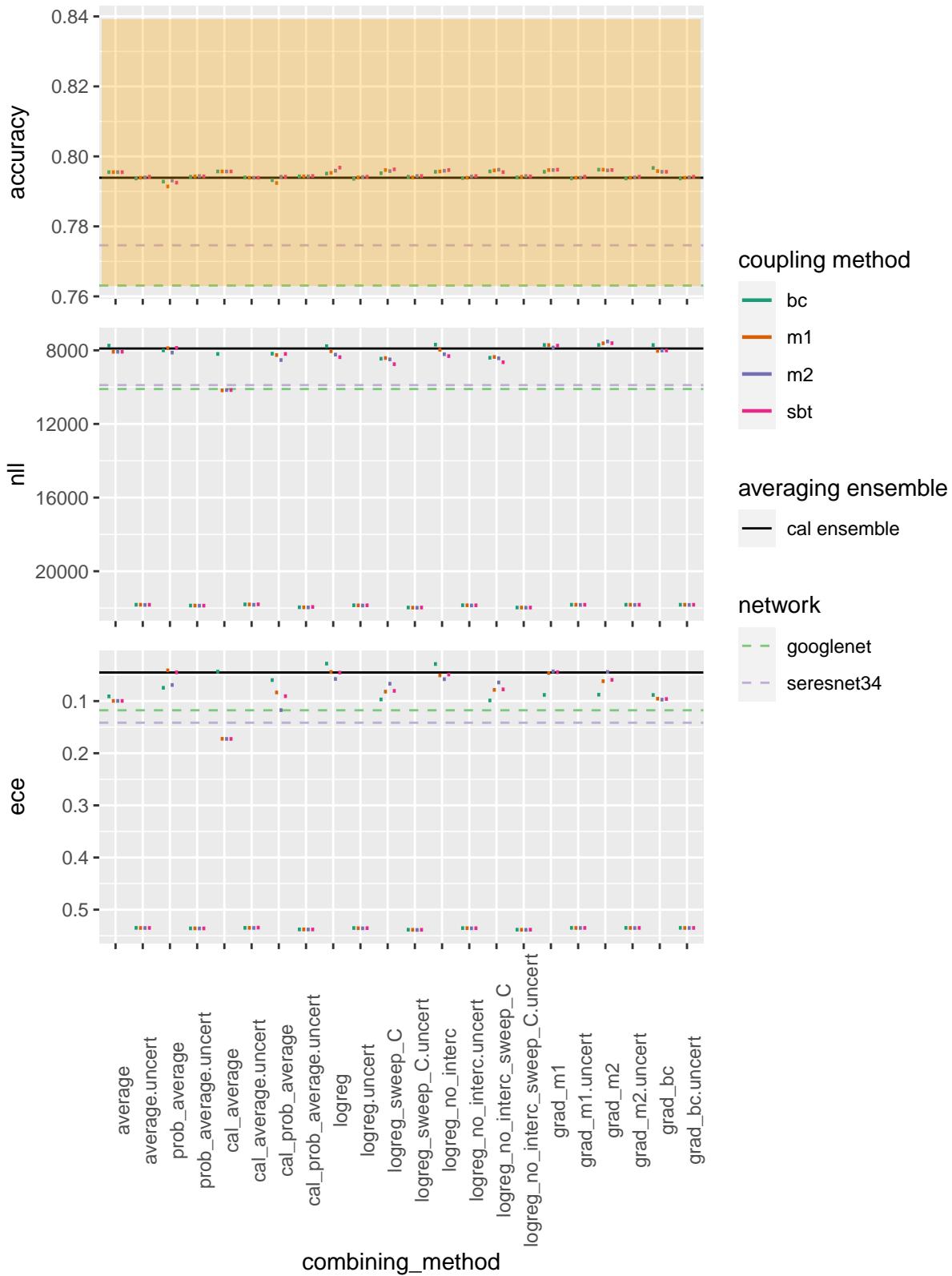
Ensemble metrics
Error inconsistency 0.214399993419647



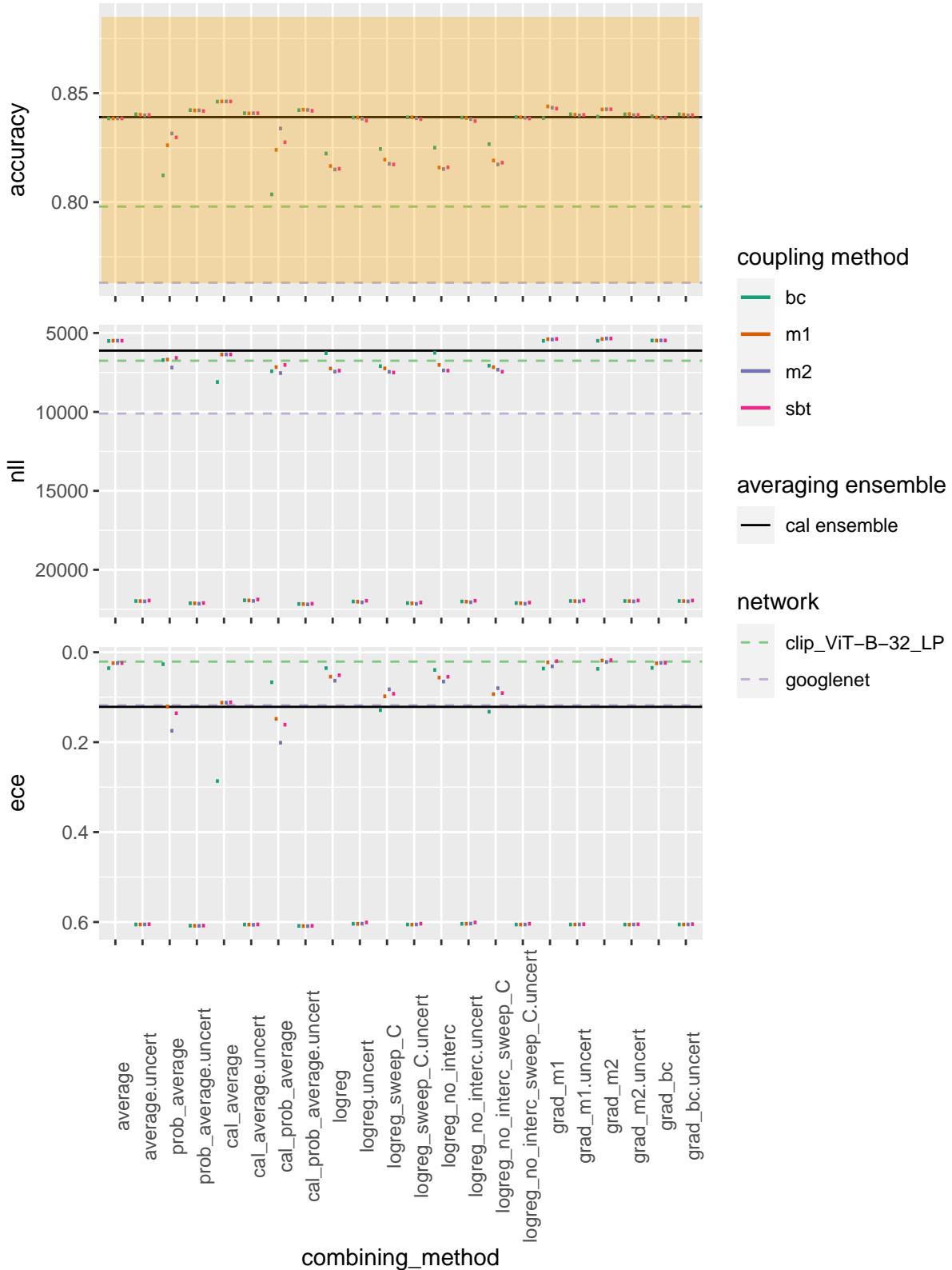
Ensemble metrics
Error inconsistency 0.142700001597404



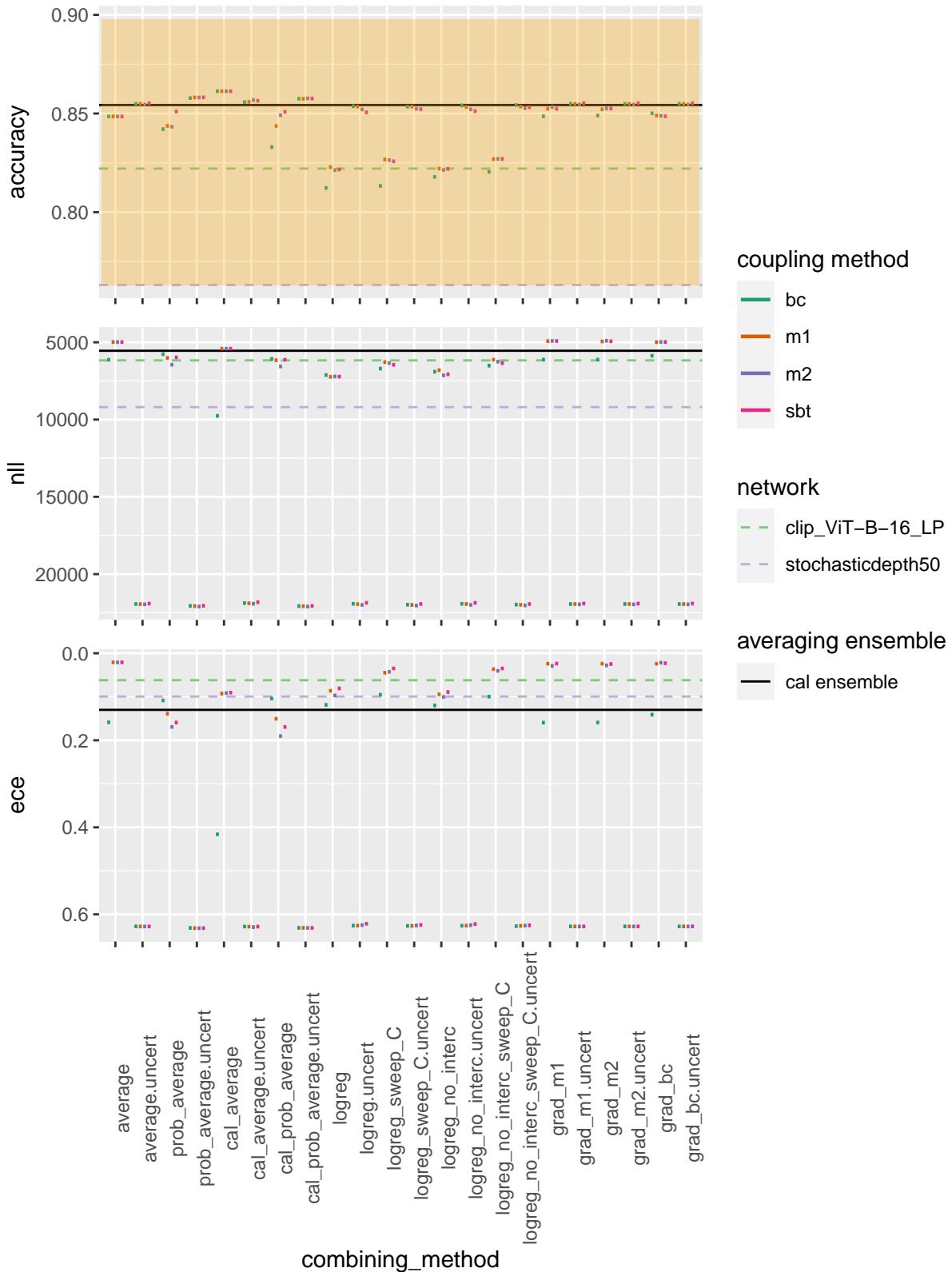
Ensemble metrics
Error inconsistency 0.140699997544289



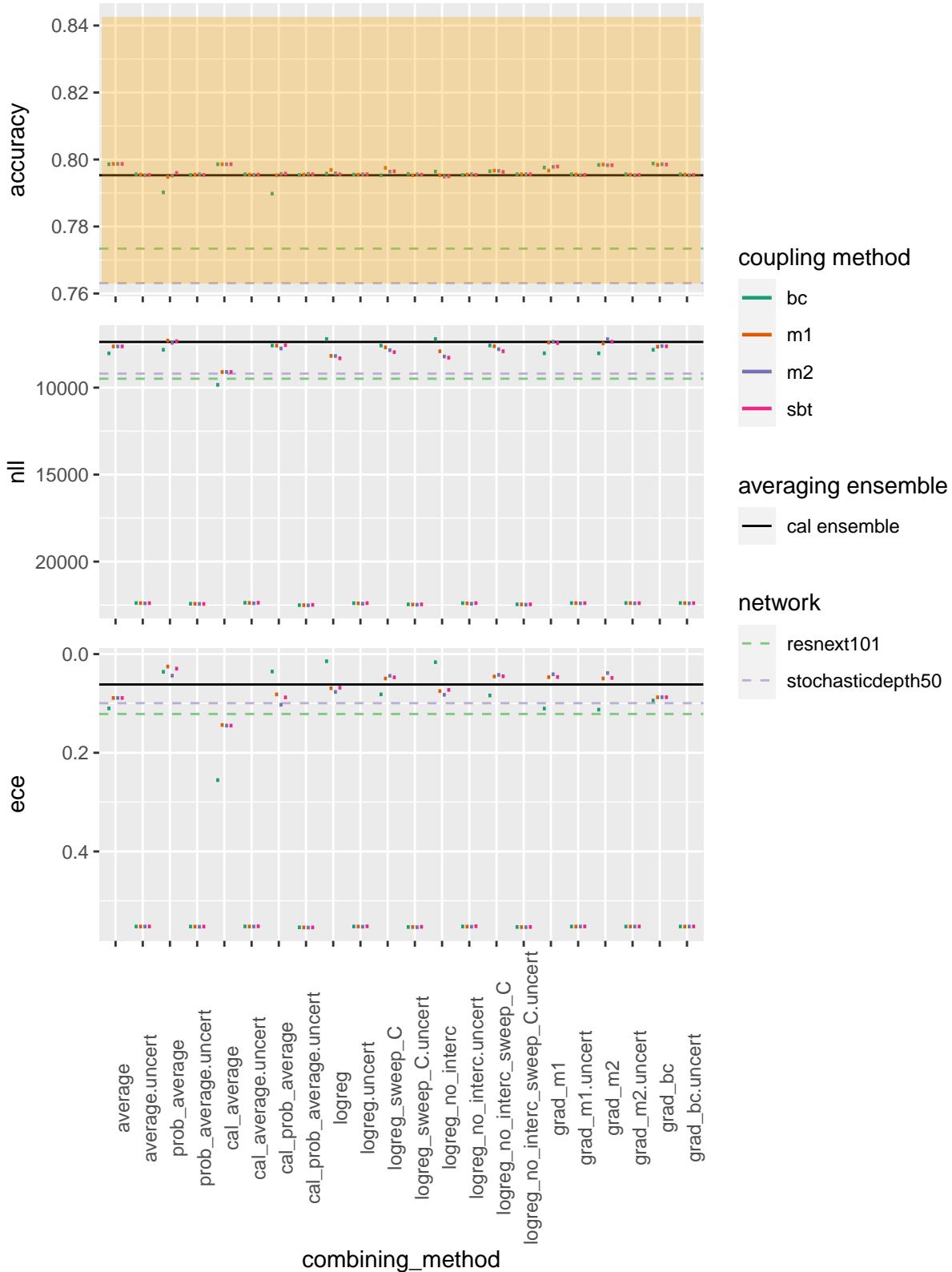
Ensemble metrics
Error inconsistency 0.208899989724159



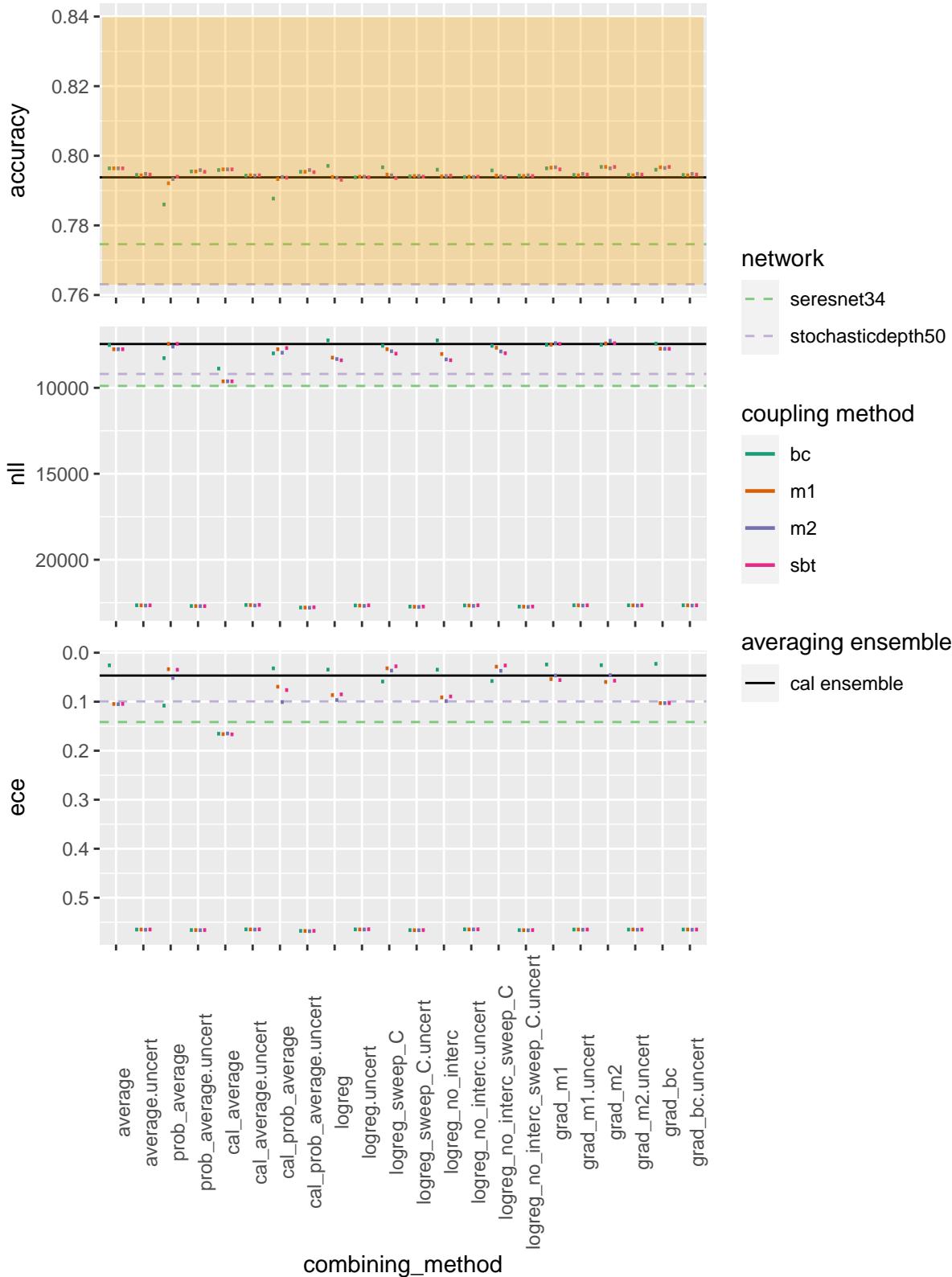
Ensemble metrics
Error inconsistency 0.210400000214577



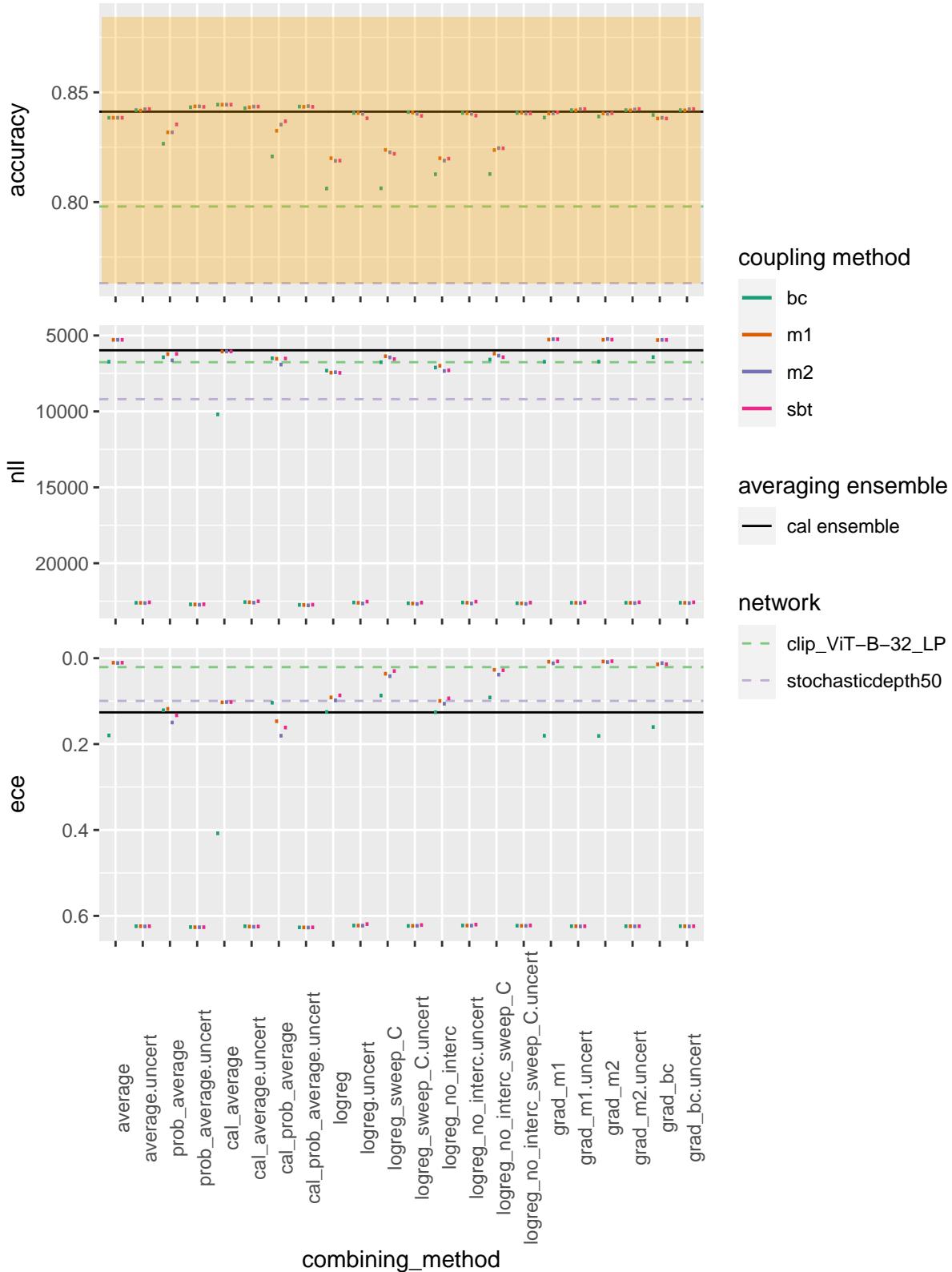
Ensemble metrics
Error inconsistency 0.148699998855591



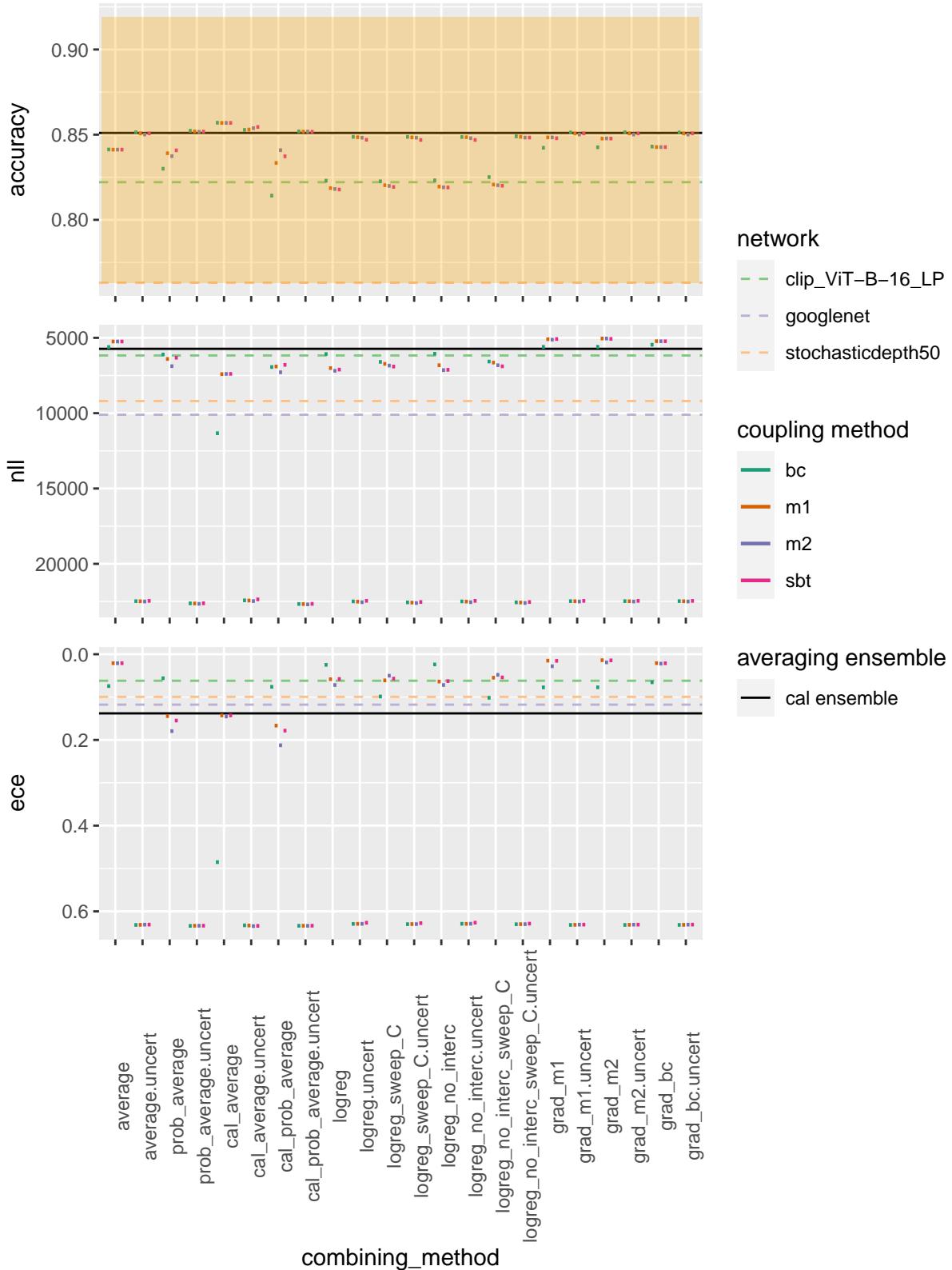
Ensemble metrics
Error inconsistency 0.142099991440773



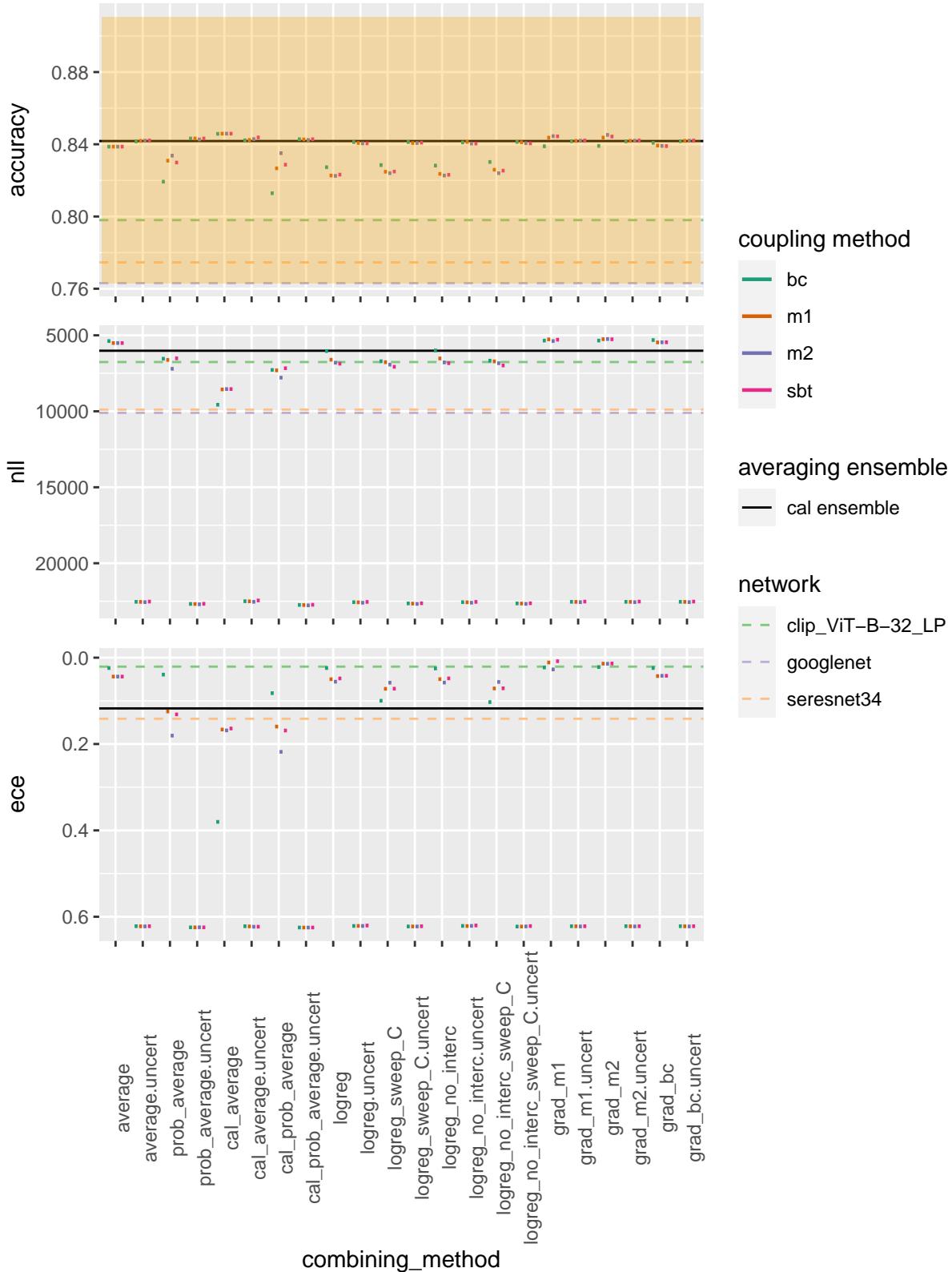
Ensemble metrics
Error inconsistency 0.207699999213219



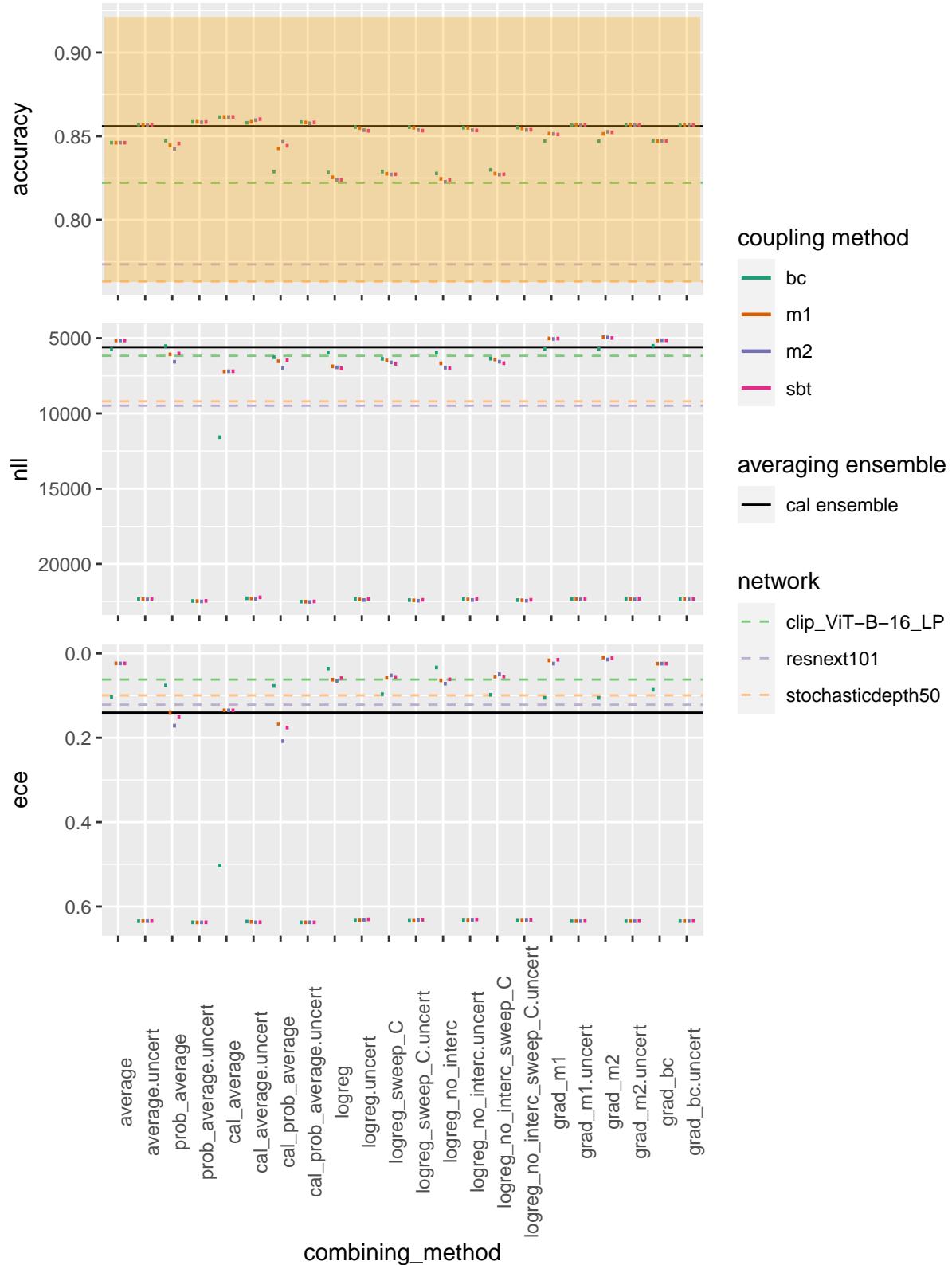
Ensemble metrics
Error inconsistency 0.286799997091293



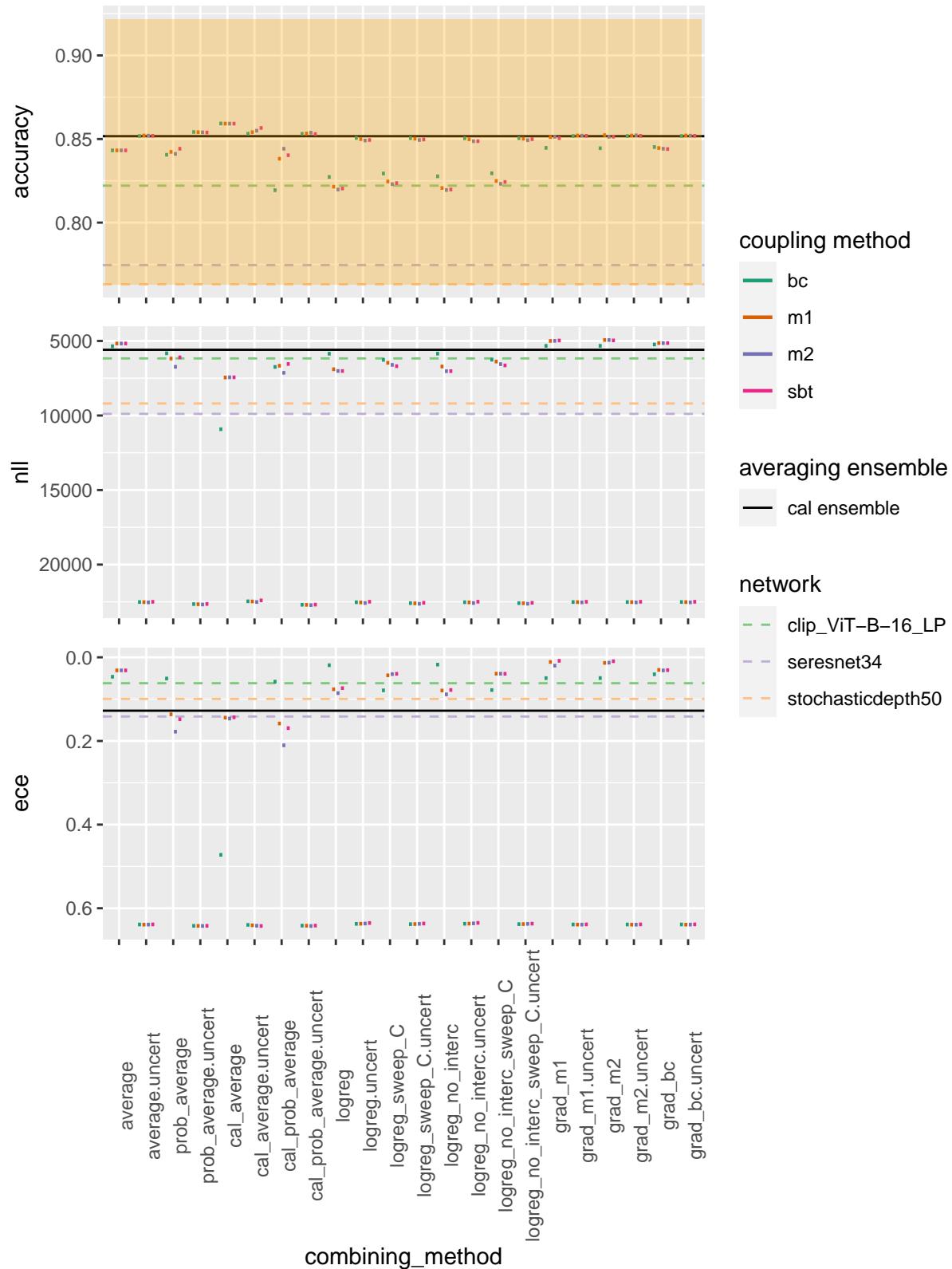
Ensemble metrics
Error inconsistency 0.27989998459816



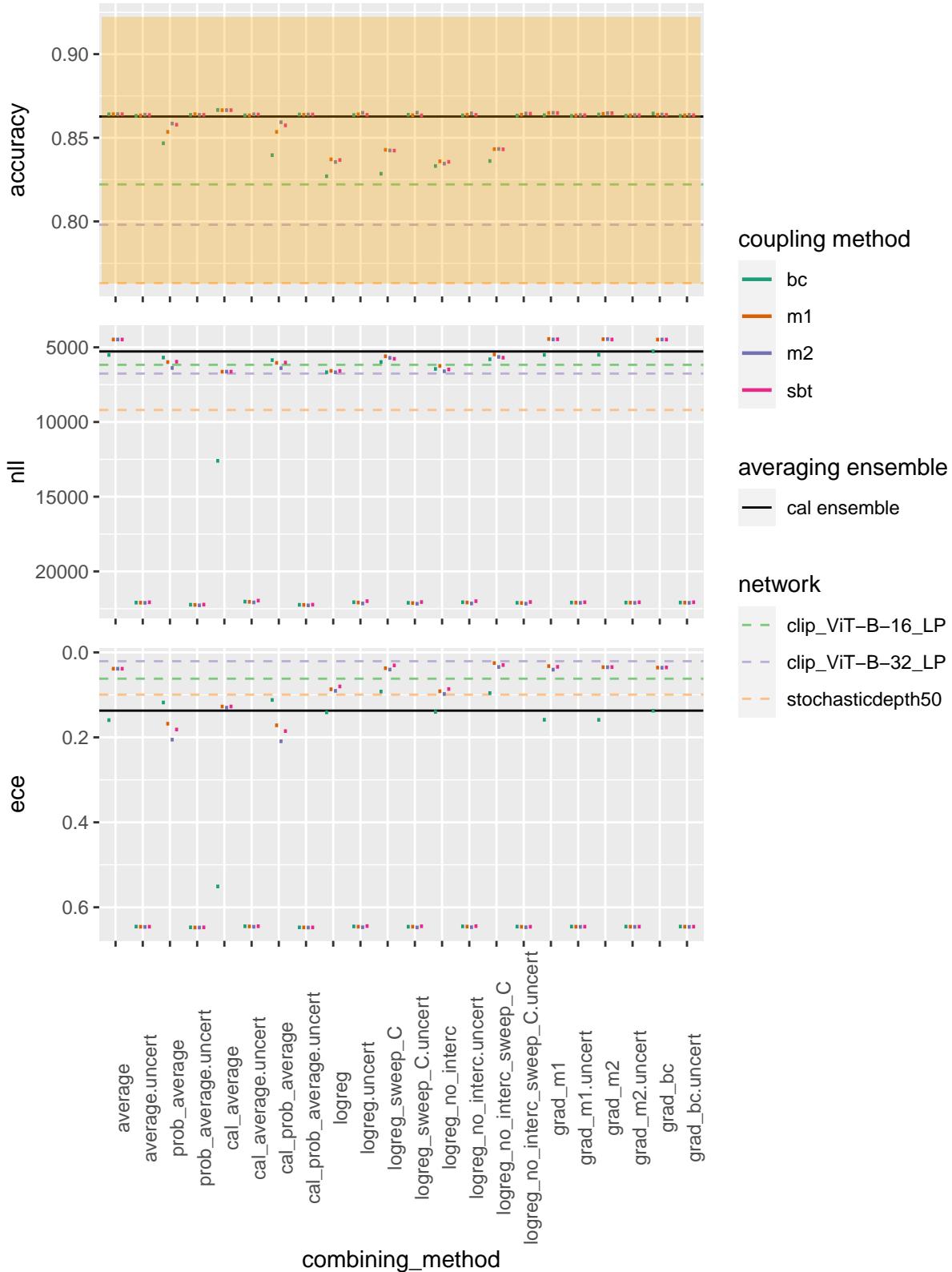
Ensemble metrics
Error inconsistency 0.282700002193451



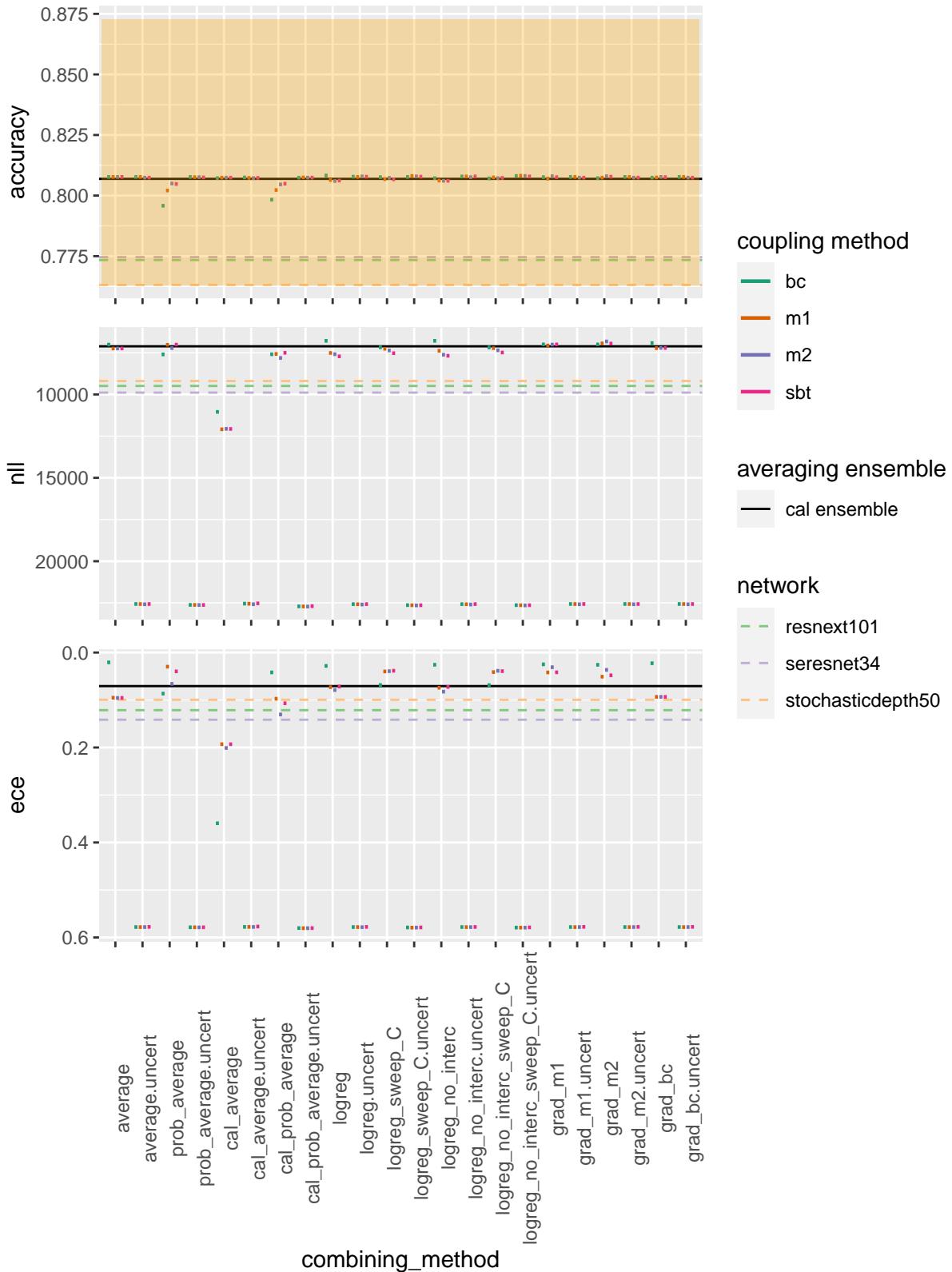
Ensemble metrics
Error inconsistency 0.282299995422363



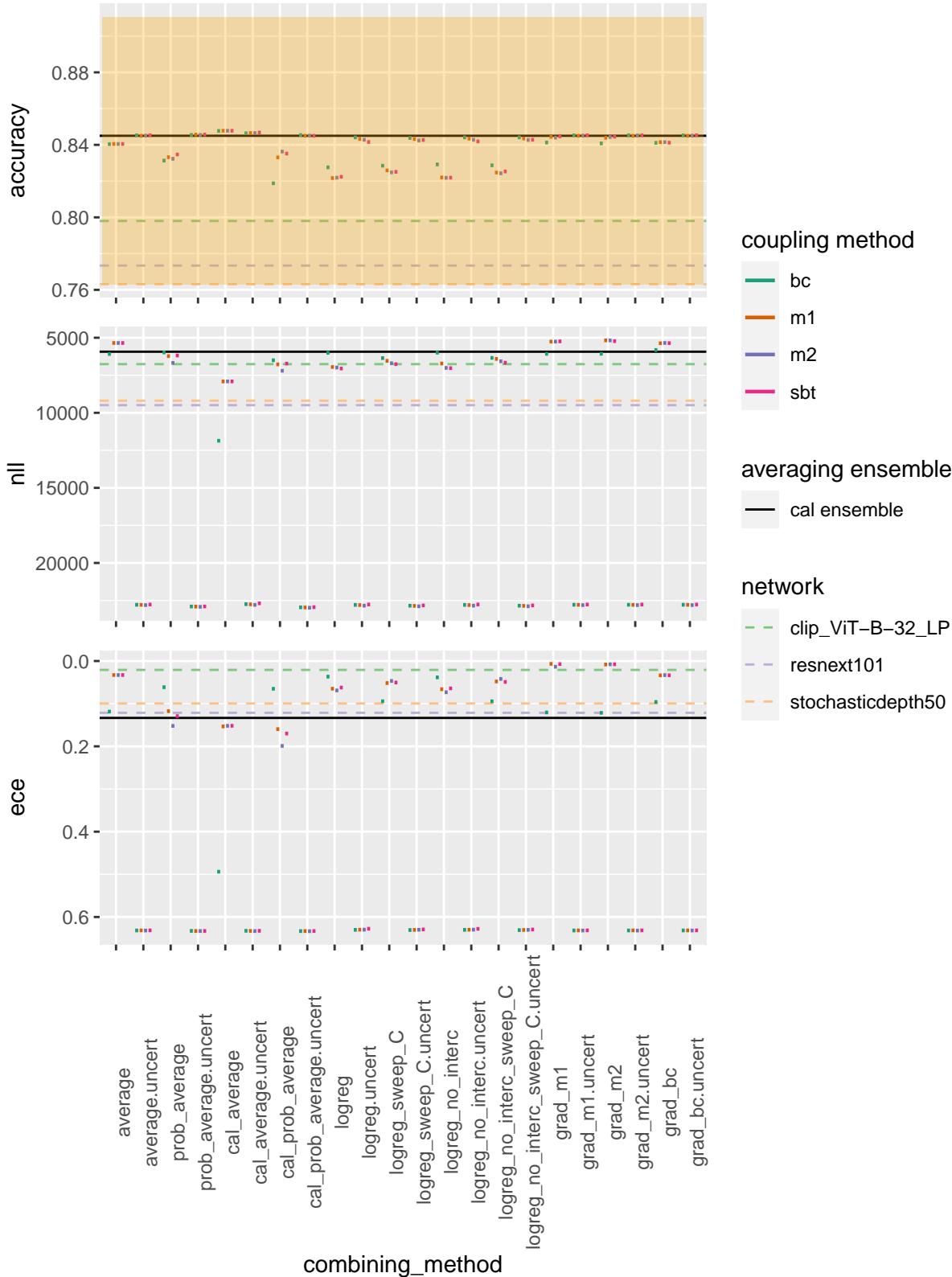
Ensemble metrics
Error inconsistency 0.284999996423721



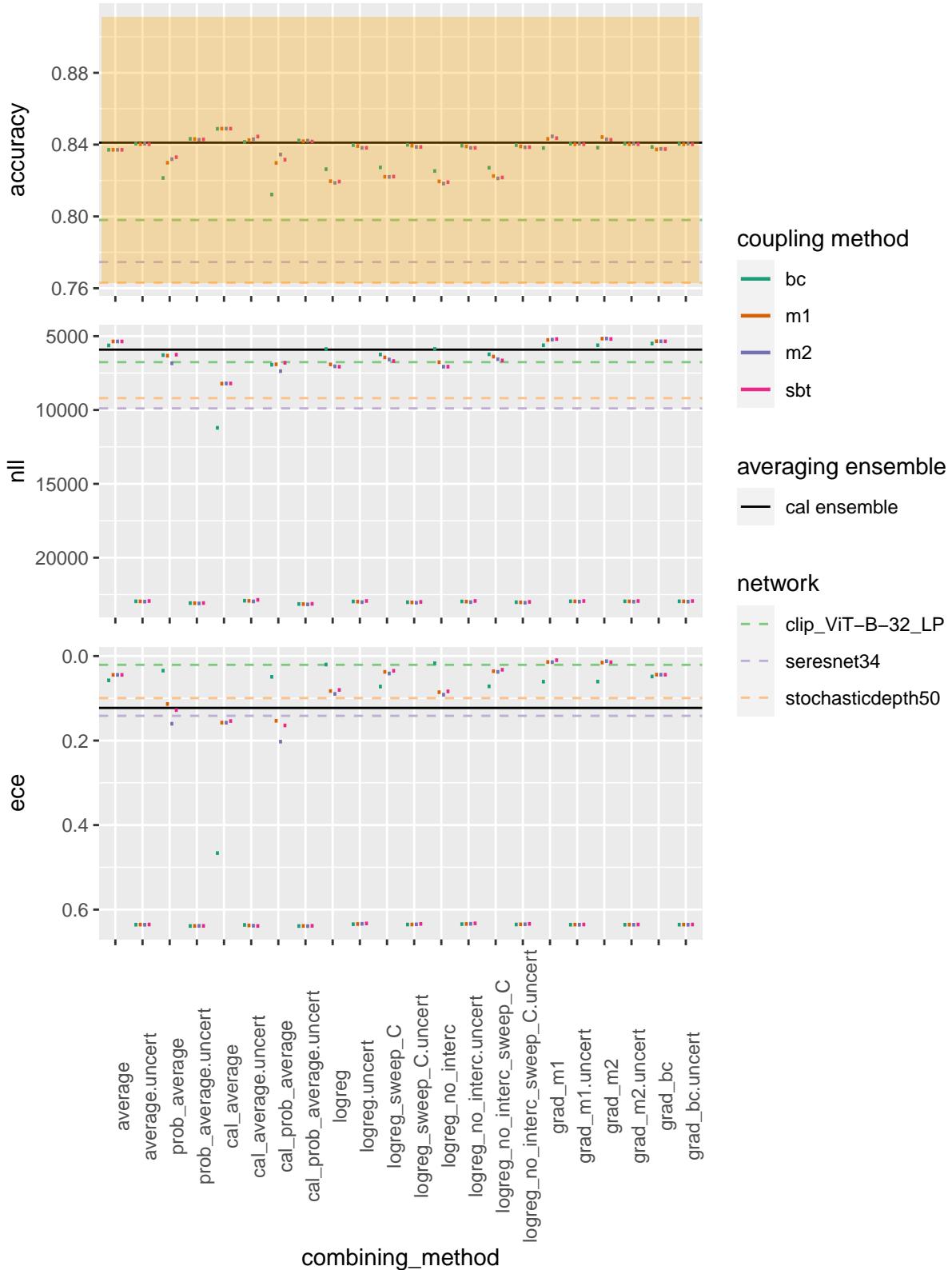
Ensemble metrics
Error inconsistency 0.212799996137619



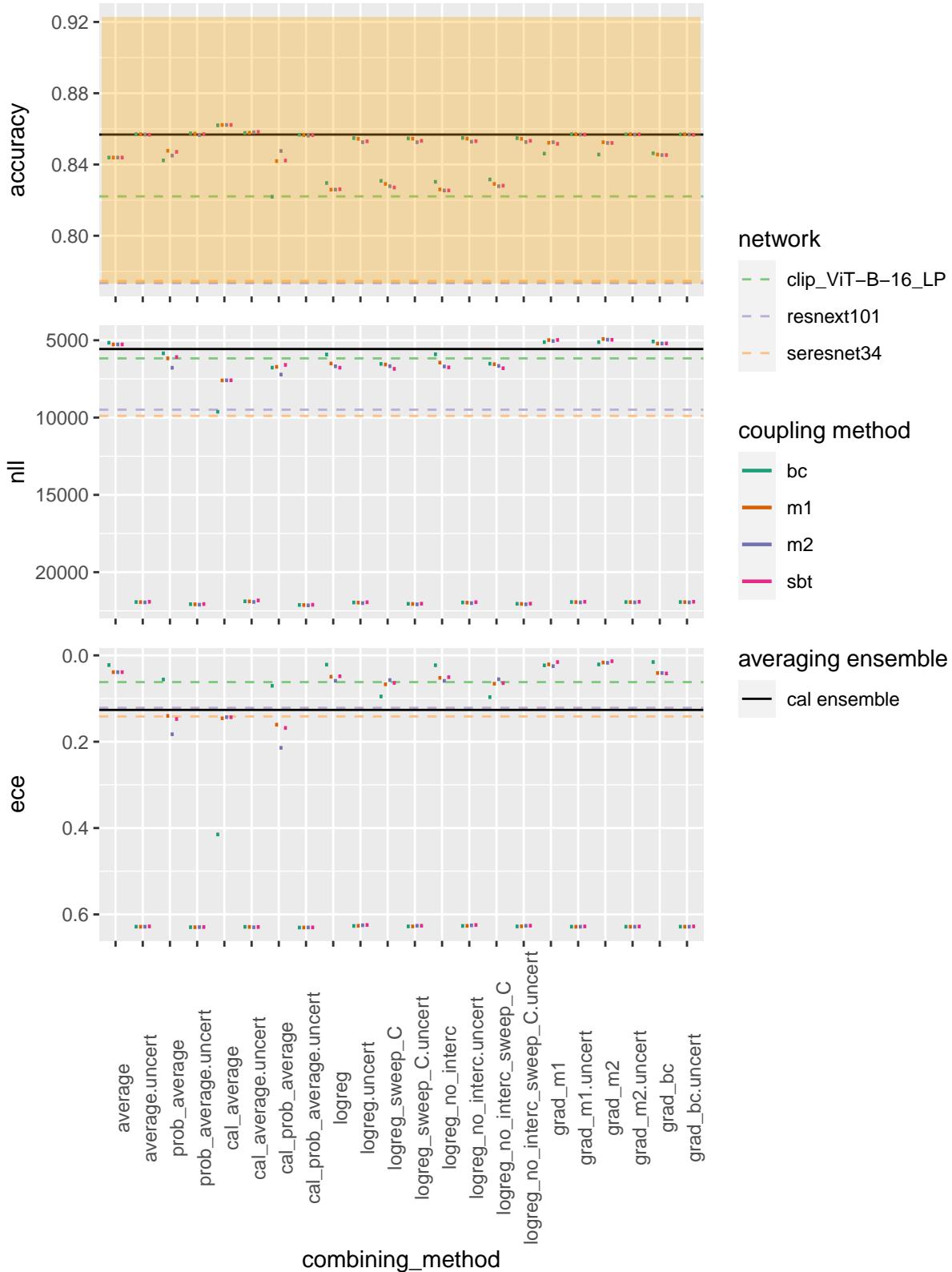
Ensemble metrics
Error inconsistency 0.280099987983704



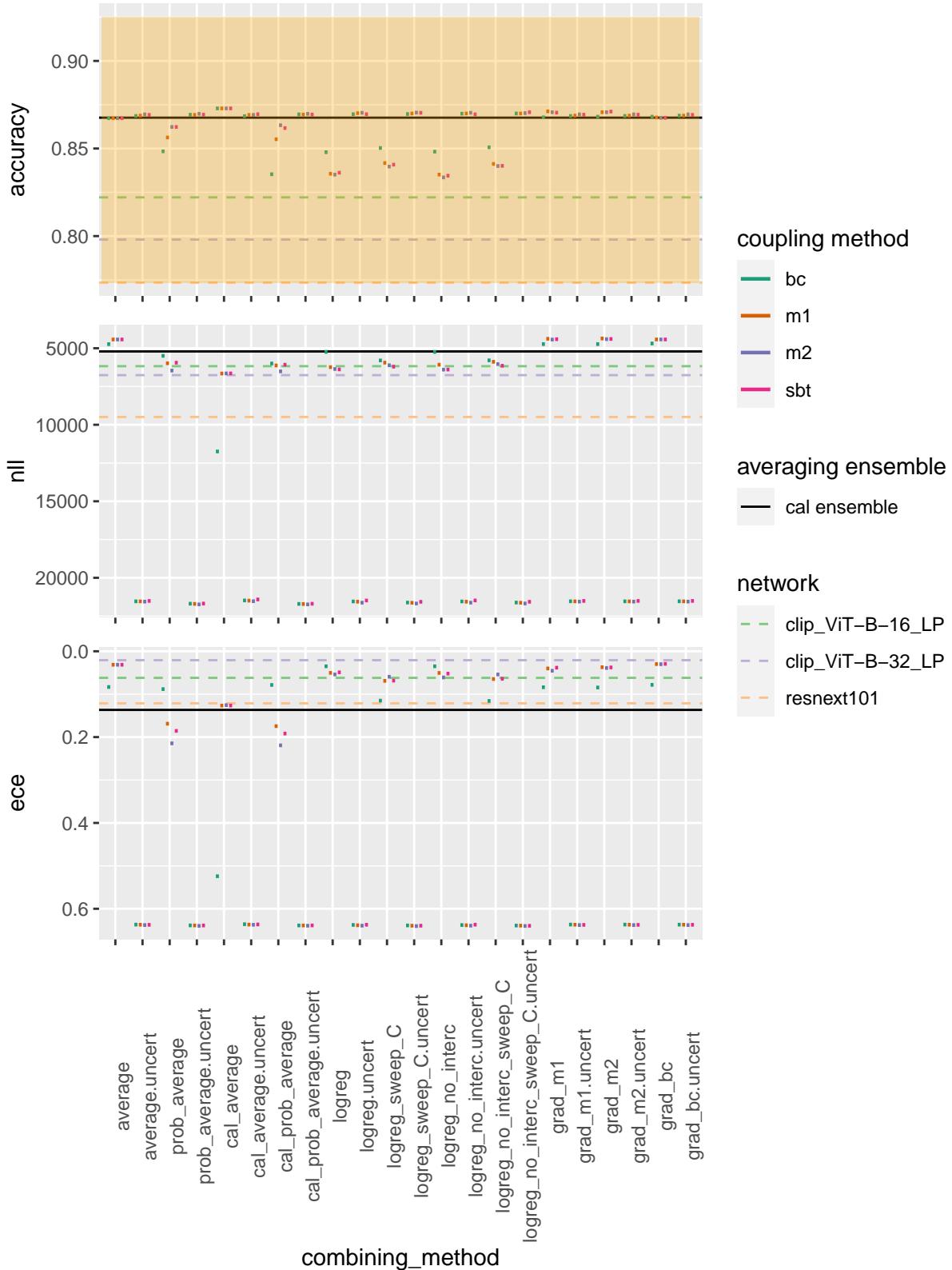
Ensemble metrics
Error inconsistency 0.280000001192093



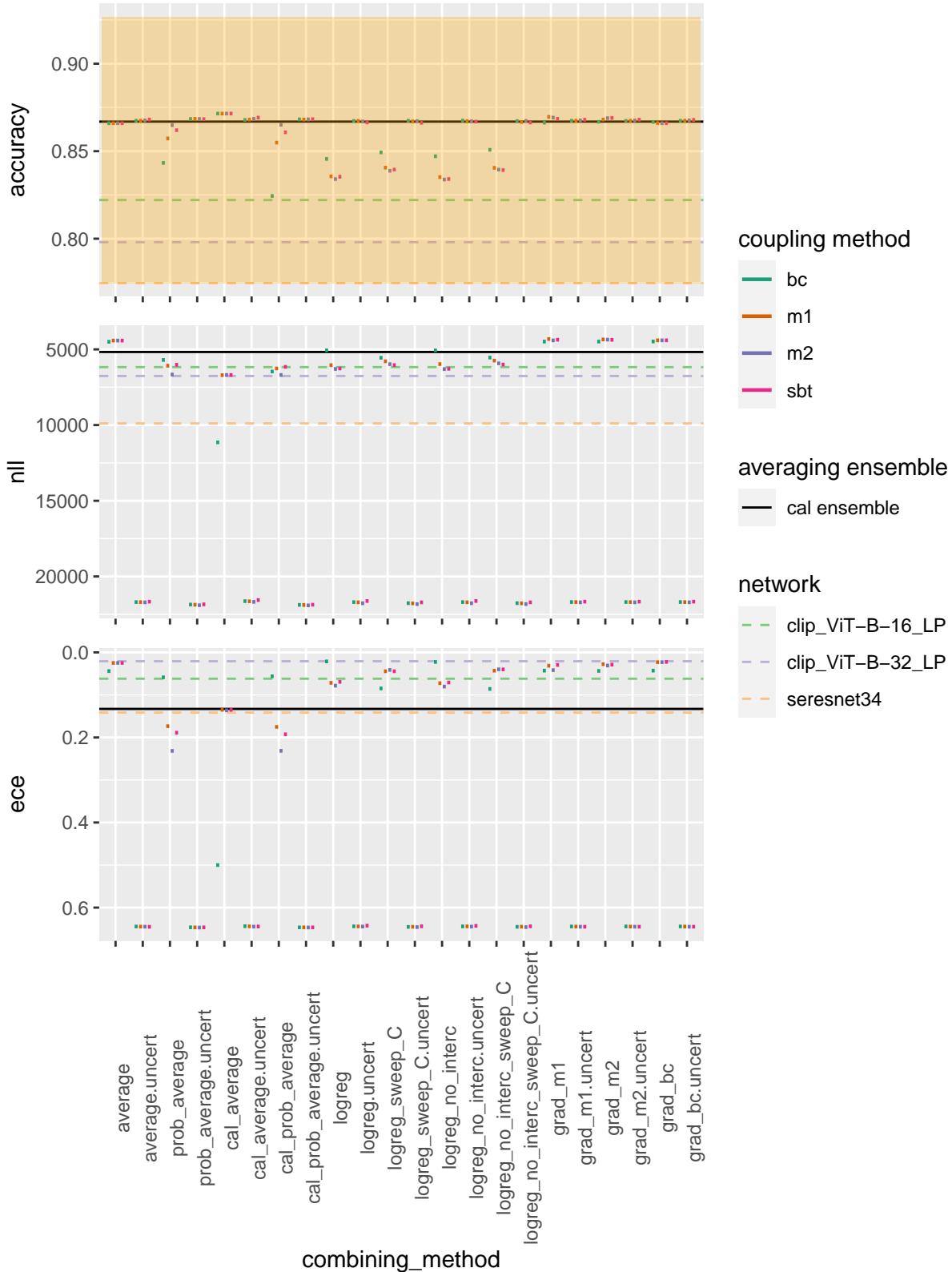
Ensemble metrics
Error inconsistency 0.276600003242493



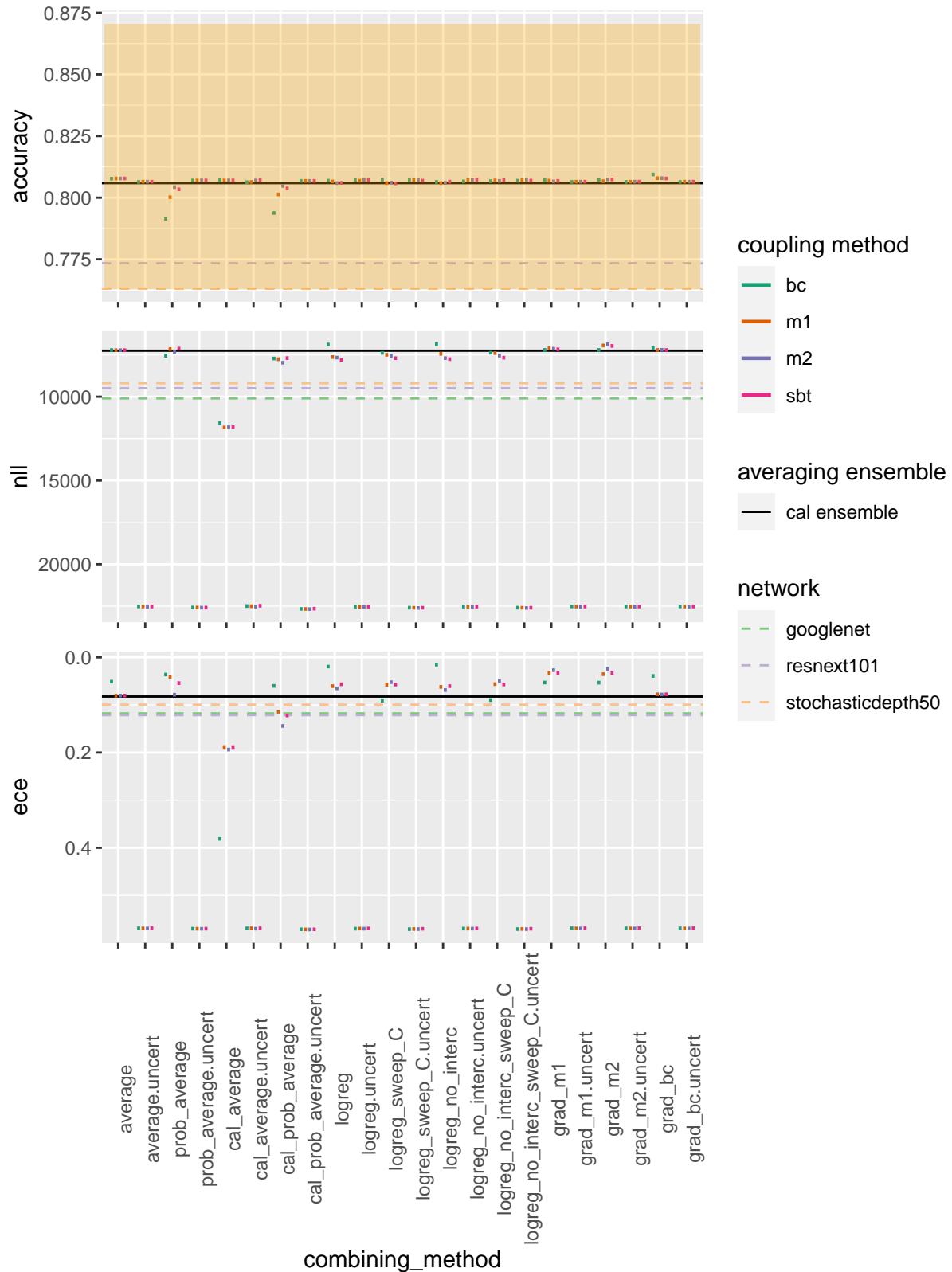
Ensemble metrics
Error inconsistency 0.28099998831749



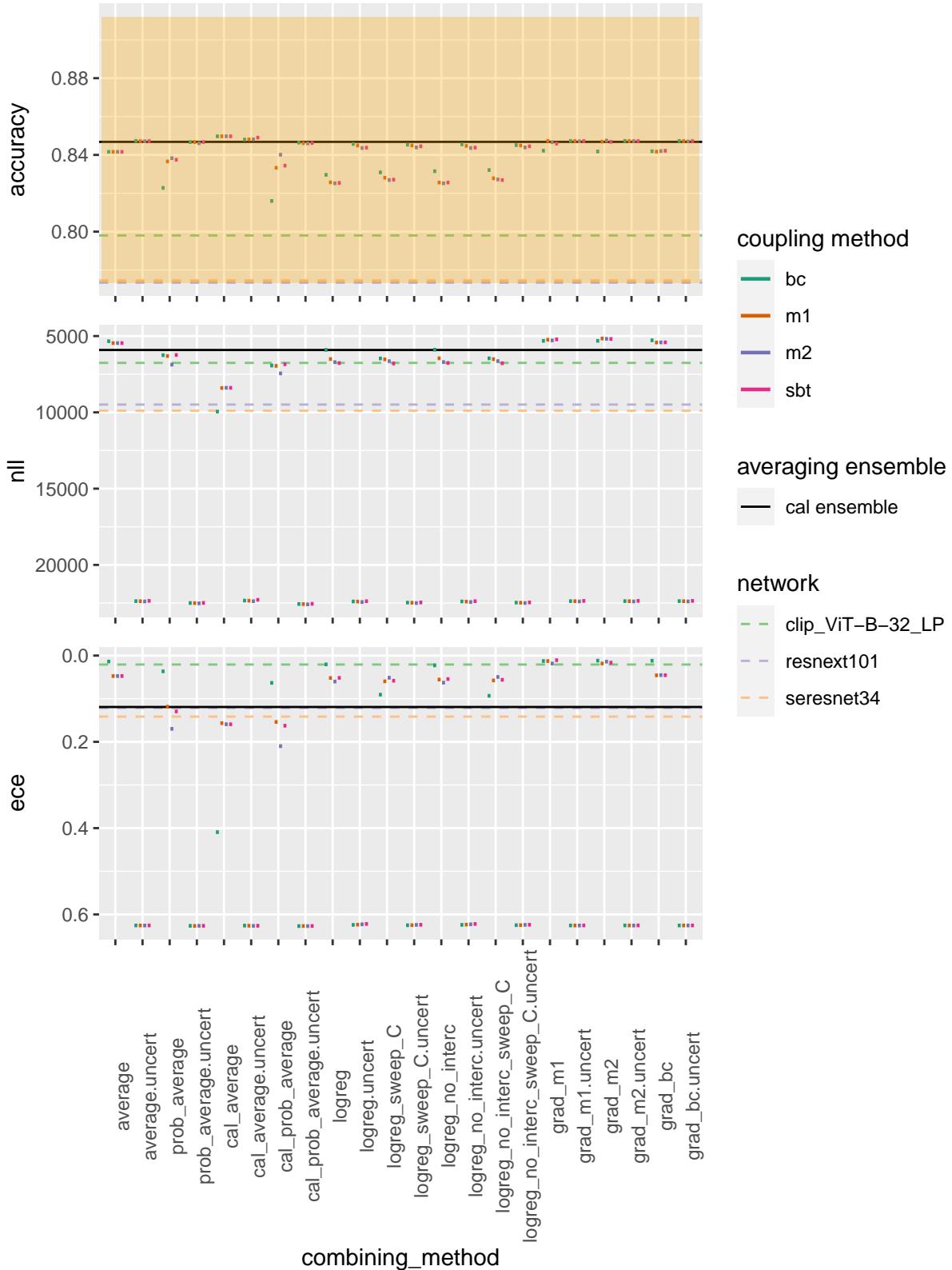
Ensemble metrics
Error inconsistency 0.287099987268448



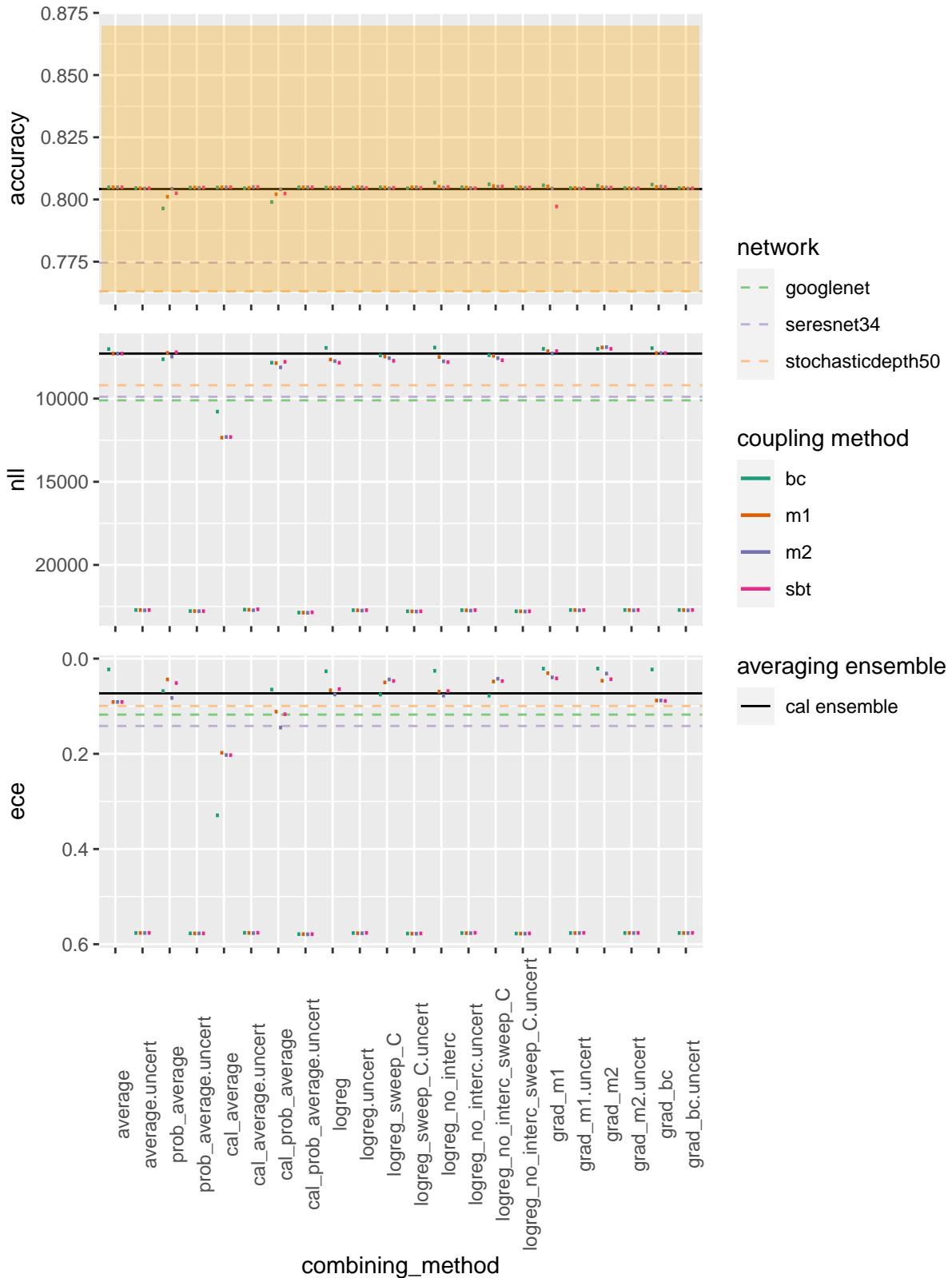
Ensemble metrics
Error inconsistency 0.220100000500679



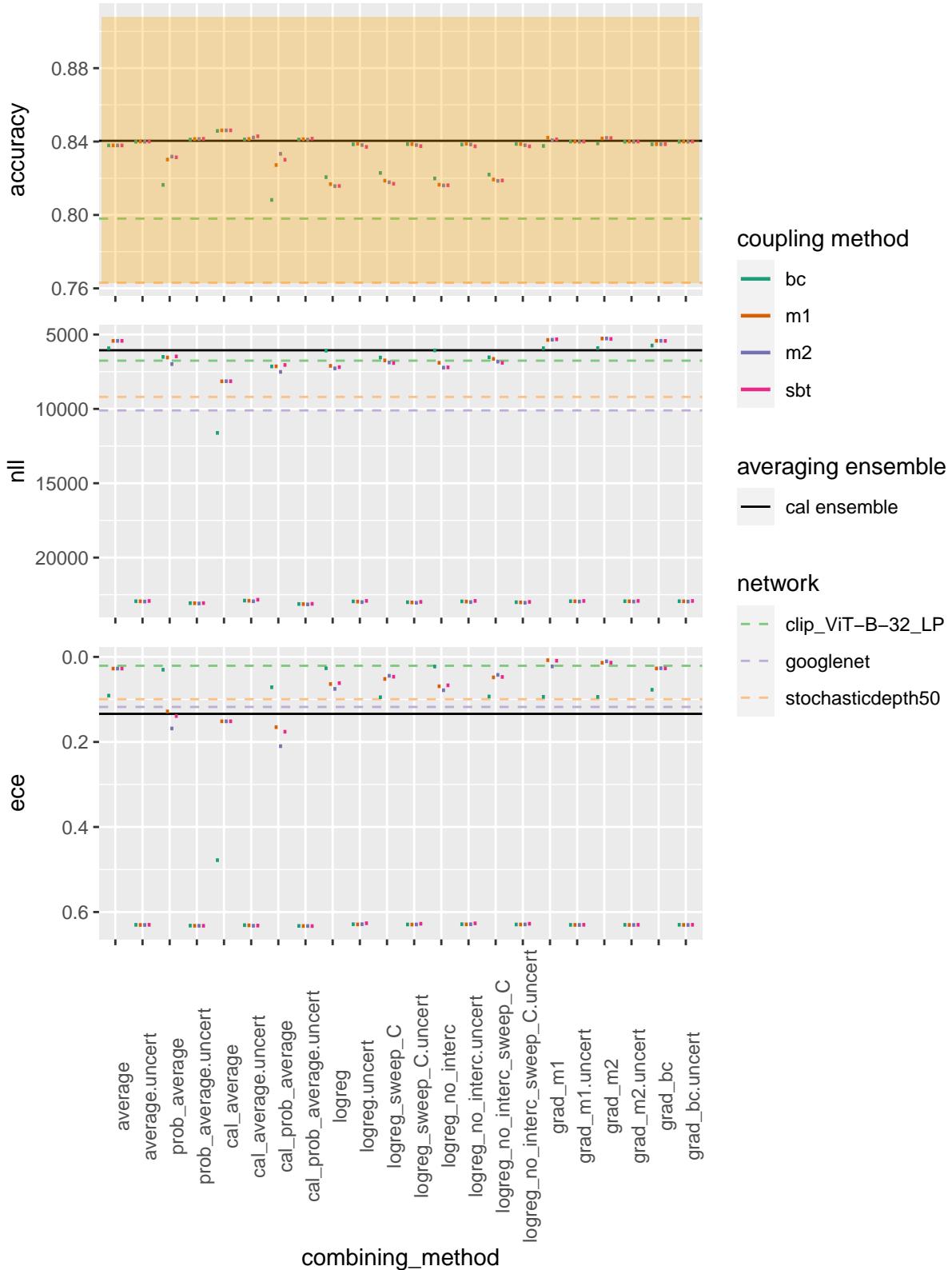
Ensemble metrics
Error inconsistency 0.274399995803833



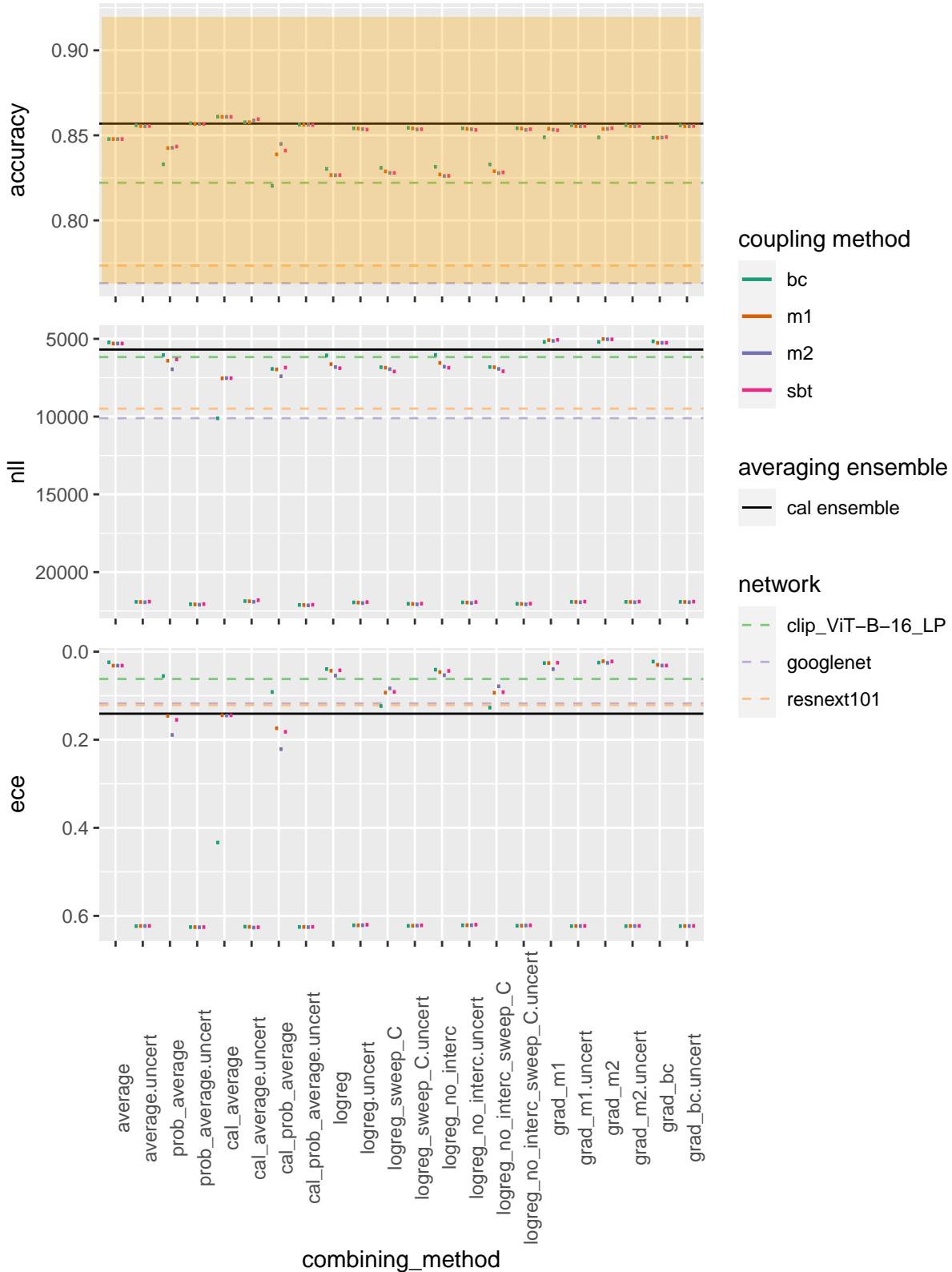
Ensemble metrics
Error inconsistency 0.215799987316132



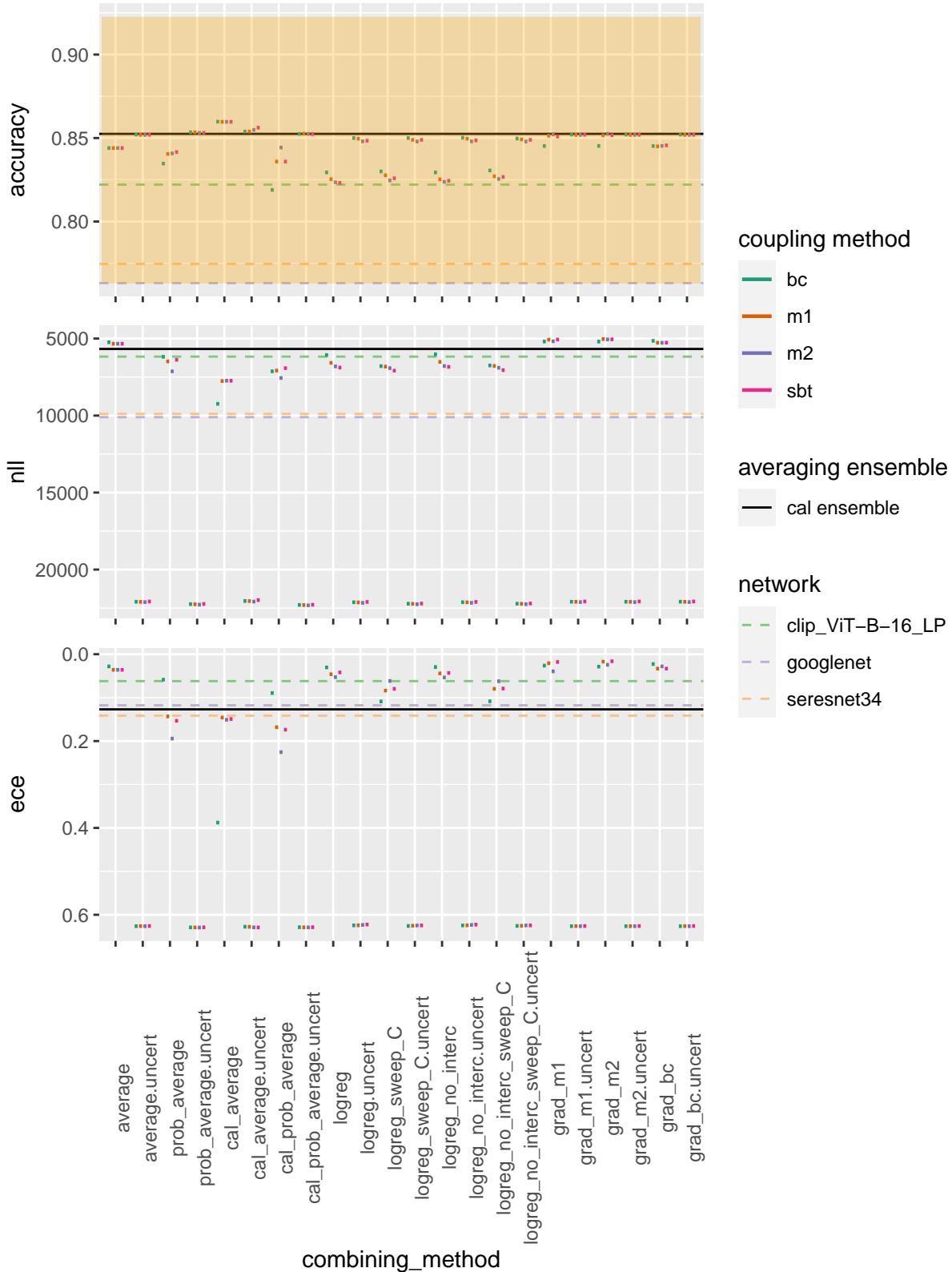
Ensemble metrics
Error inconsistency 0.282700002193451



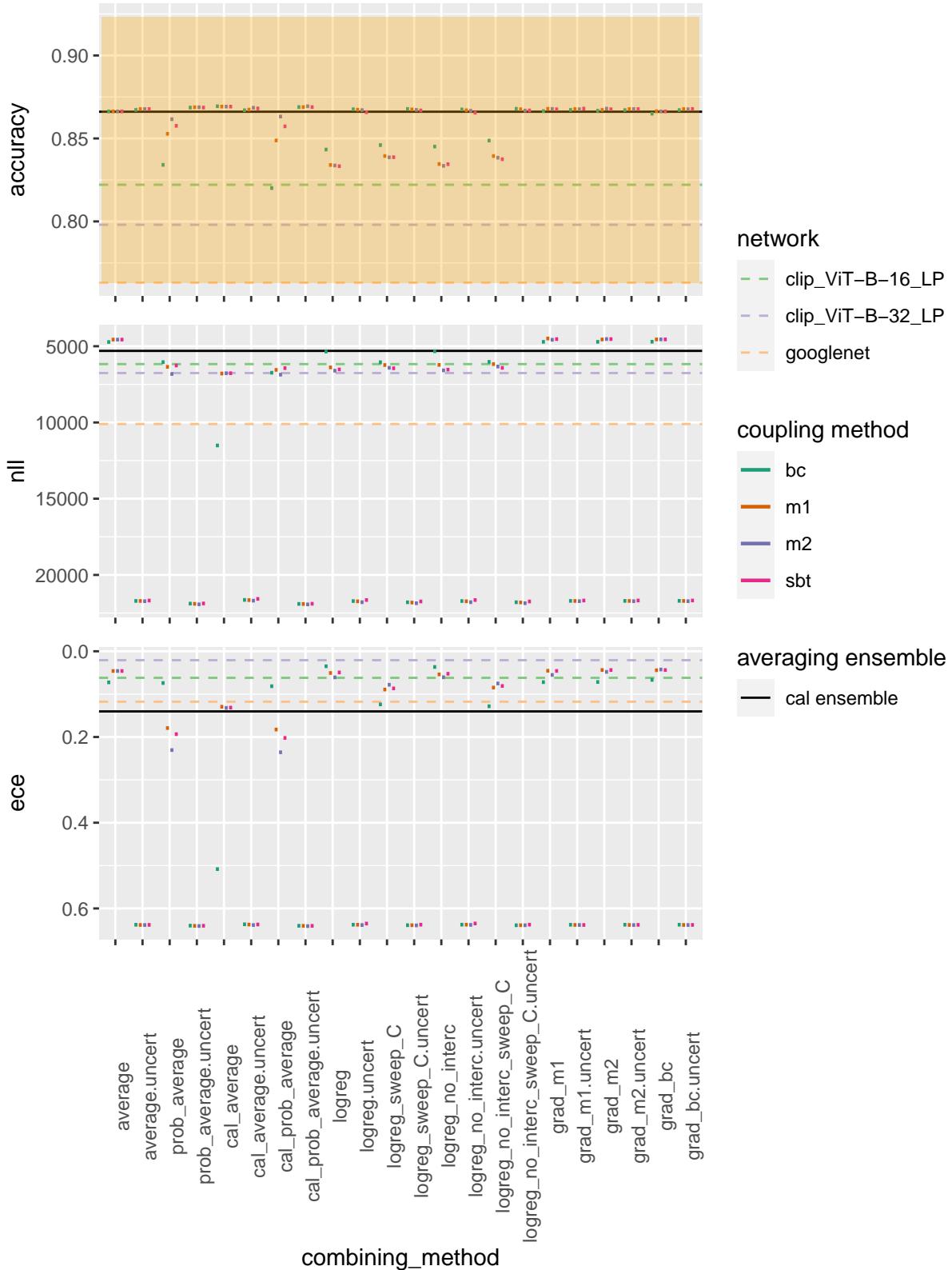
Ensemble metrics
Error inconsistency 0.281699985265732



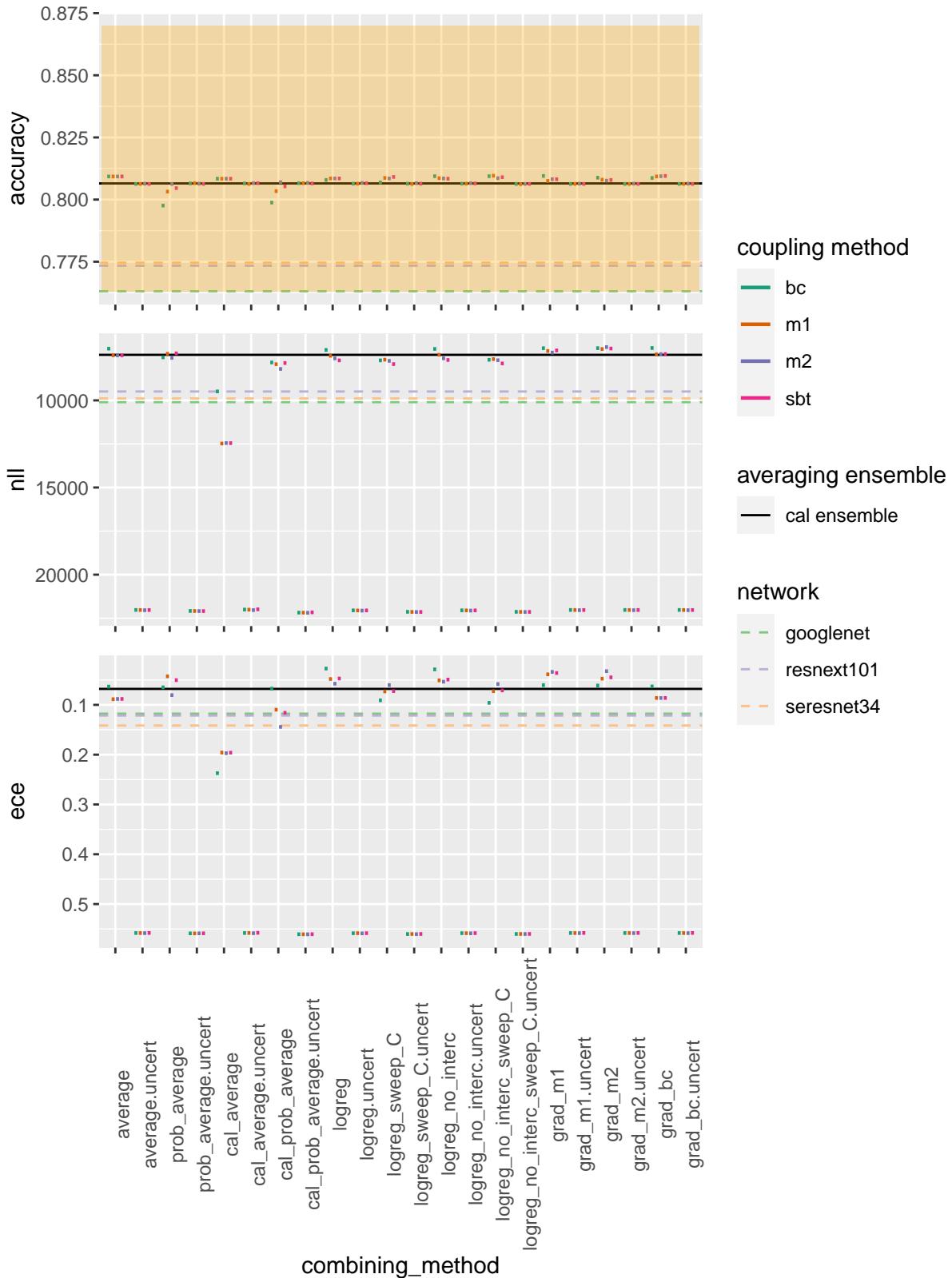
Ensemble metrics
Error inconsistency 0.283600002527237



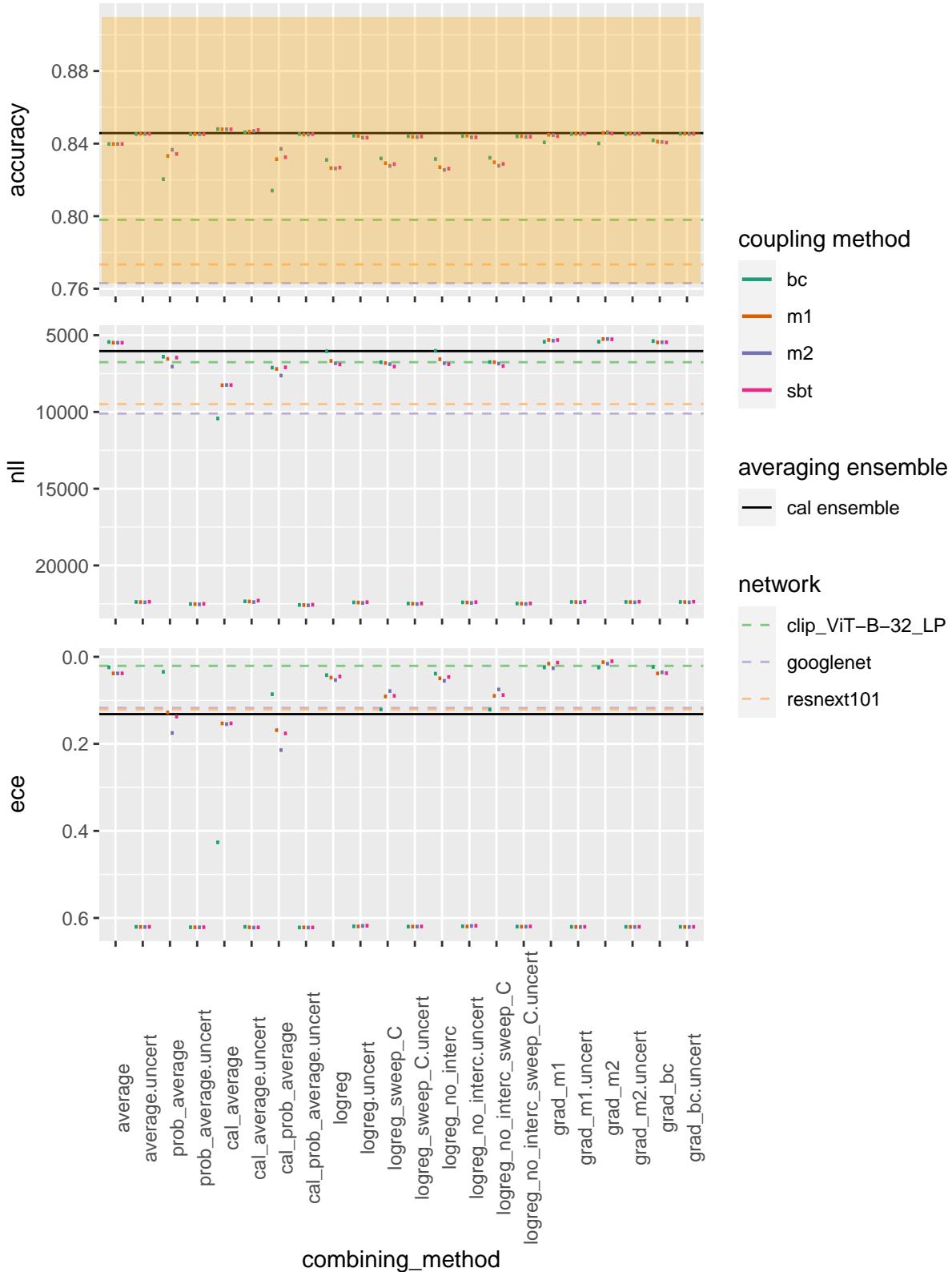
Ensemble metrics
Error inconsistency 0.287599980831146



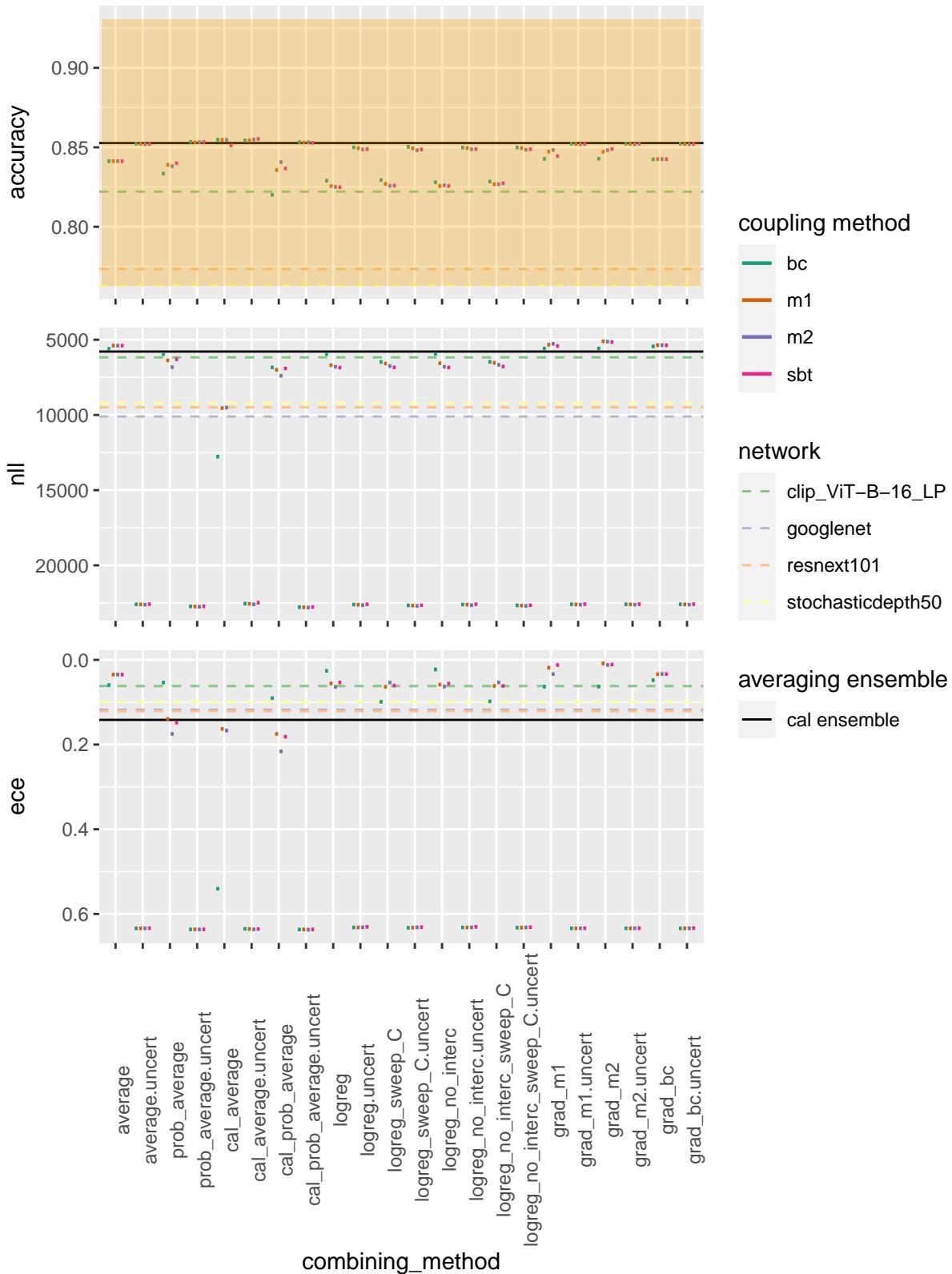
Ensemble metrics
Error inconsistency 0.209099993109703



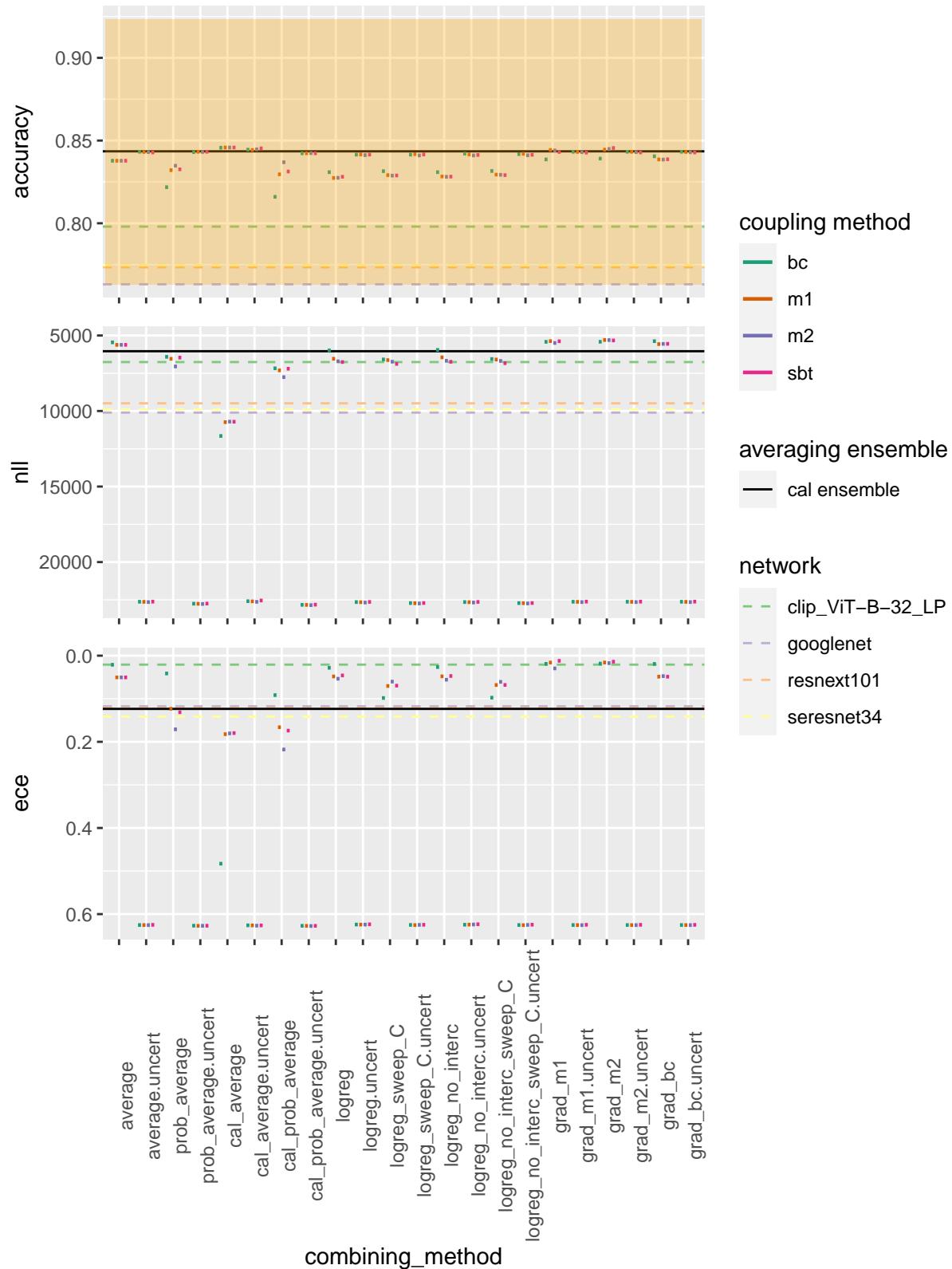
Ensemble metrics
Error inconsistency 0.277700006961823



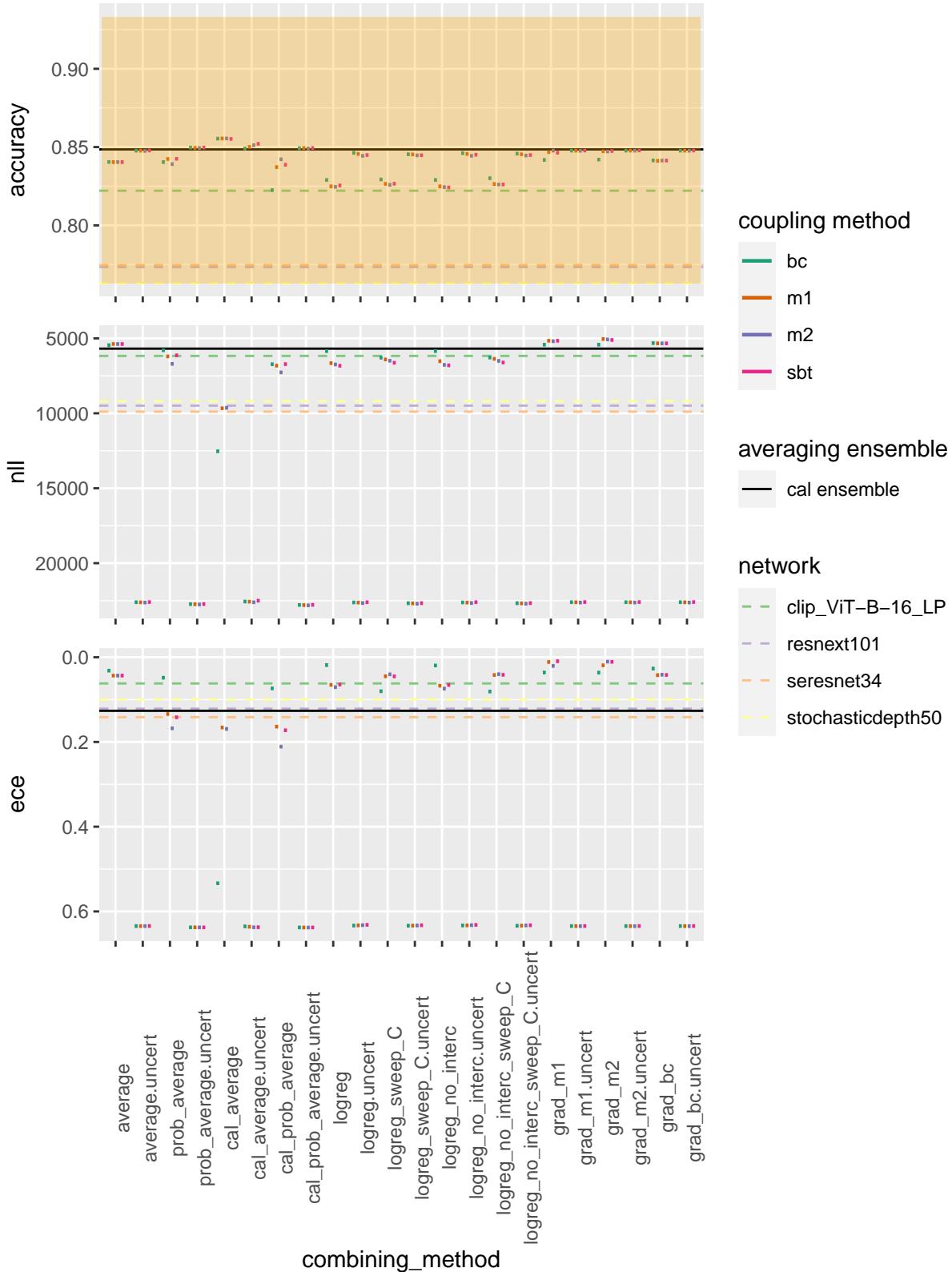
Ensemble metrics
Error inconsistency 0.326900005340576



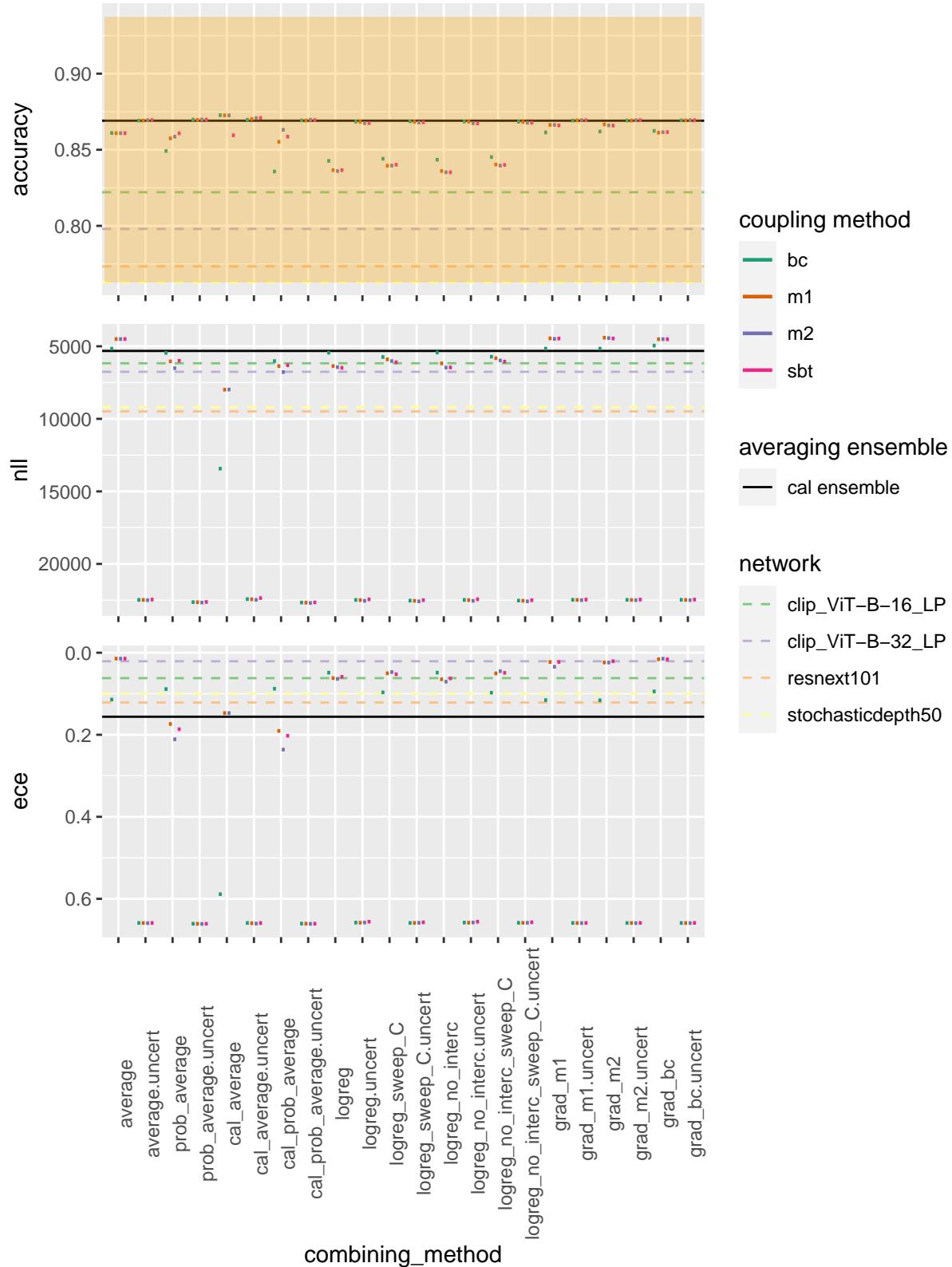
Ensemble metrics
Error inconsistency 0.319399982690811



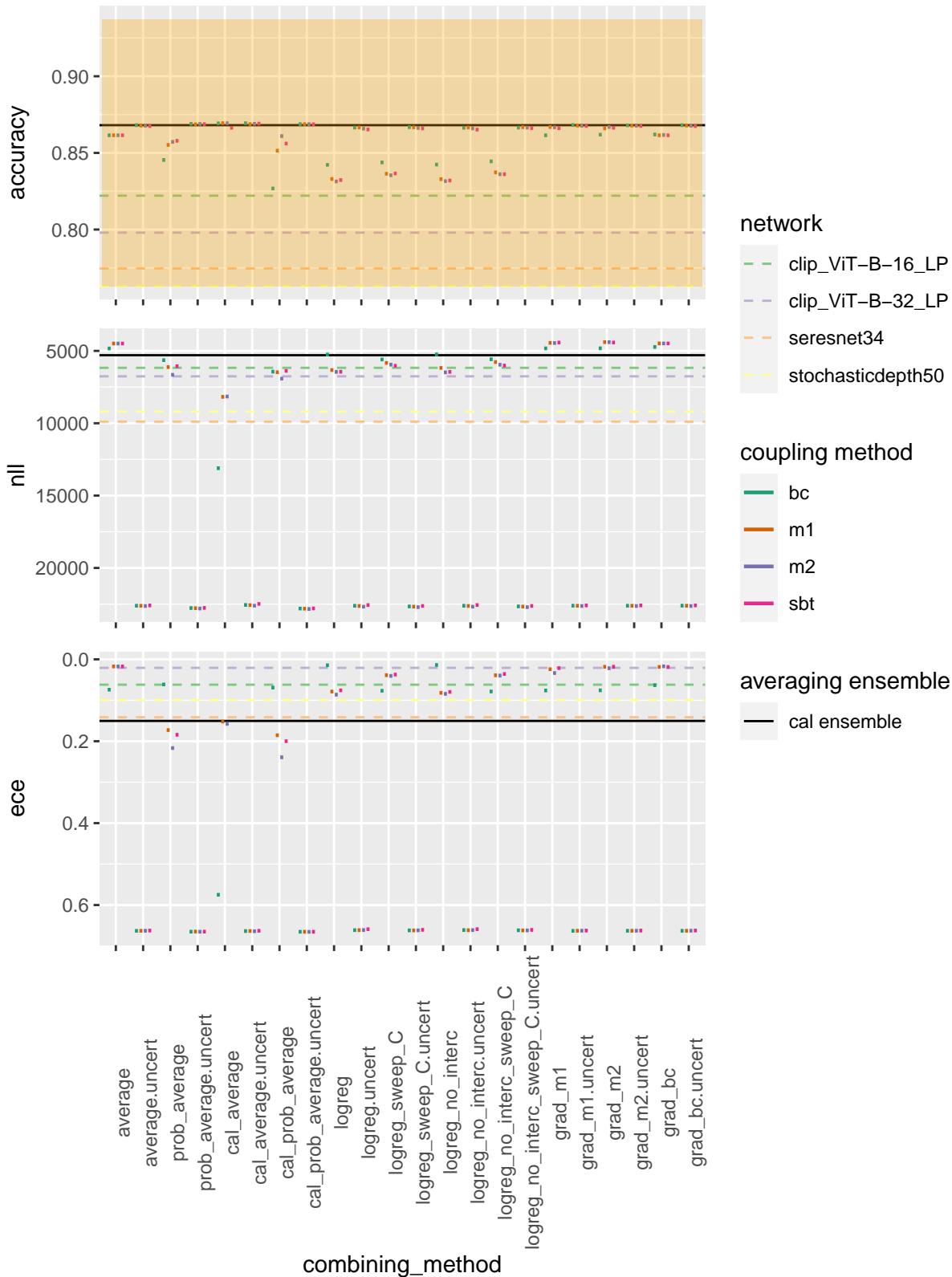
Ensemble metrics
Error inconsistency 0.321399986743927



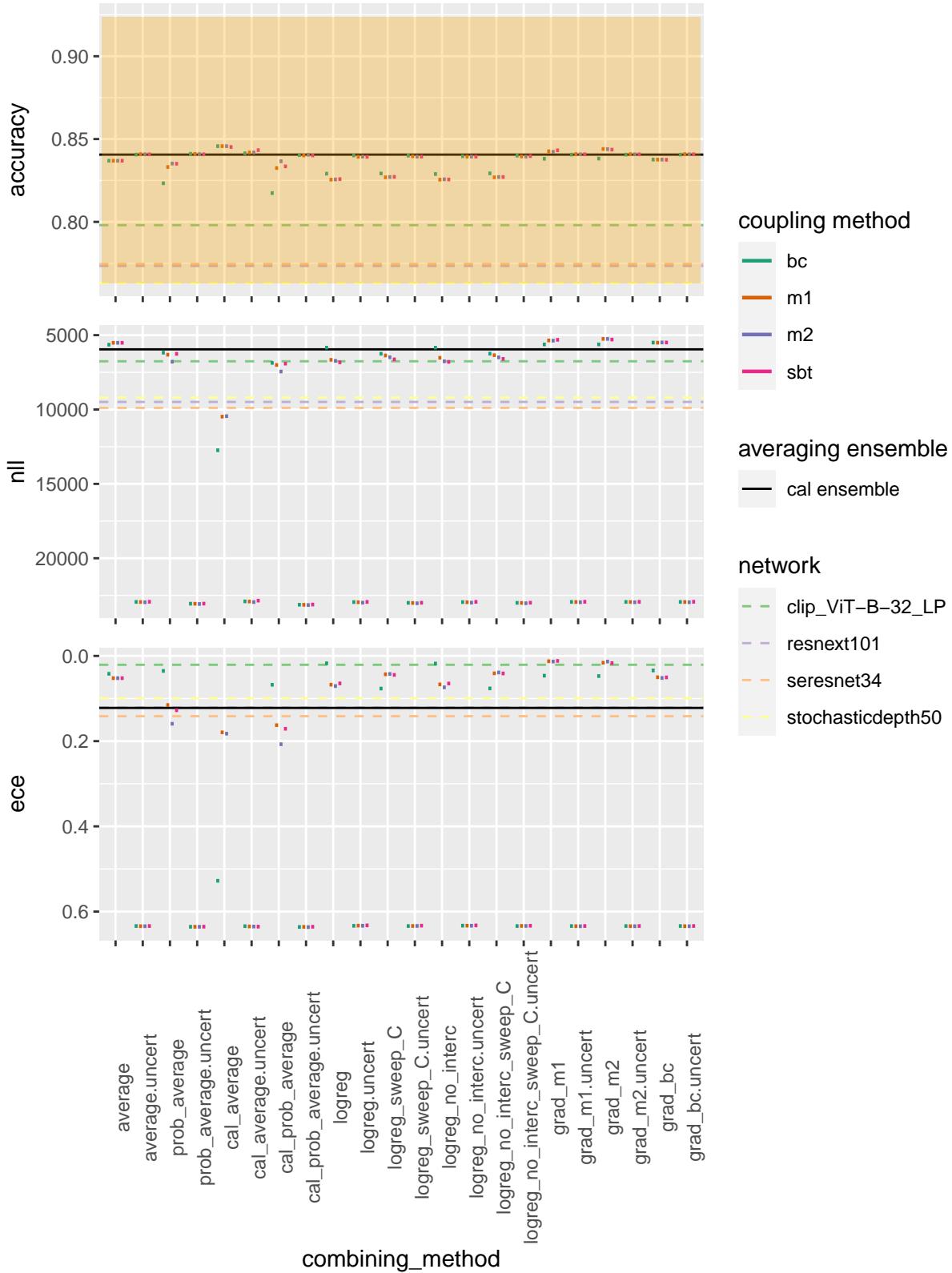
Ensemble metrics
Error inconsistency 0.337999999523163



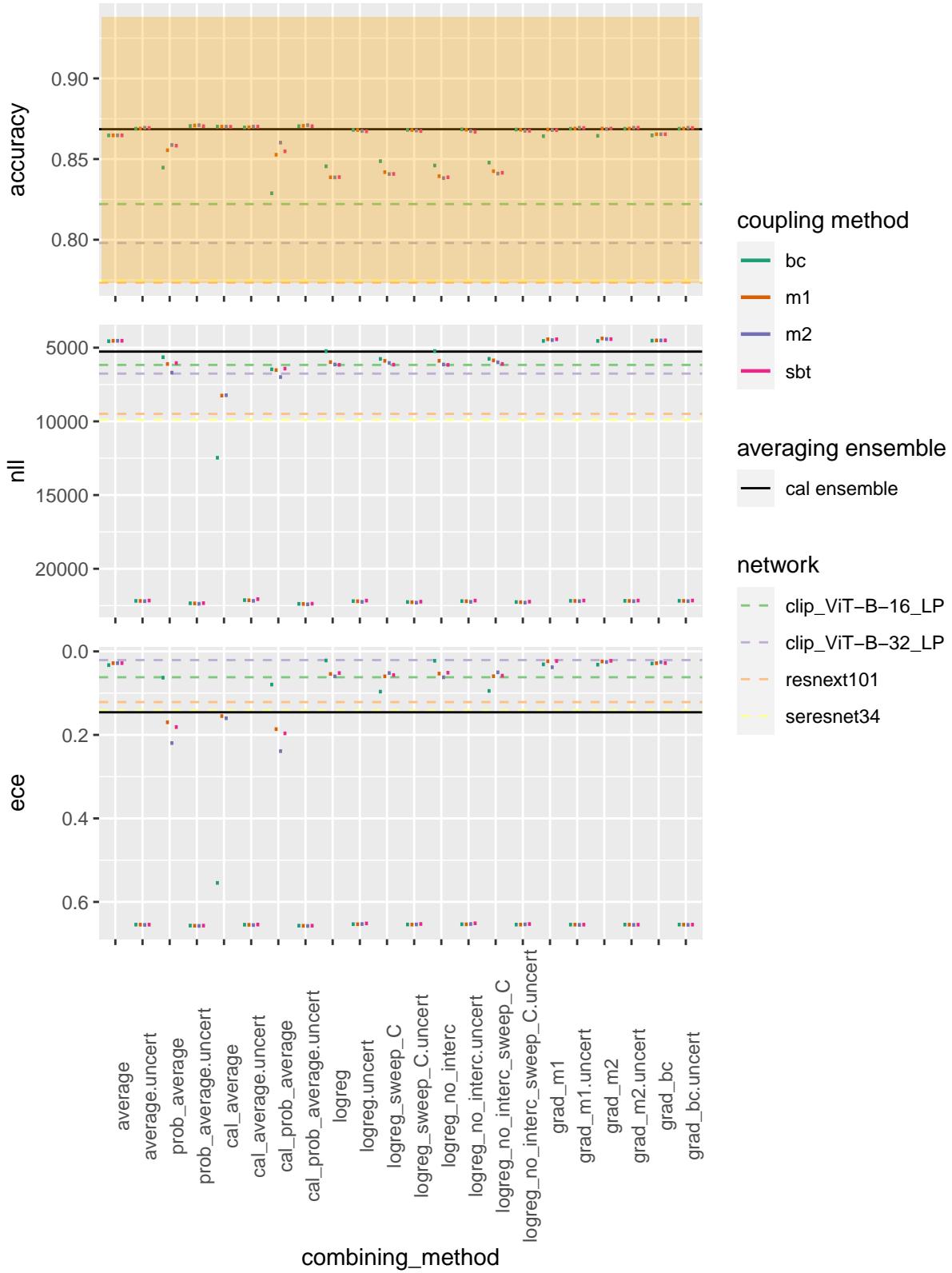
Ensemble metrics
Error inconsistency 0.338499993085861



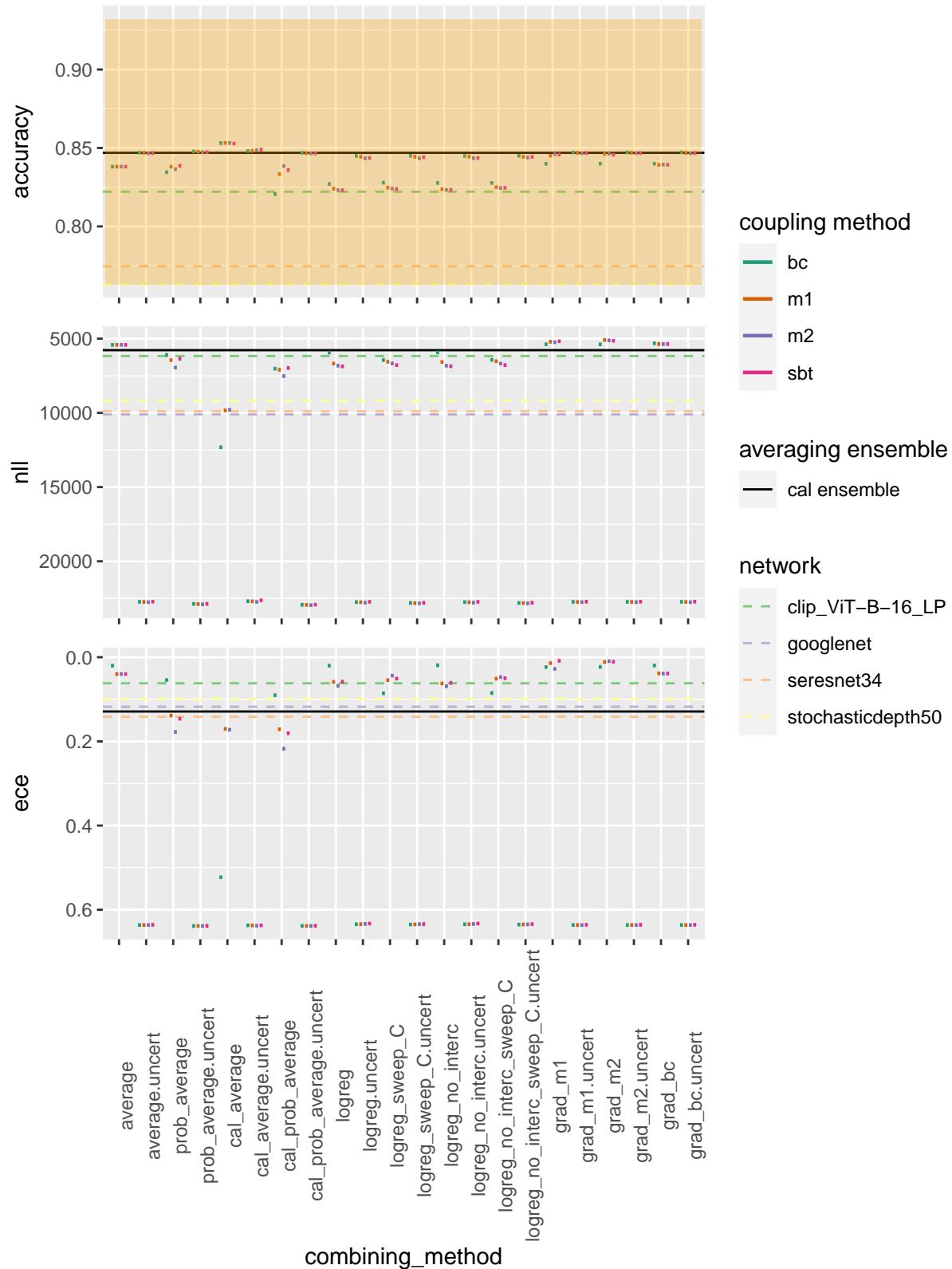
Ensemble metrics
Error inconsistency 0.319199979305267



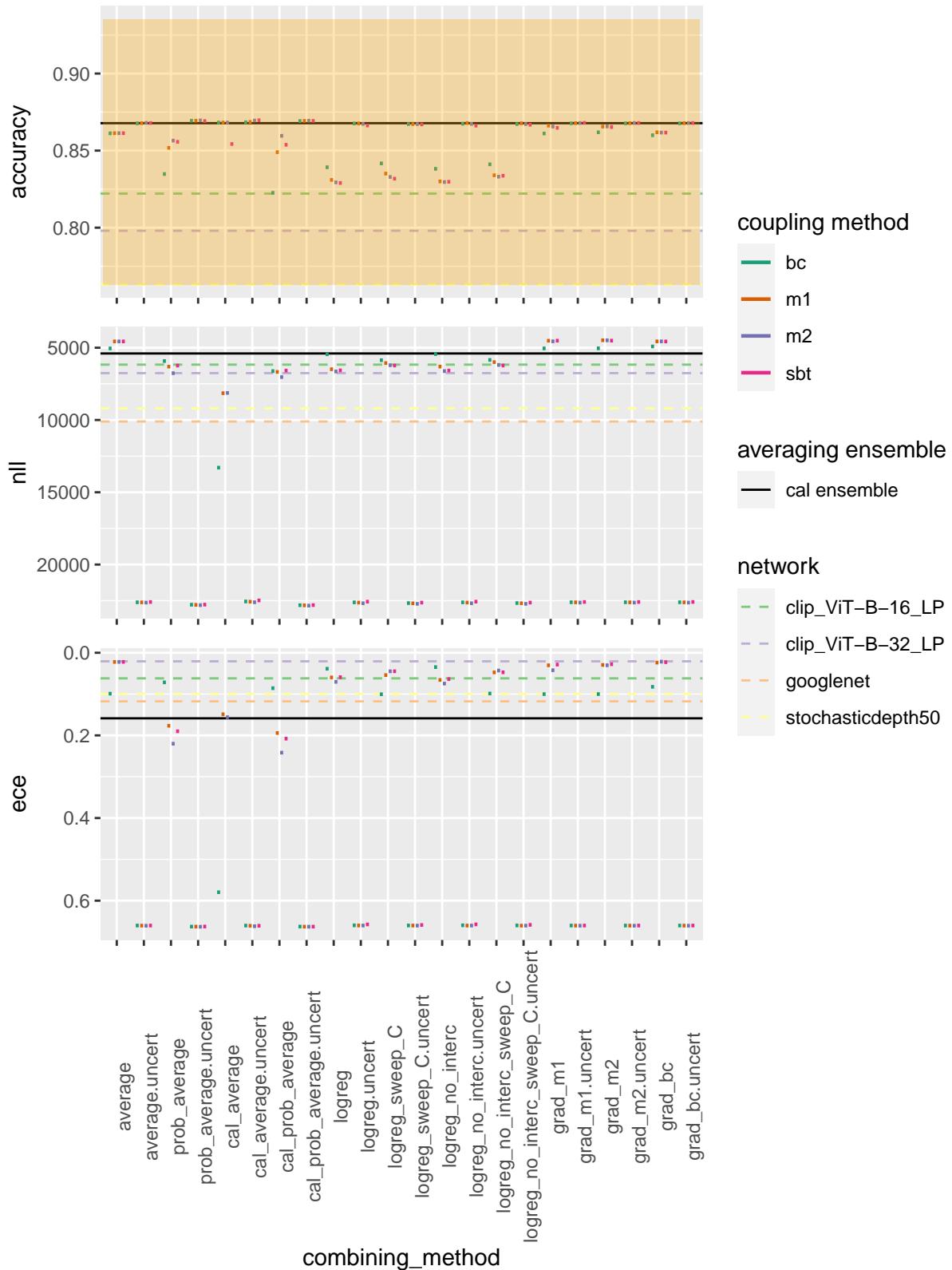
Ensemble metrics
Error inconsistency 0.333099991083145



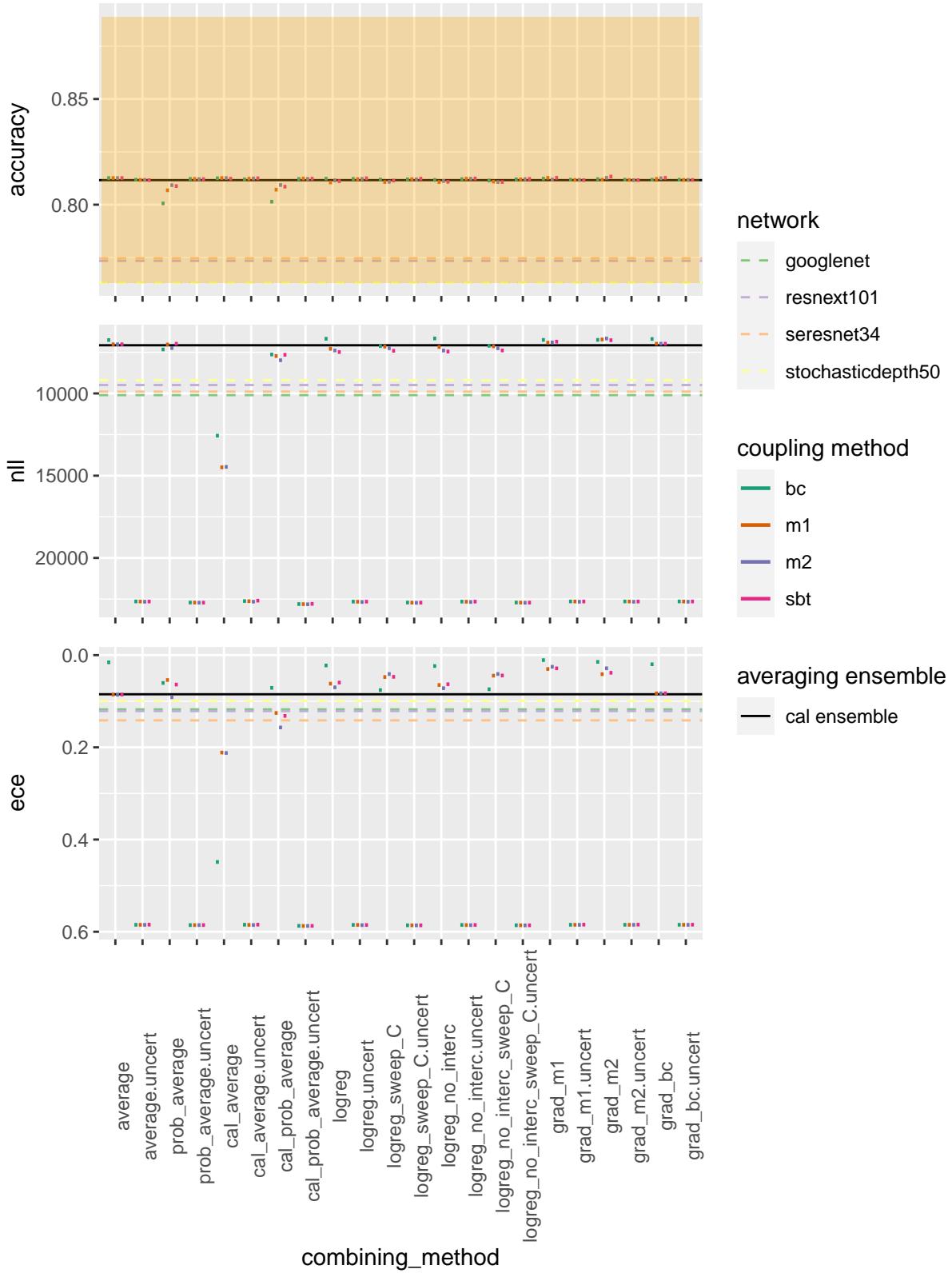
Ensemble metrics
Error inconsistency 0.326499998569489



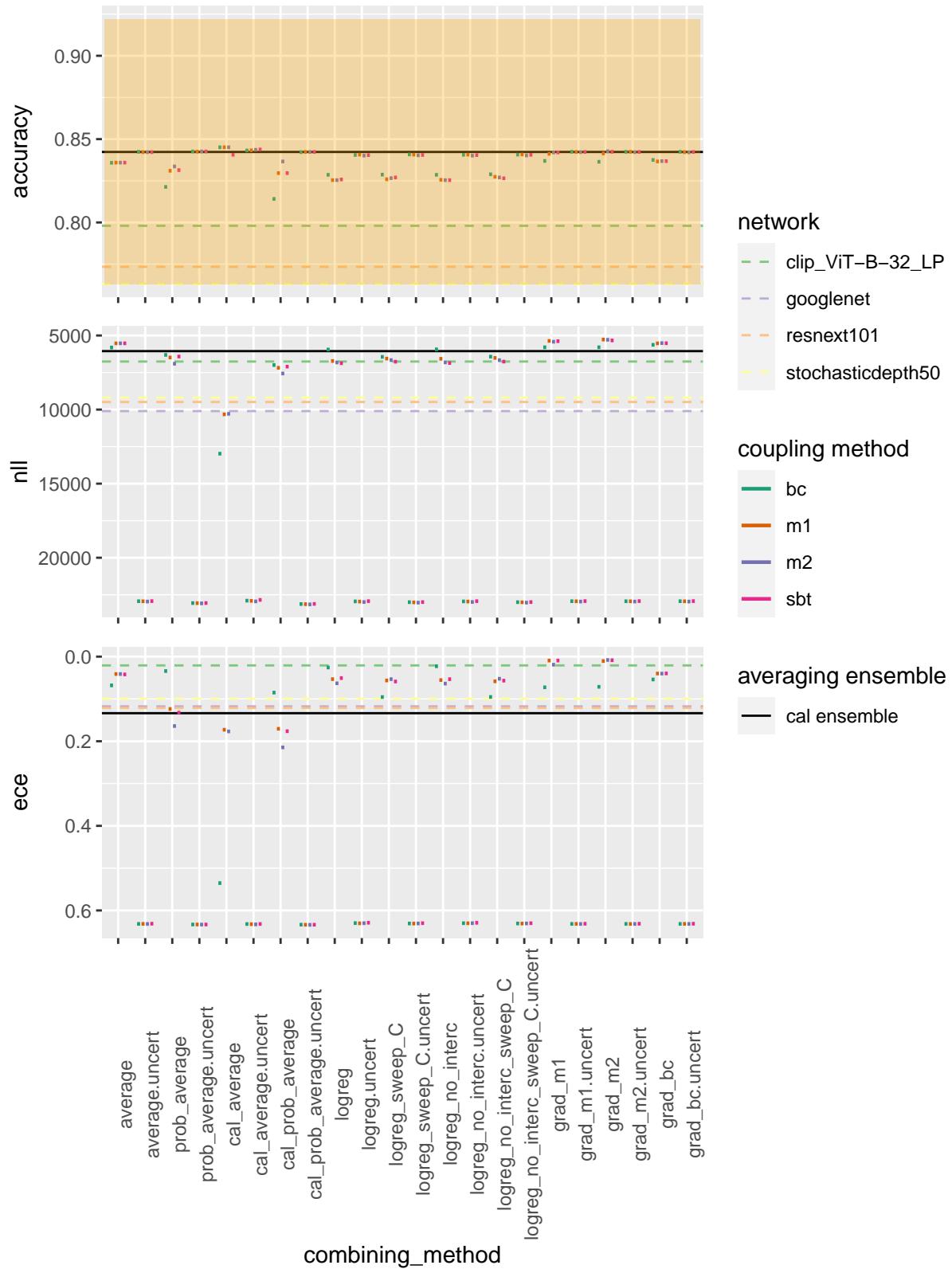
Ensemble metrics
Error inconsistency 0.341999977827072



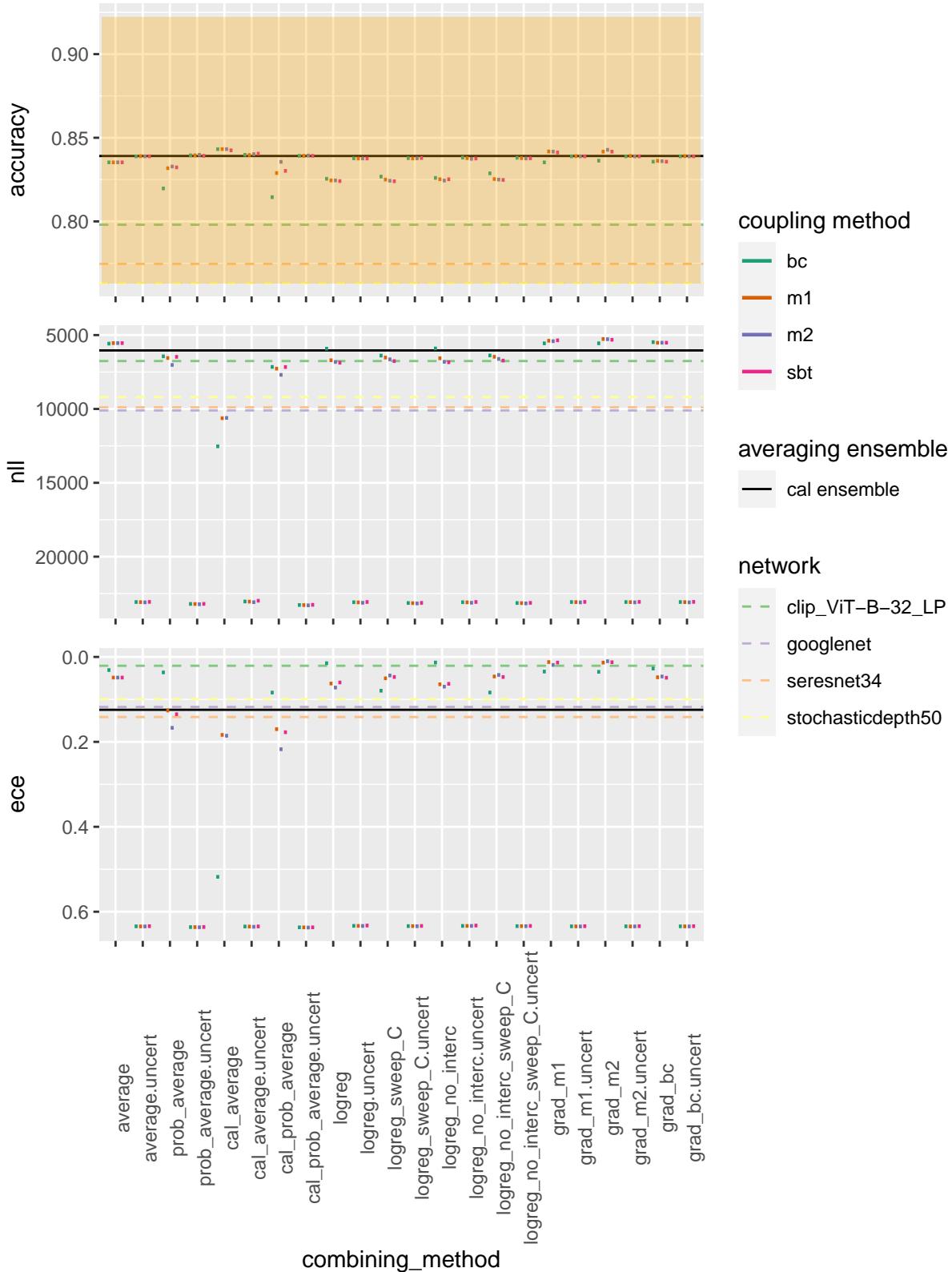
Ensemble metrics
Error inconsistency 0.260600000619888



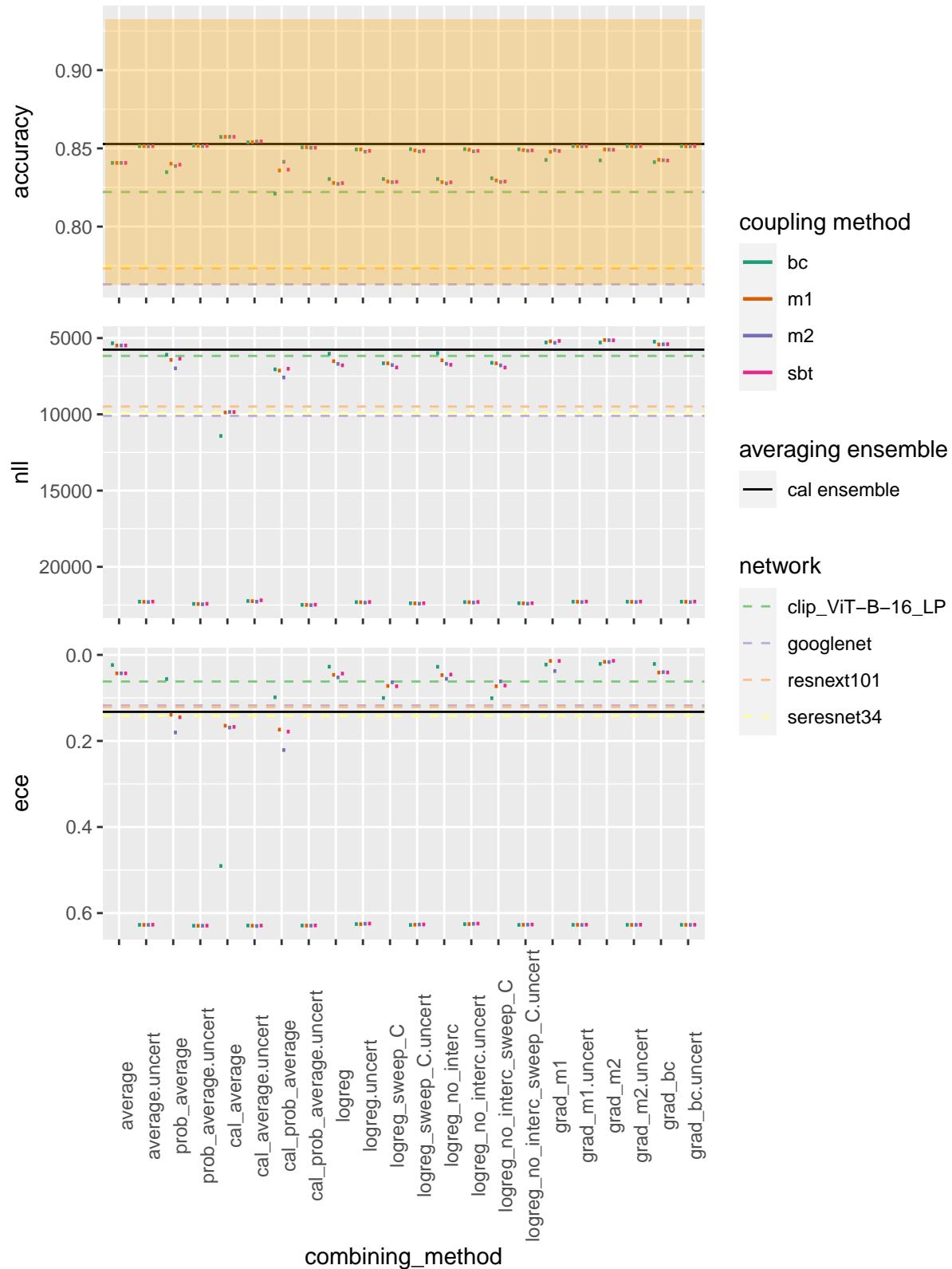
Ensemble metrics
Error inconsistency 0.324099987745285



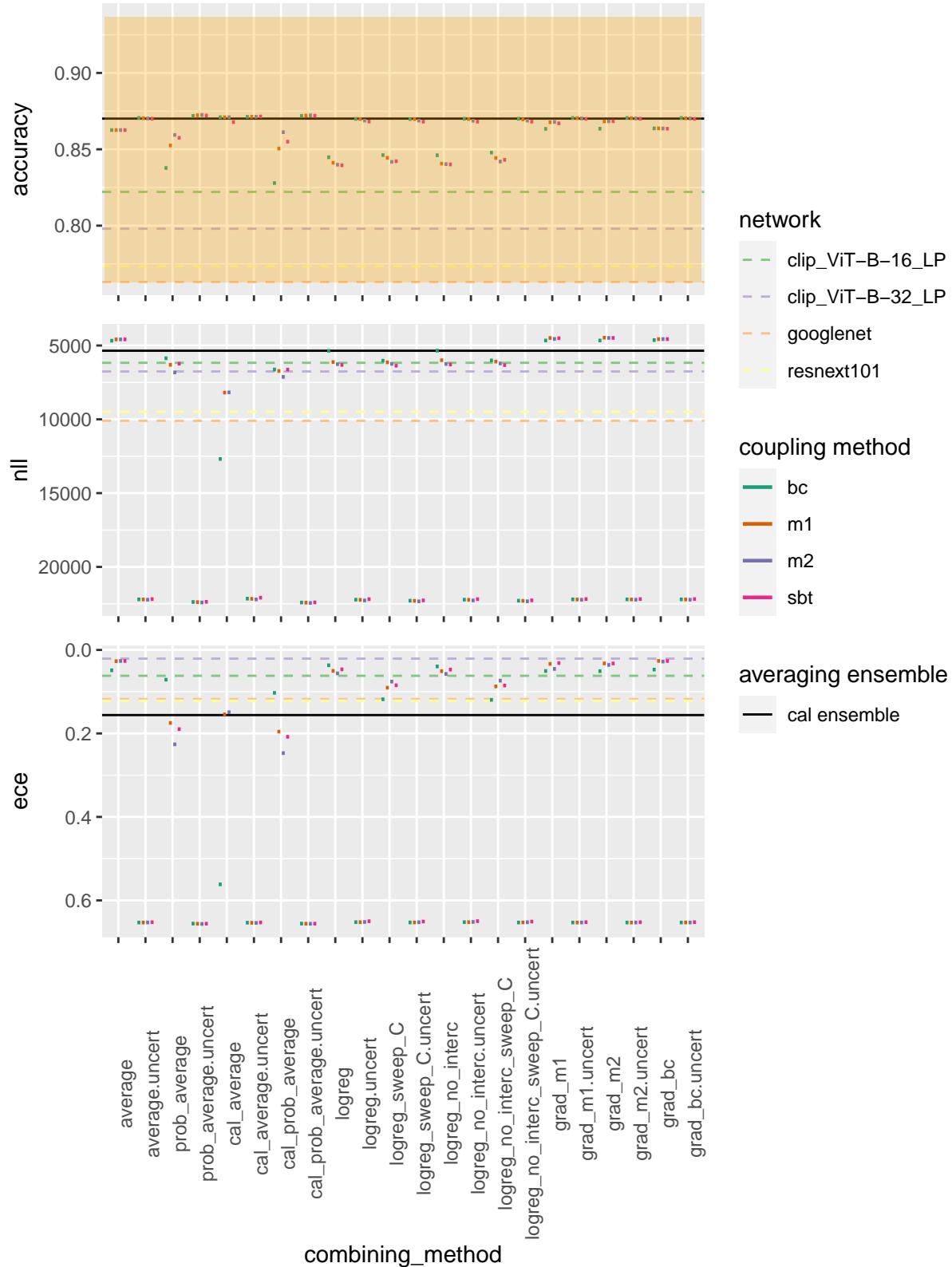
Ensemble metrics
Error inconsistency 0.323499977588654



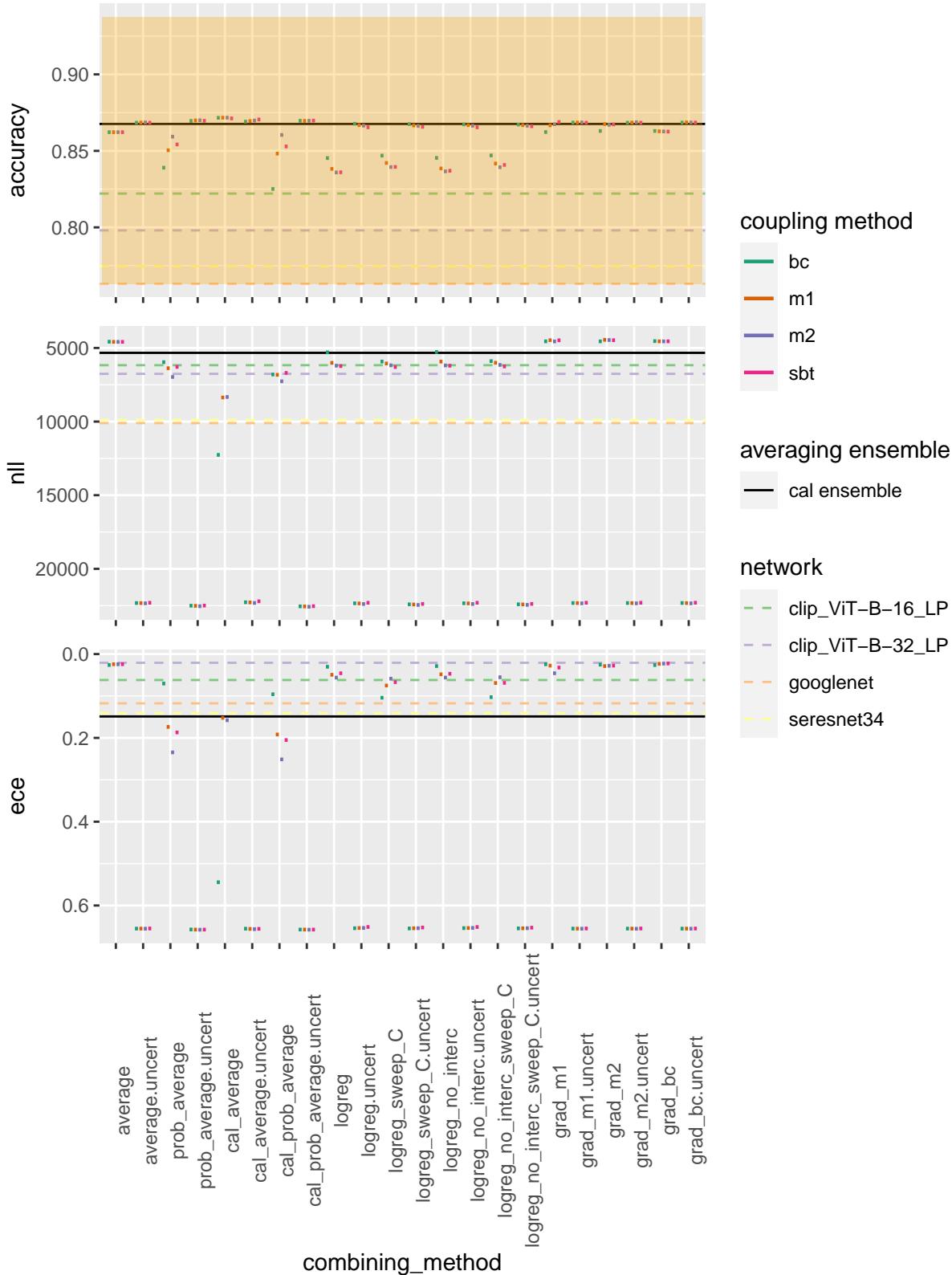
Ensemble metrics
Error inconsistency 0.322099983692169



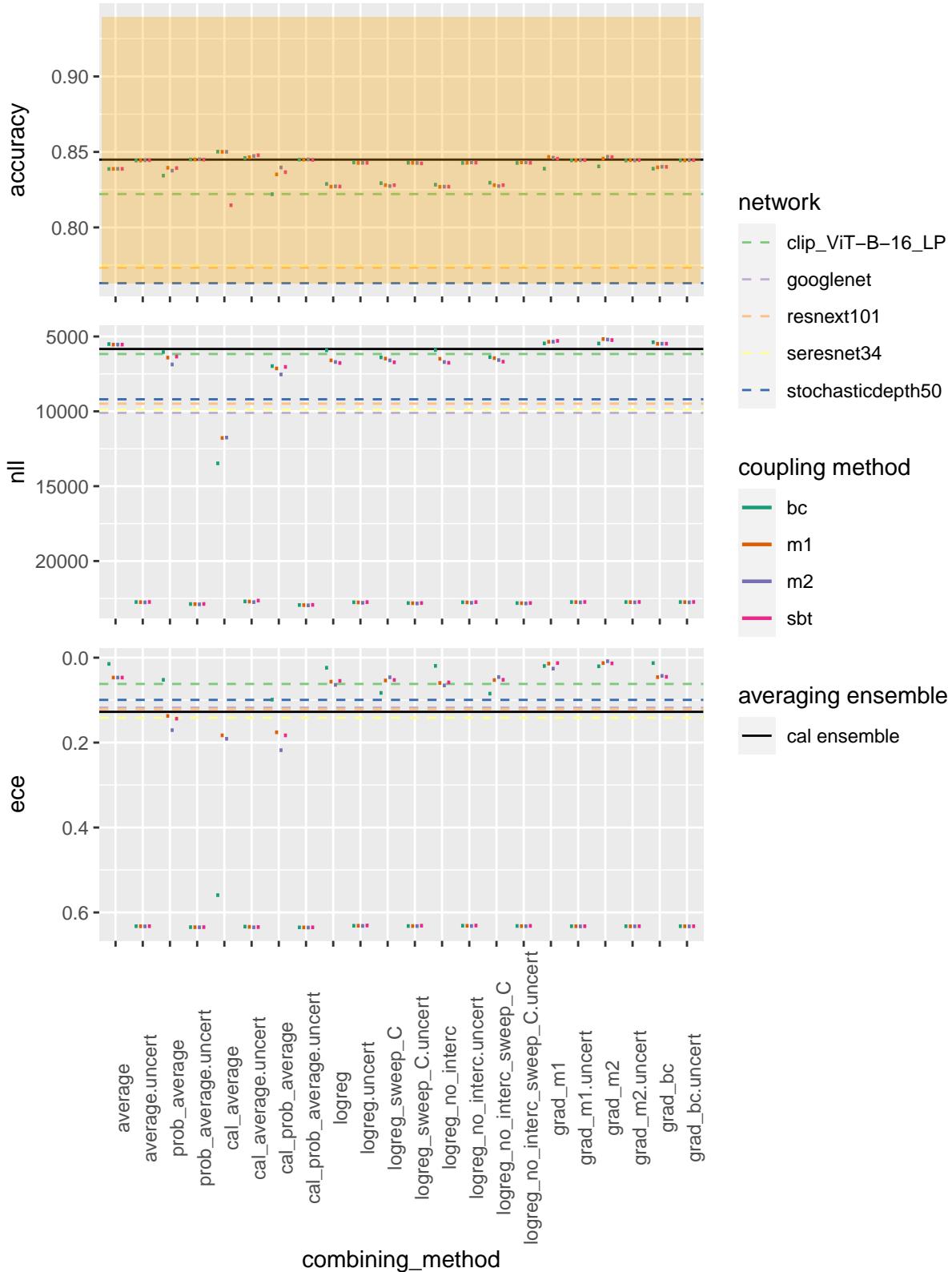
Ensemble metrics
Error inconsistency 0.337799996137619



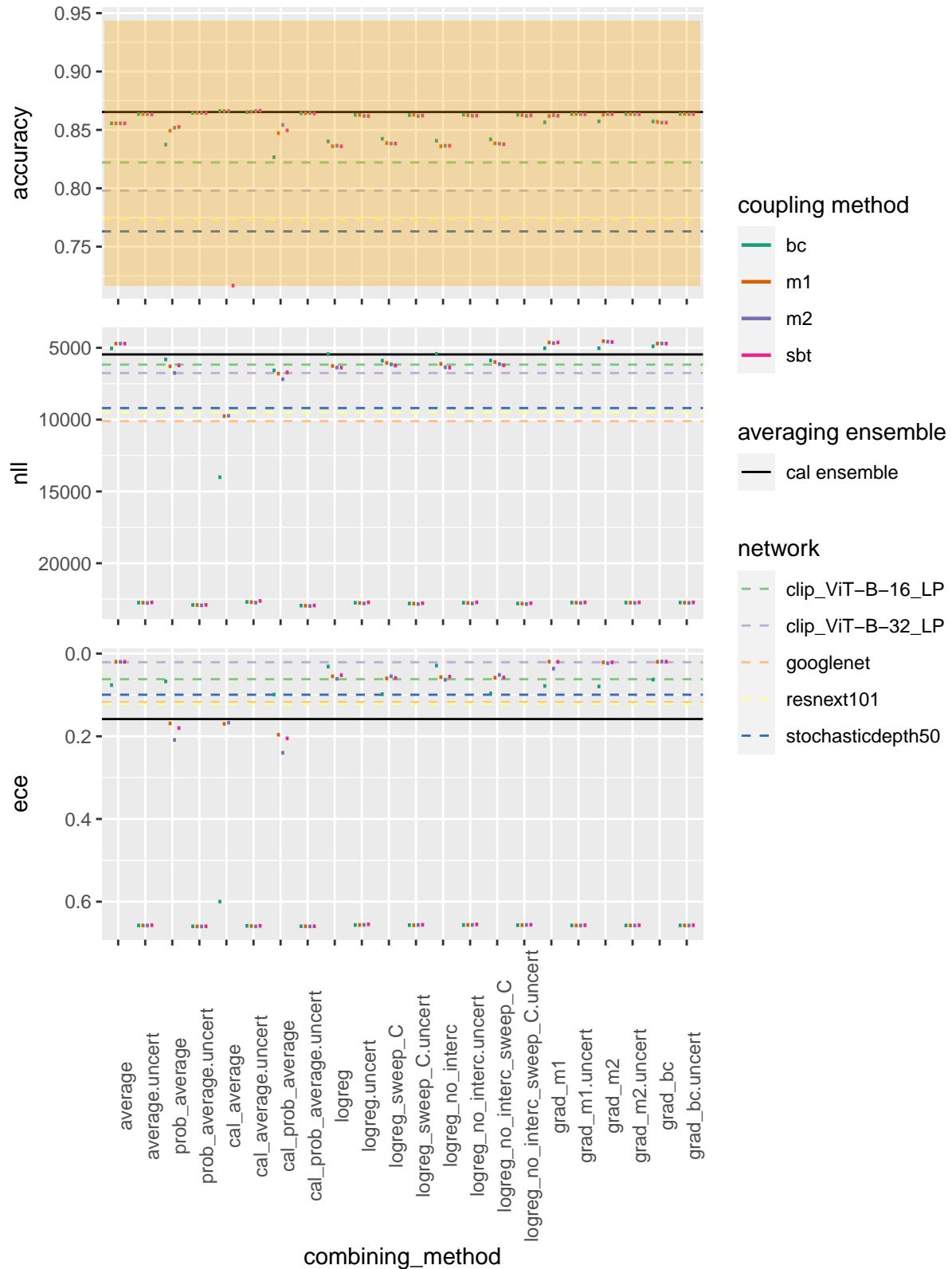
Ensemble metrics
Error inconsistency 0.339899986982346



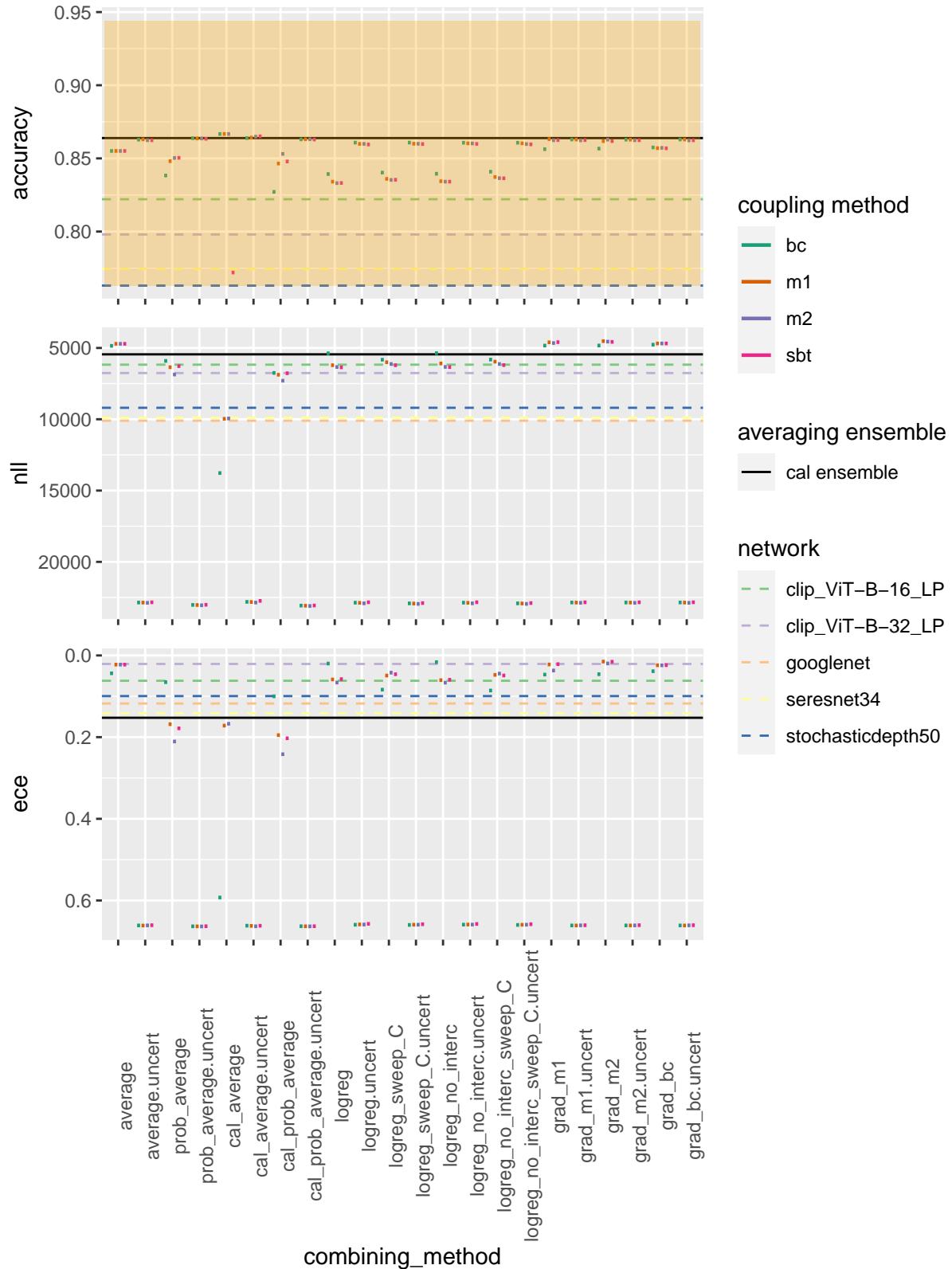
Ensemble metrics
Error inconsistency 0.353499978780746



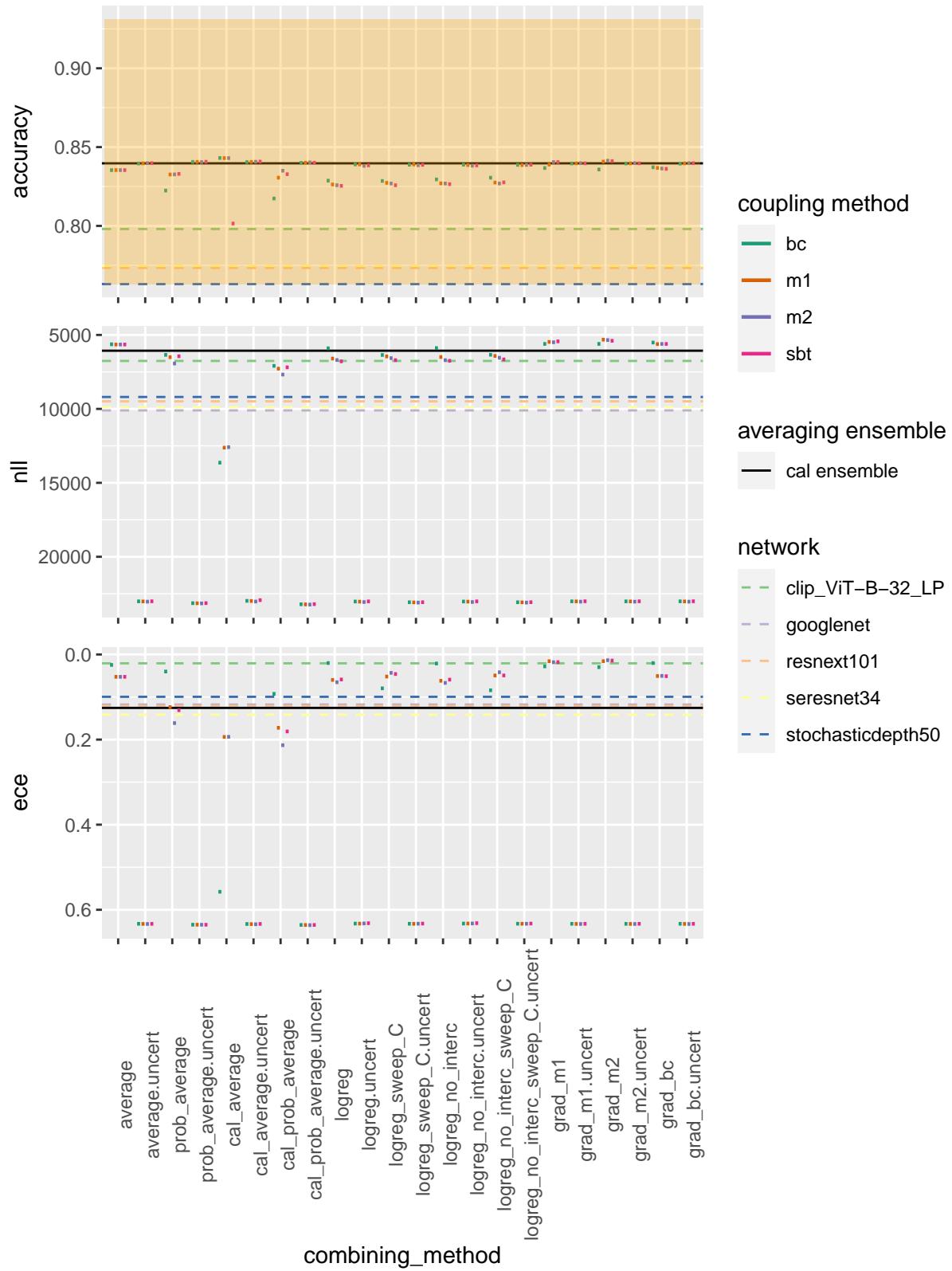
Ensemble metrics
Error inconsistency 0.373499989509583



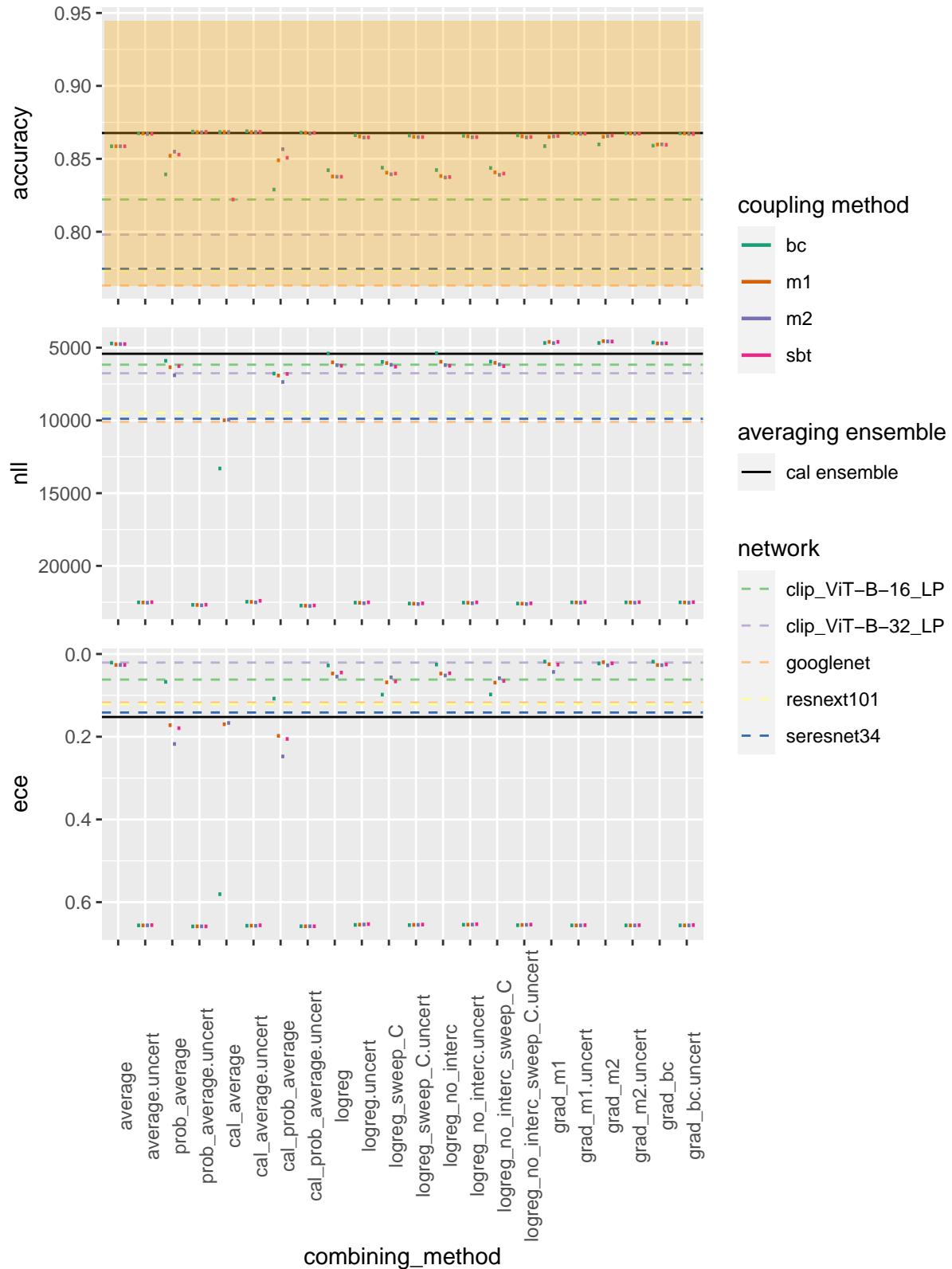
Ensemble metrics
Error inconsistency 0.373400002717972



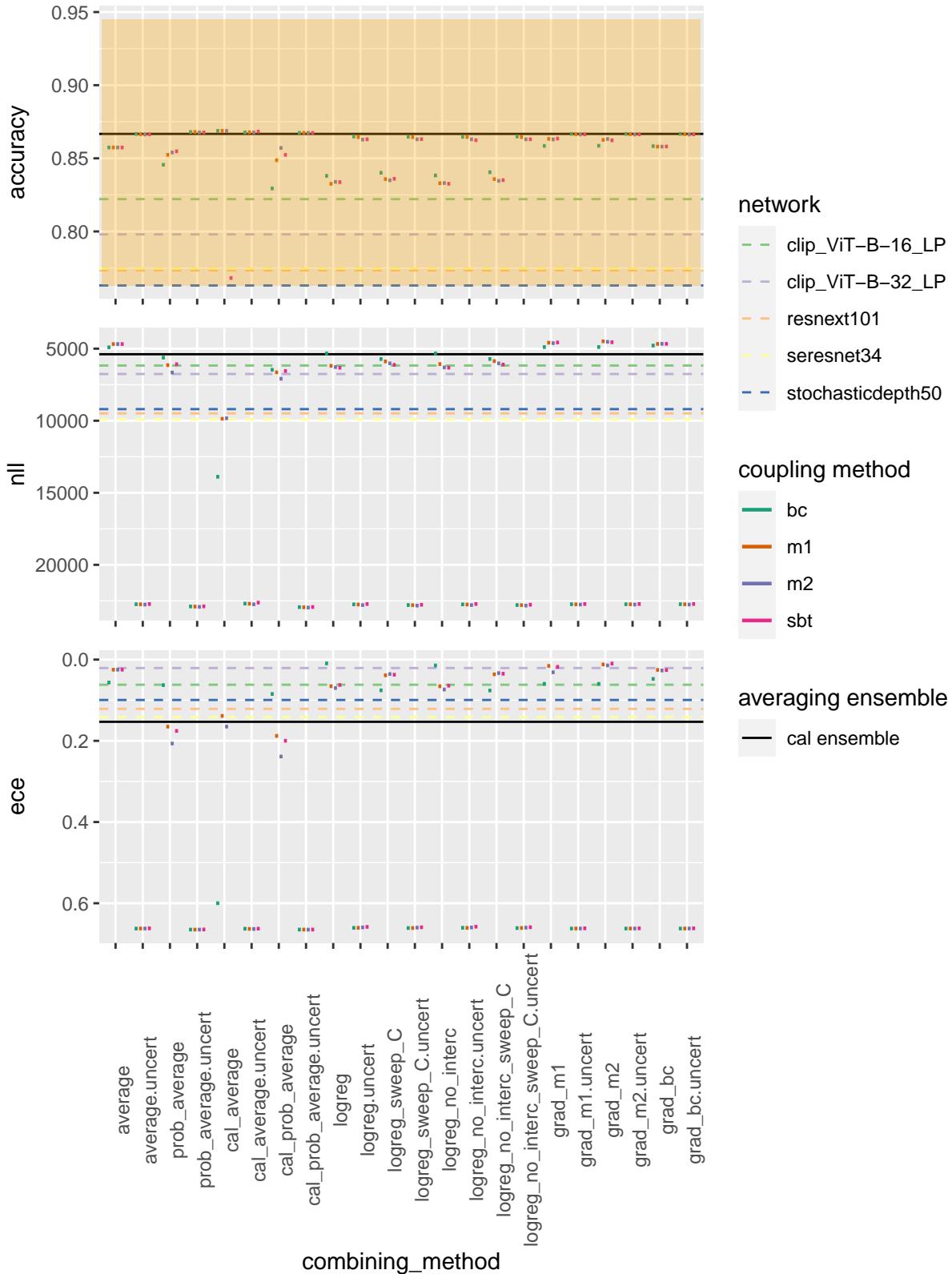
Ensemble metrics
Error inconsistency 0.351300001144409



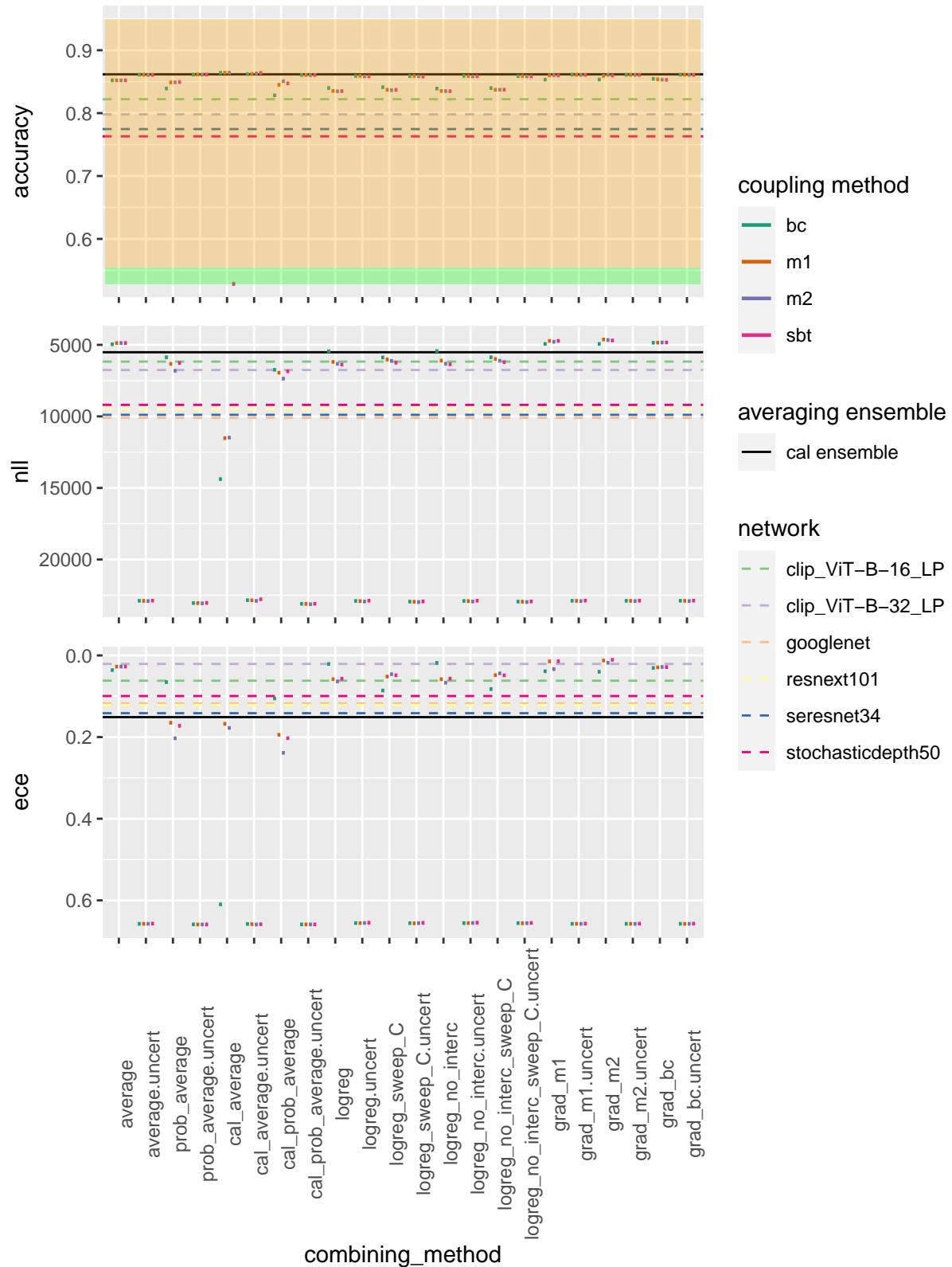
Ensemble metrics
Error inconsistency 0.369699984788895



Ensemble metrics
Error inconsistency 0.368099987506866



Ensemble metrics
Error inconsistency 0.394899994134903



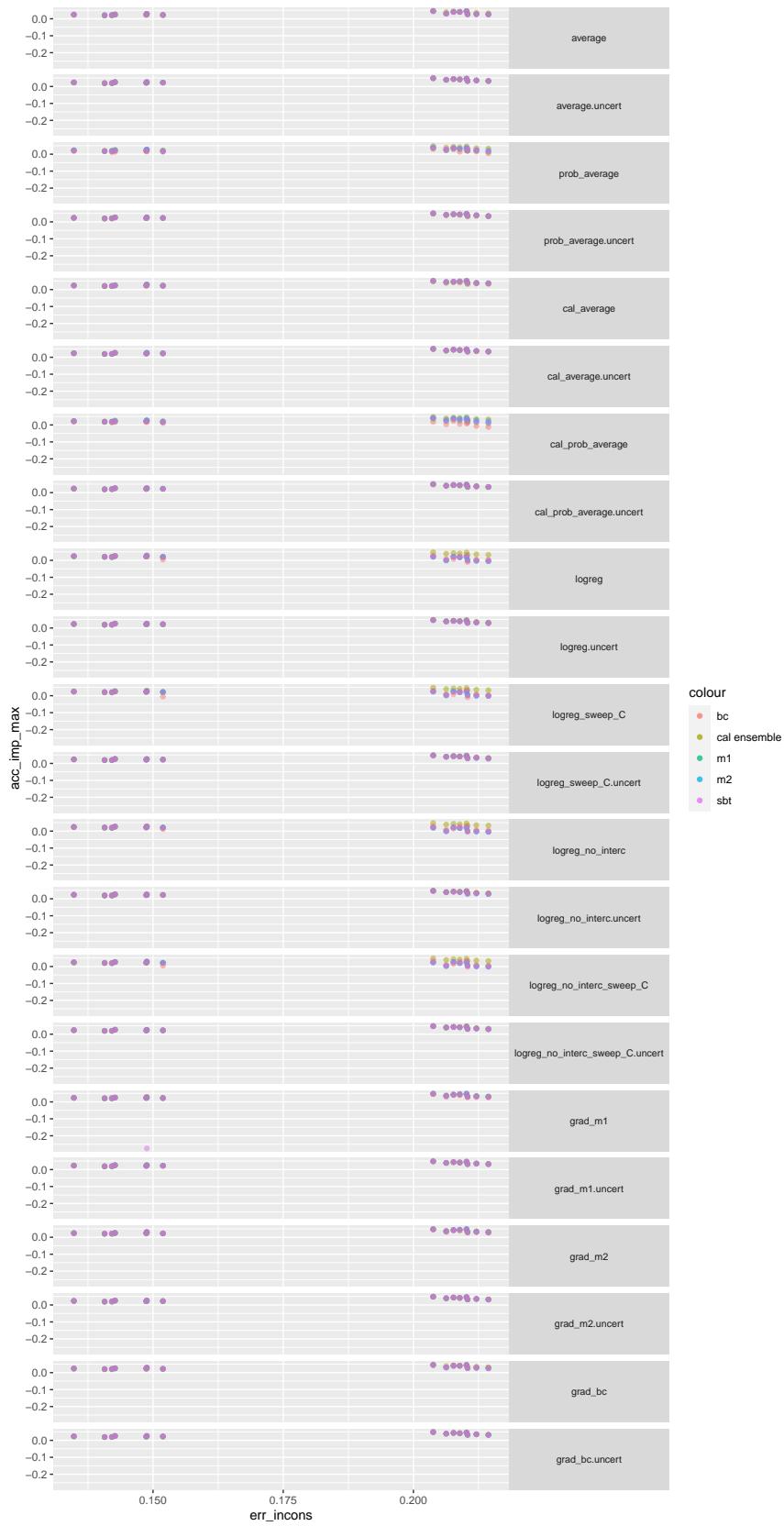
```

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
        )) +
        theme(strip.text.y = element_text(size = 8, angle = 0))

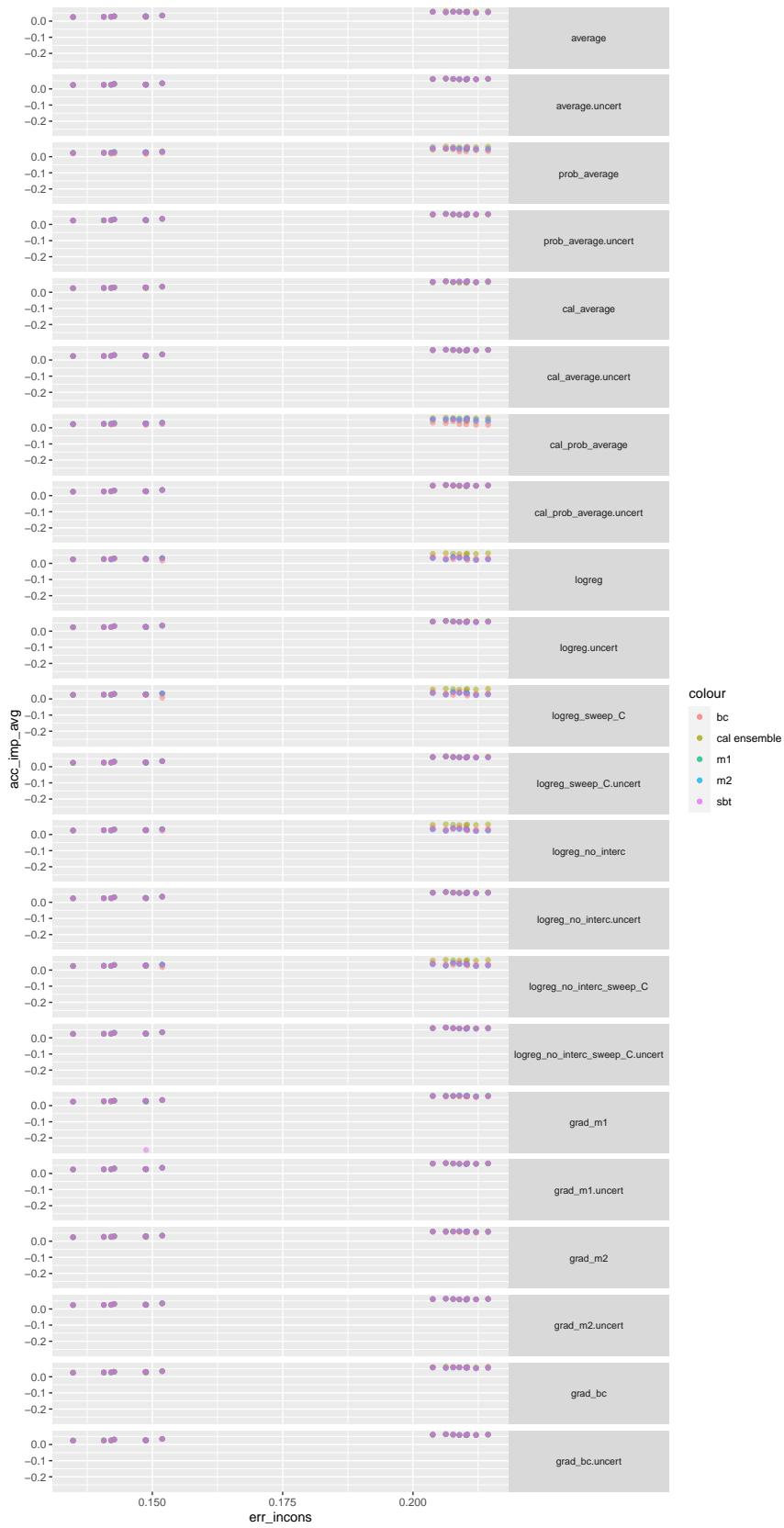
      print(cur_plot)
    }
  }
}

```

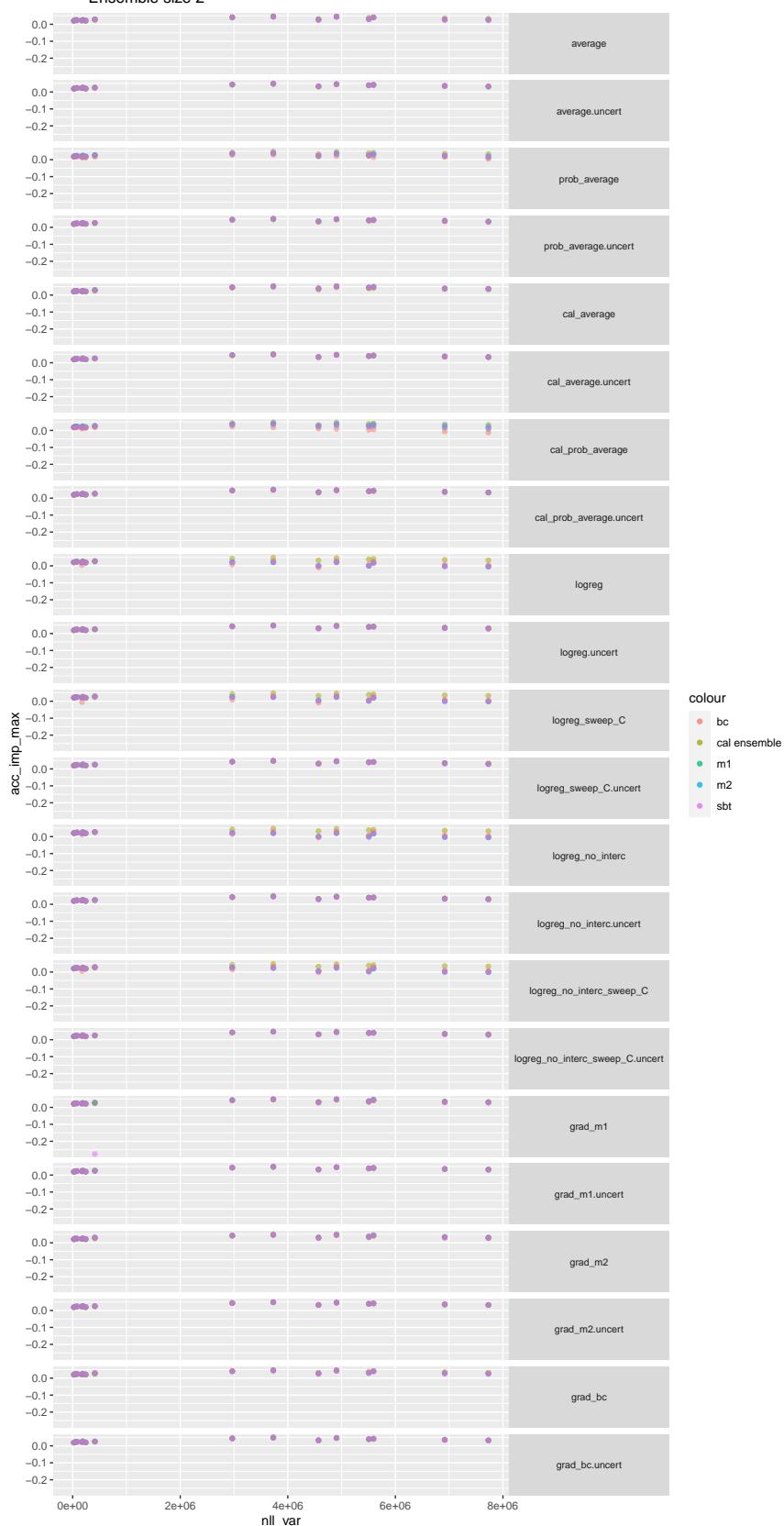
Accuracy improvement of ensemble over
the best of networks vs err_incons.
Ensemble size 2



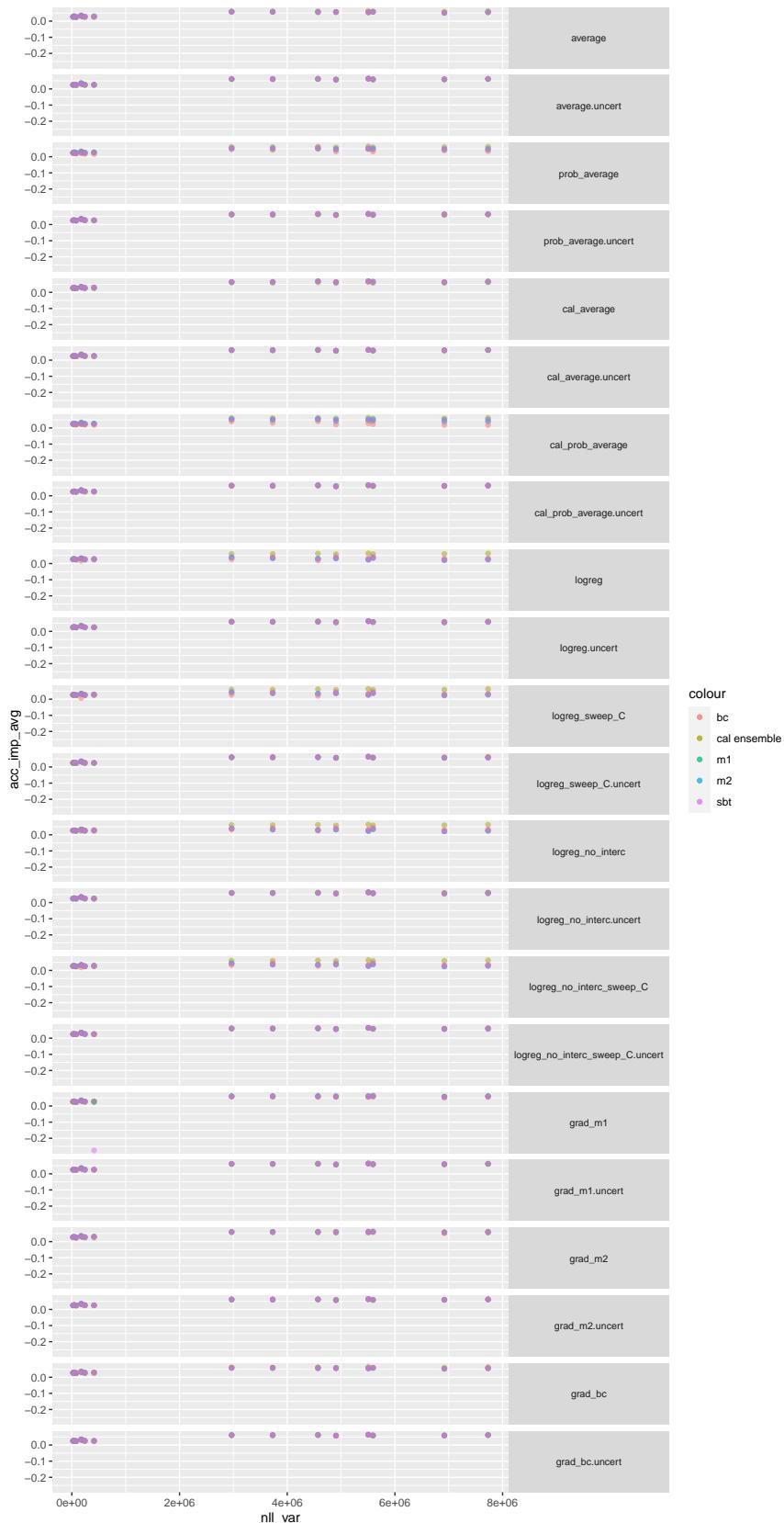
Accuracy improvement of ensemble over
the average of networks vs err_incons.
Ensemble size 2



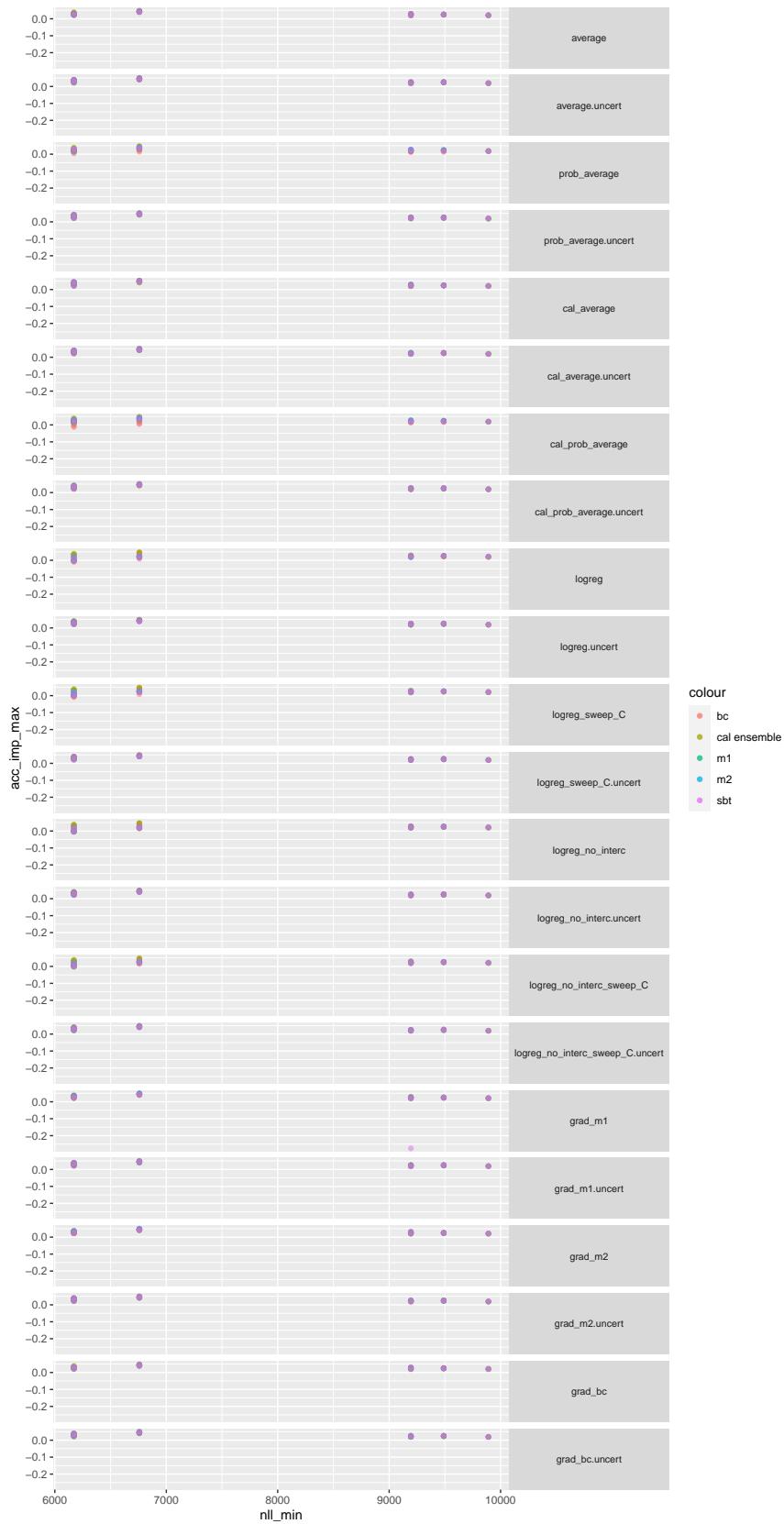
Accuracy improvement of ensemble over
the best of networks vs nll_var.
Ensemble size 2



Accuracy improvement of ensemble over
the average of networks vs nll_var.
Ensemble size 2



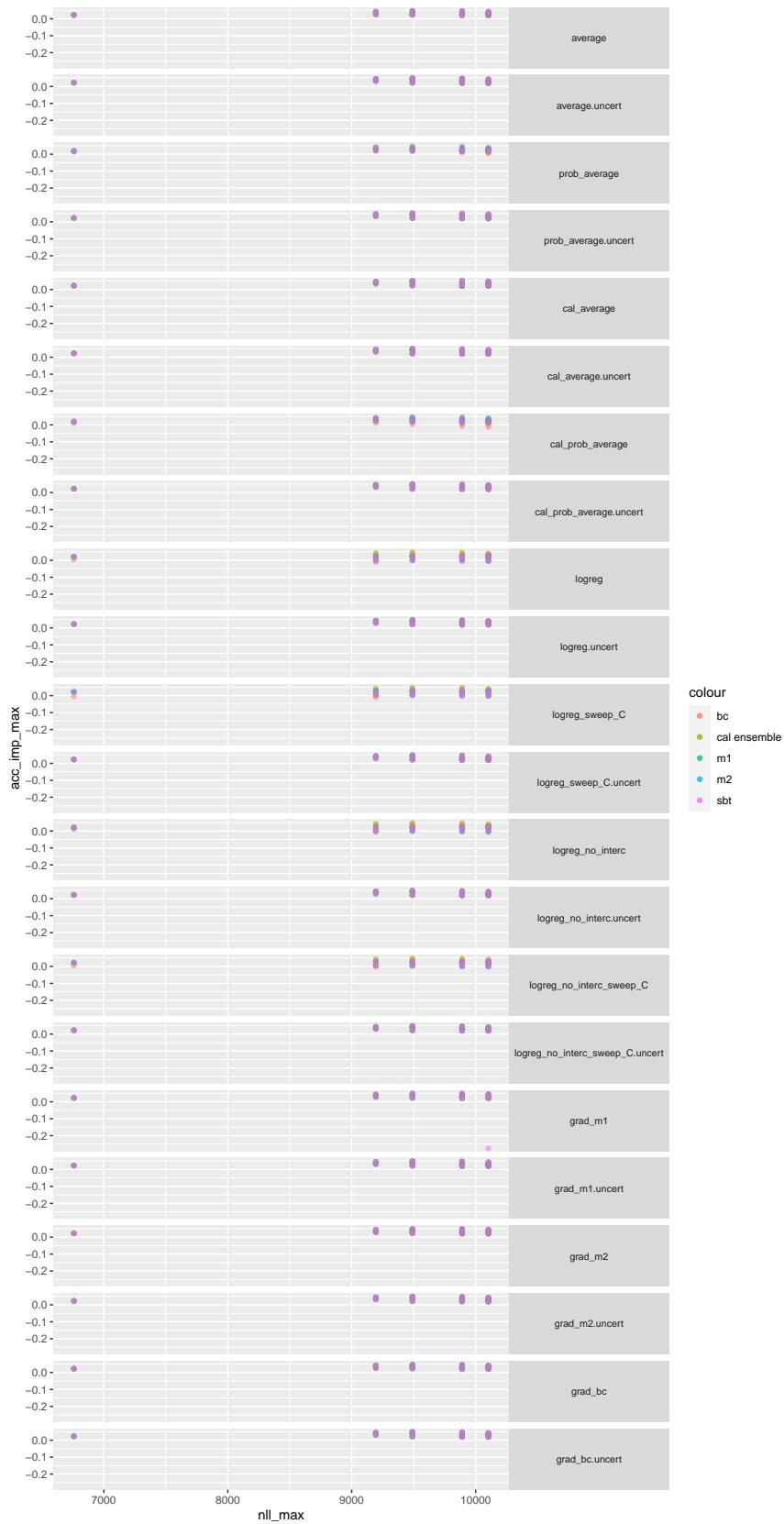
Accuracy improvement of ensemble over
the best of networks vs nll_min.
Ensemble size 2



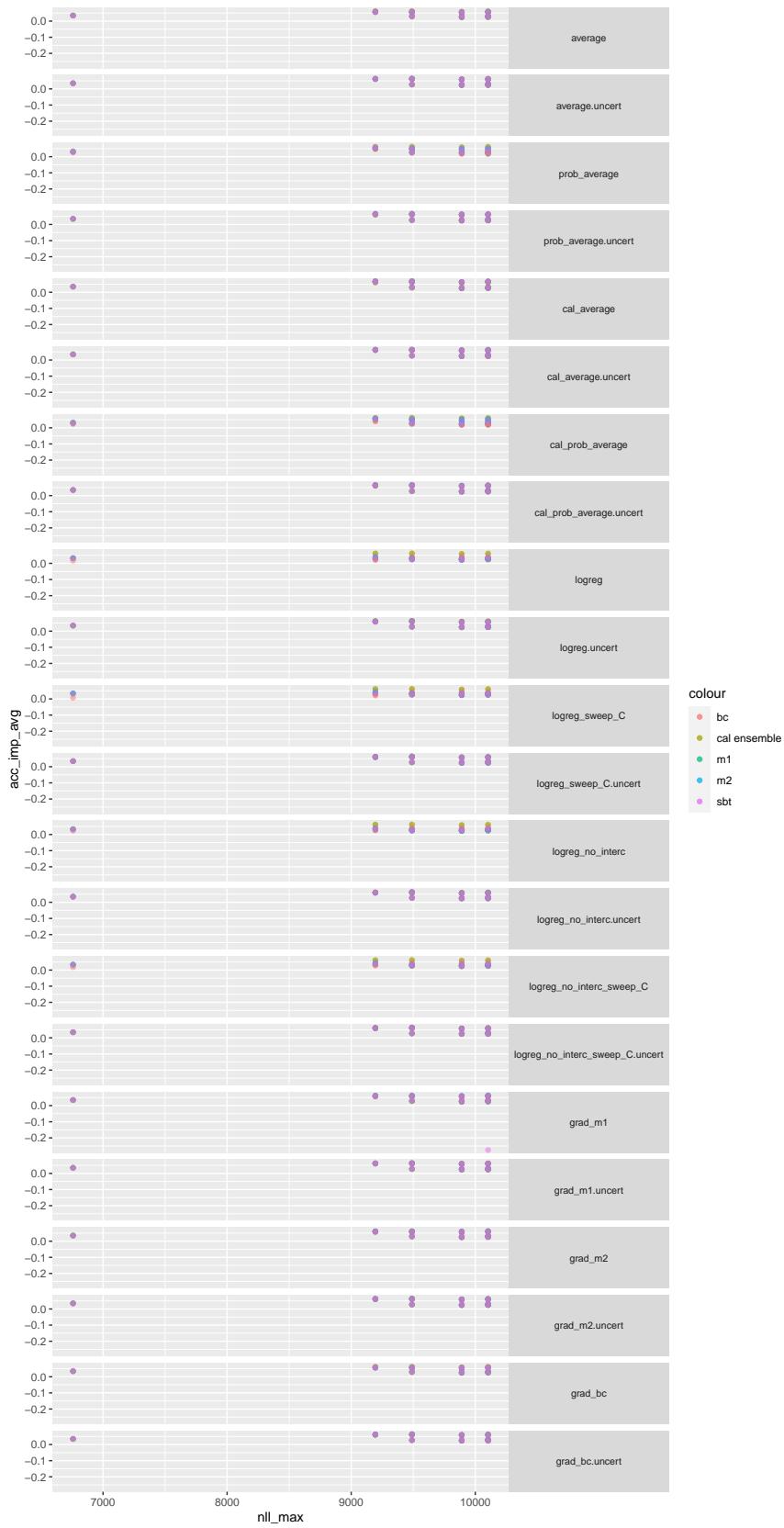
Accuracy improvement of ensemble over
the average of networks vs nll_min.
Ensemble size 2



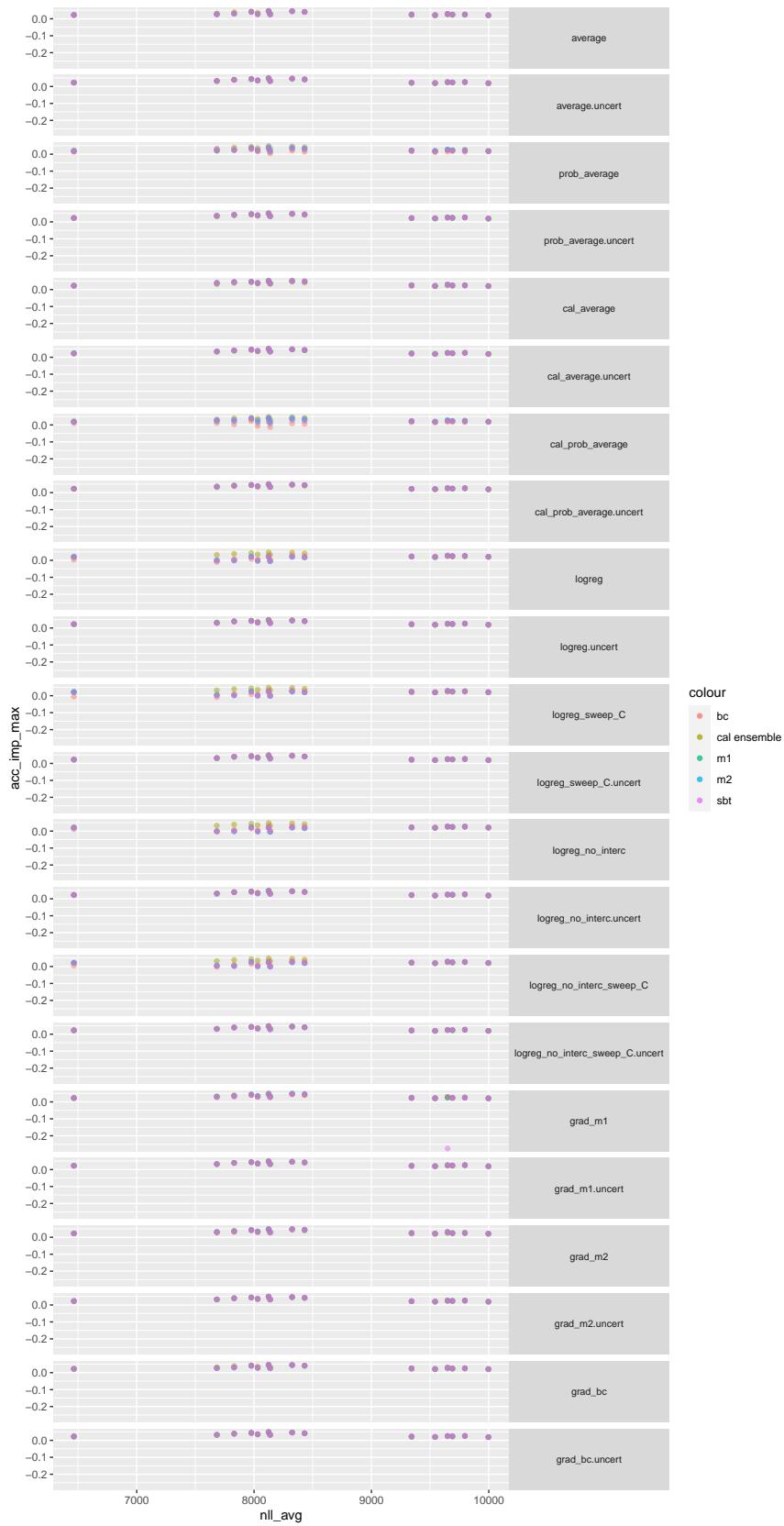
Accuracy improvement of ensemble over
the best of networks vs nll_max.
Ensemble size 2



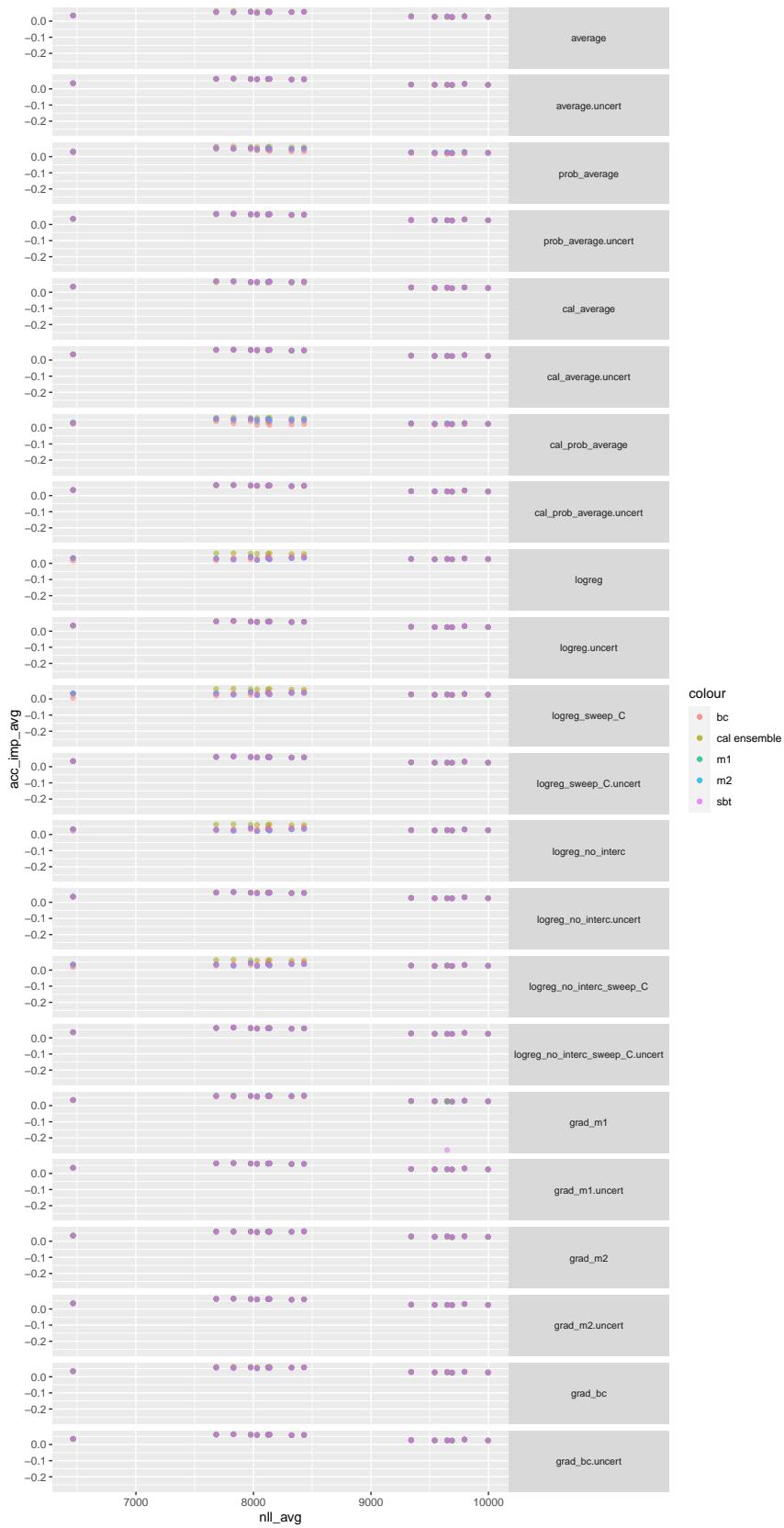
Accuracy improvement of ensemble over
the average of networks vs nll_max.
Ensemble size 2



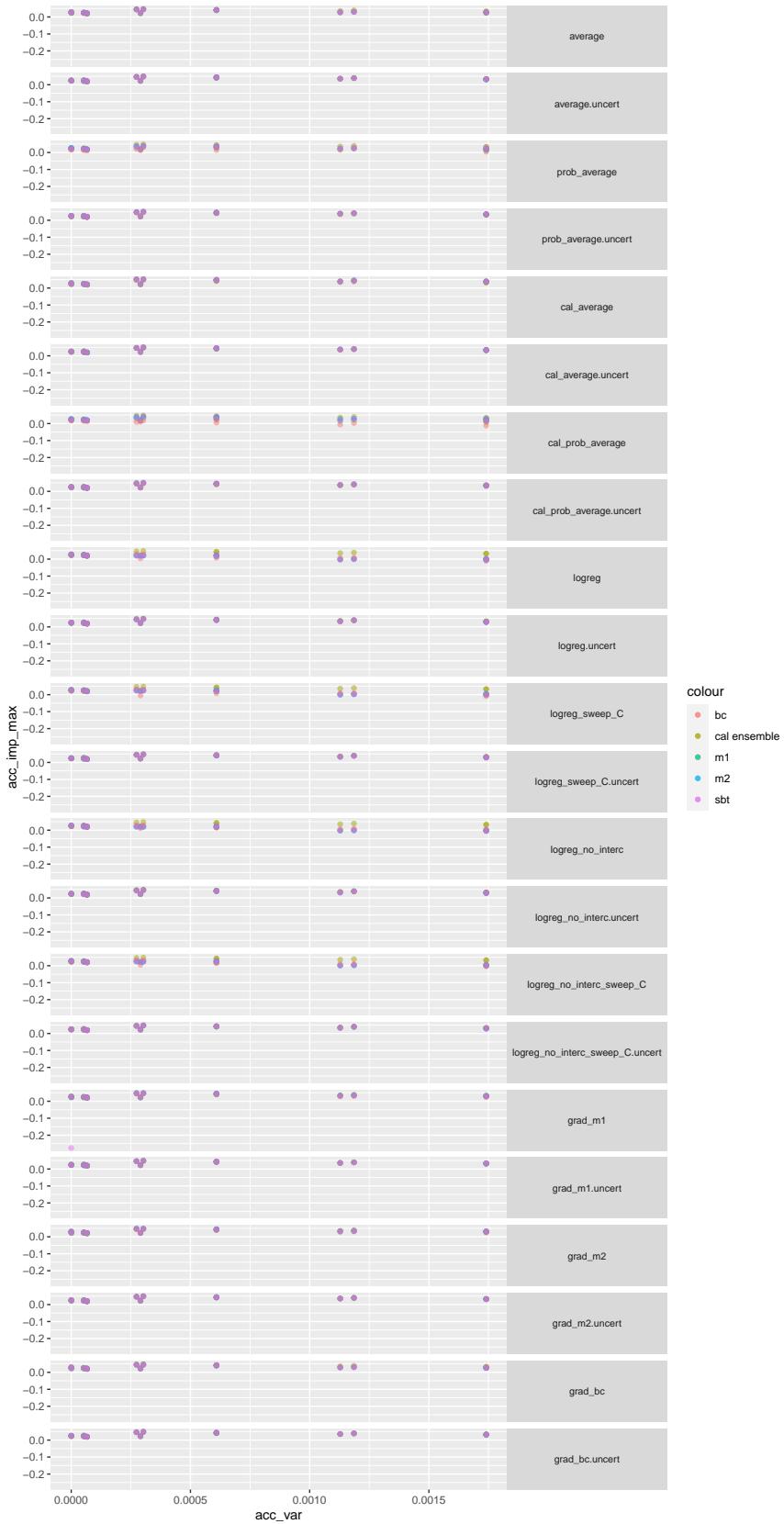
Accuracy improvement of ensemble over
the best of networks vs nll_avg.
Ensemble size 2



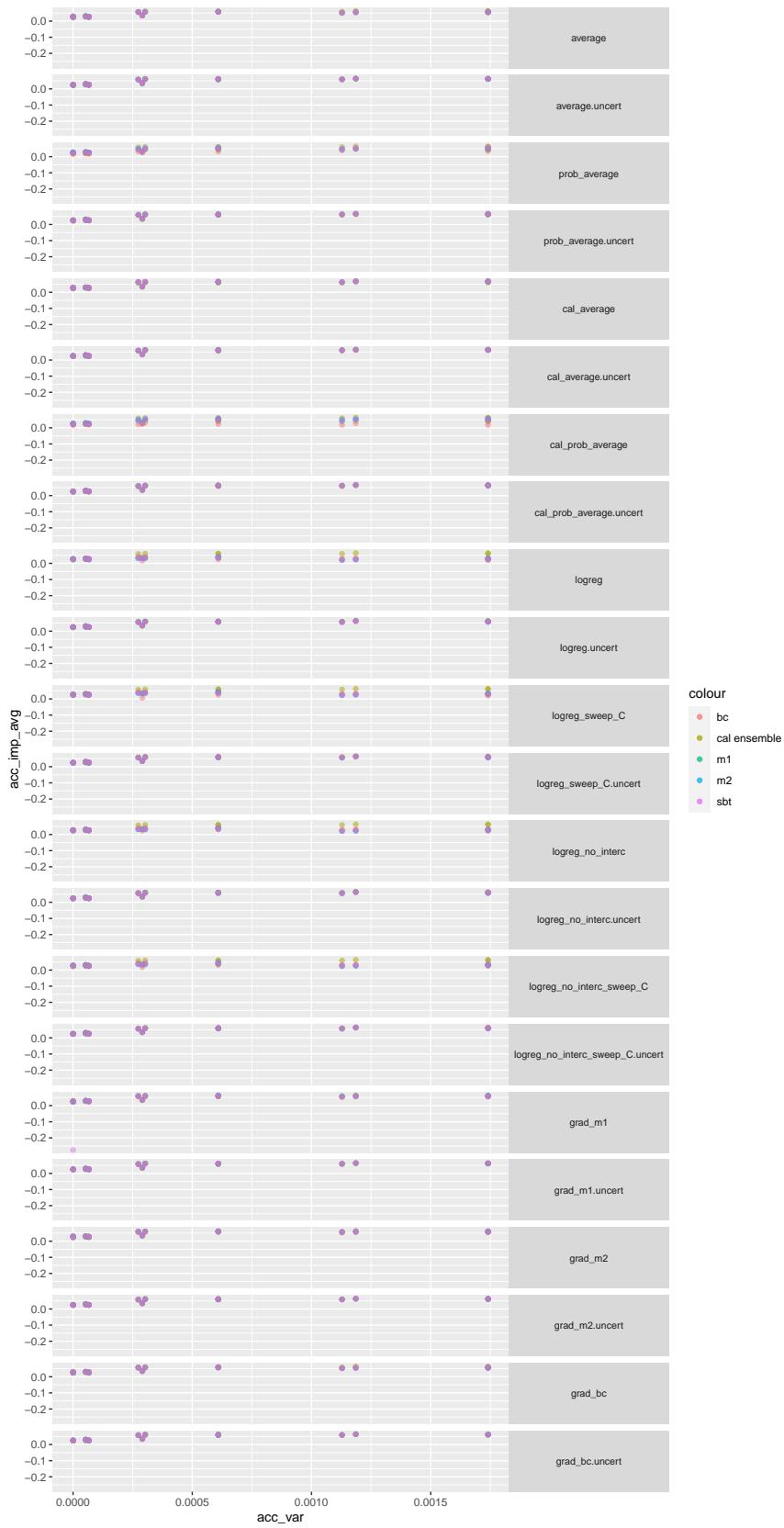
Accuracy improvement of ensemble over
the average of networks vs nll_avg.
Ensemble size 2



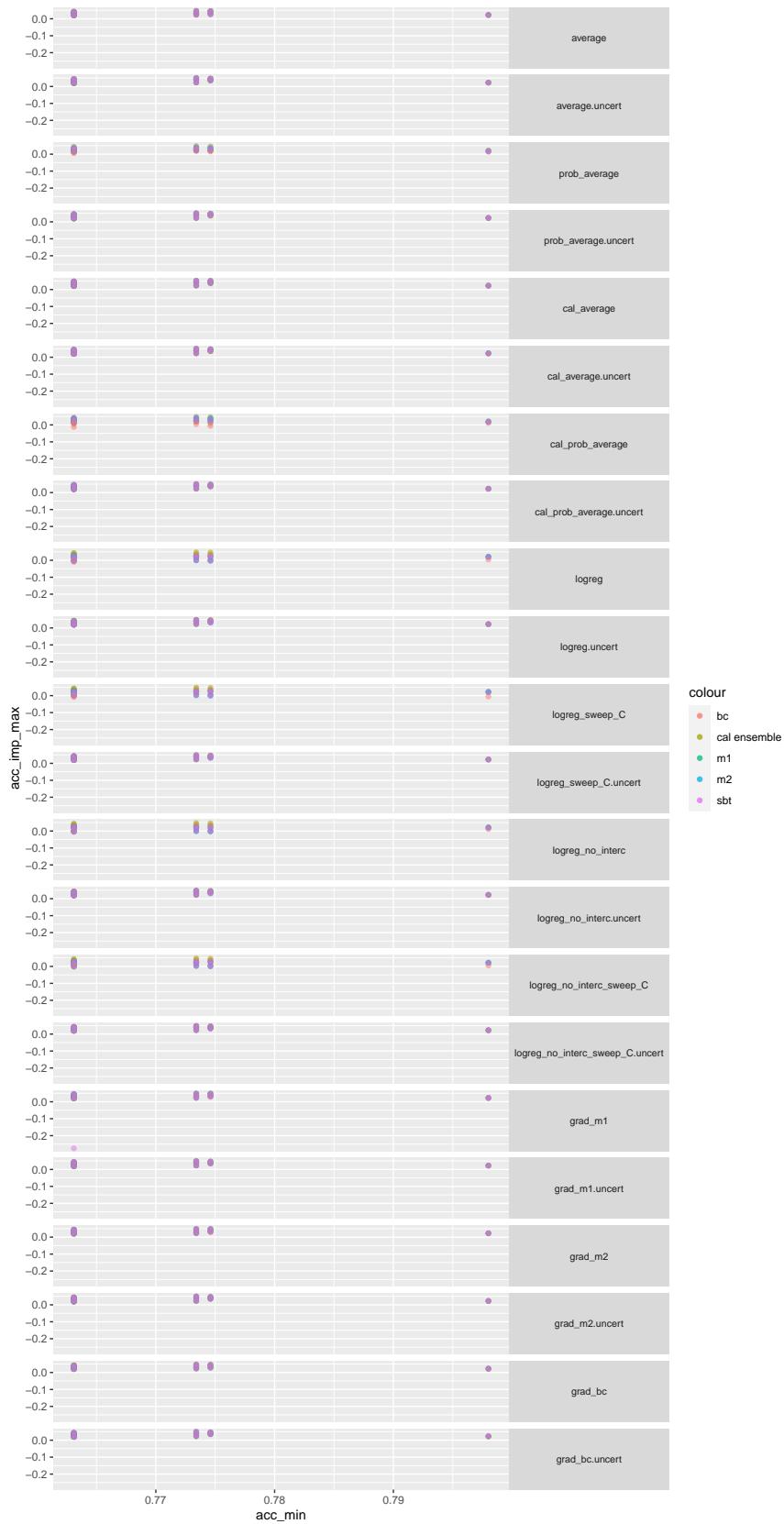
Accuracy improvement of ensemble over
the best of networks vs acc_var.
Ensemble size 2



Accuracy improvement of ensemble over
the average of networks vs acc_var.
Ensemble size 2



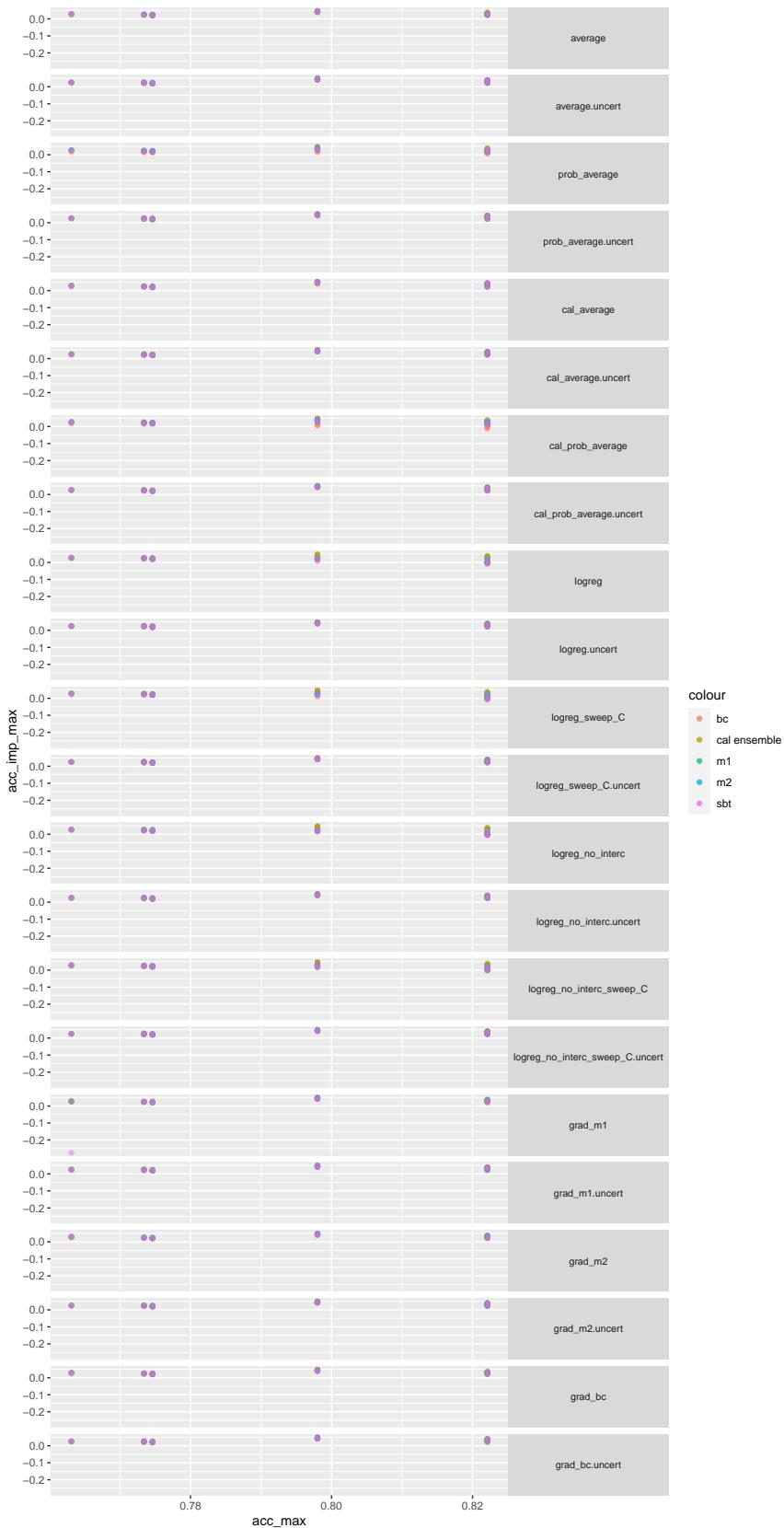
Accuracy improvement of ensemble over
the best of networks vs acc_min.
Ensemble size 2



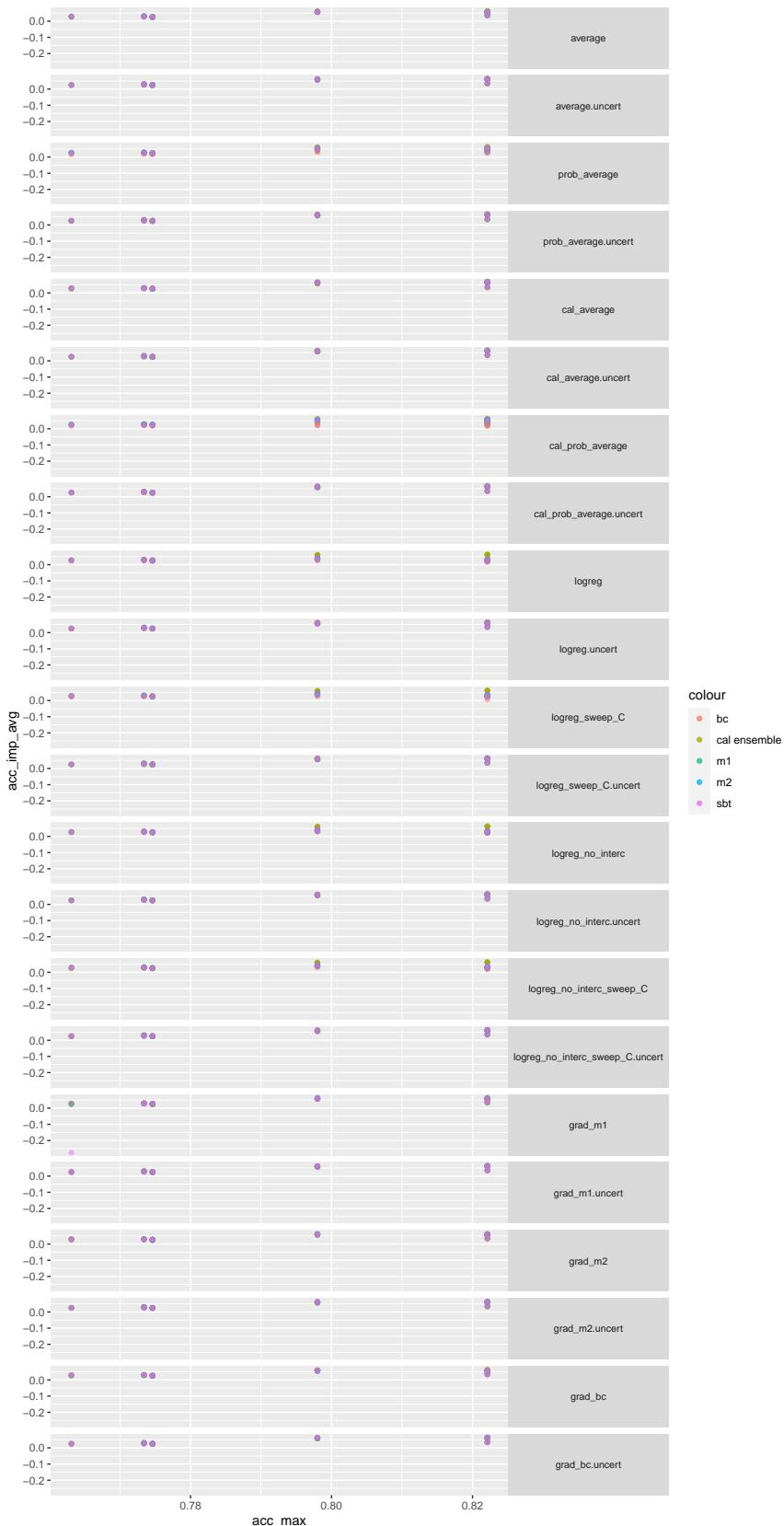
Accuracy improvement of ensemble over
the average of networks vs acc_min.
Ensemble size 2



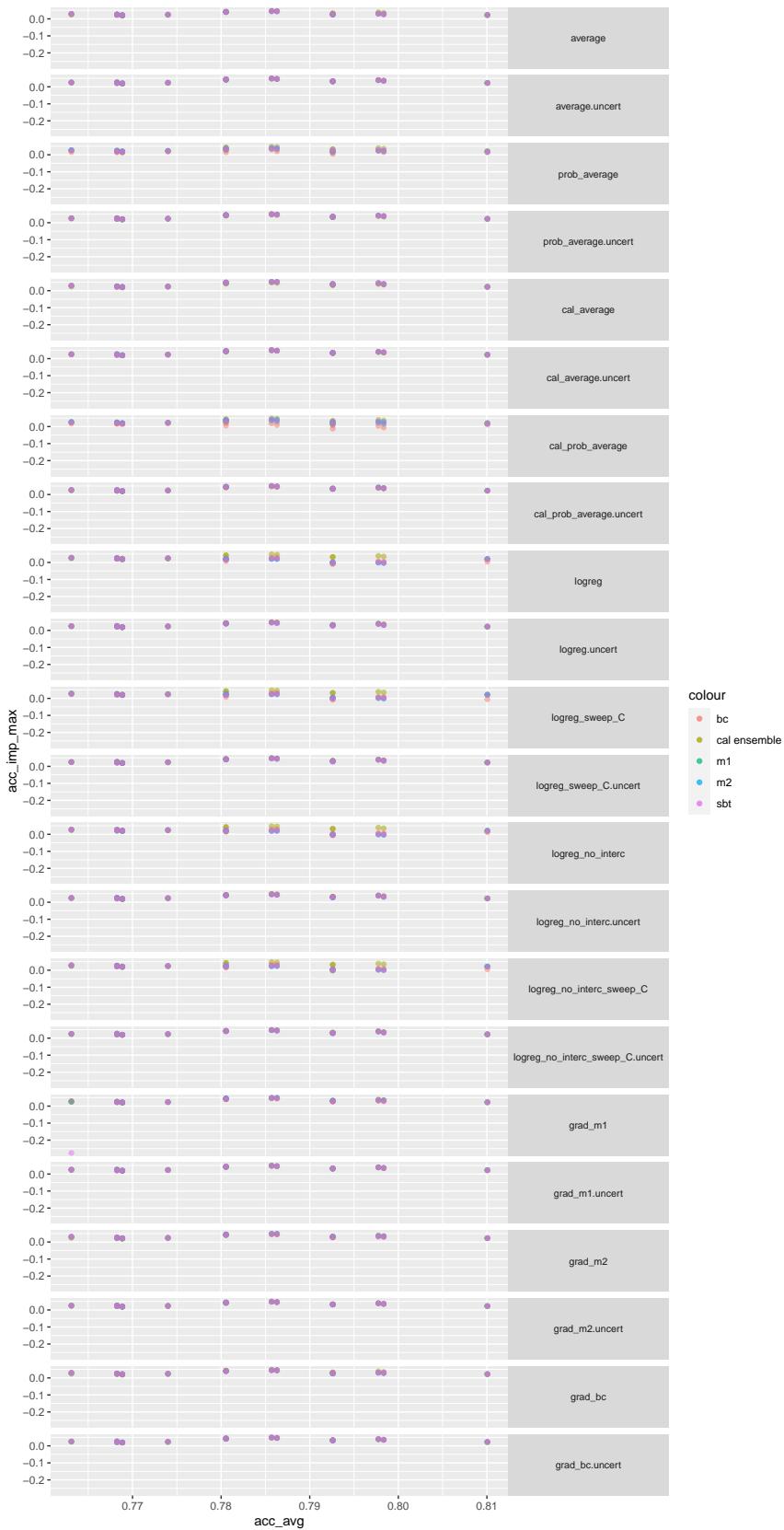
Accuracy improvement of ensemble over
the best of networks vs acc_max.
Ensemble size 2



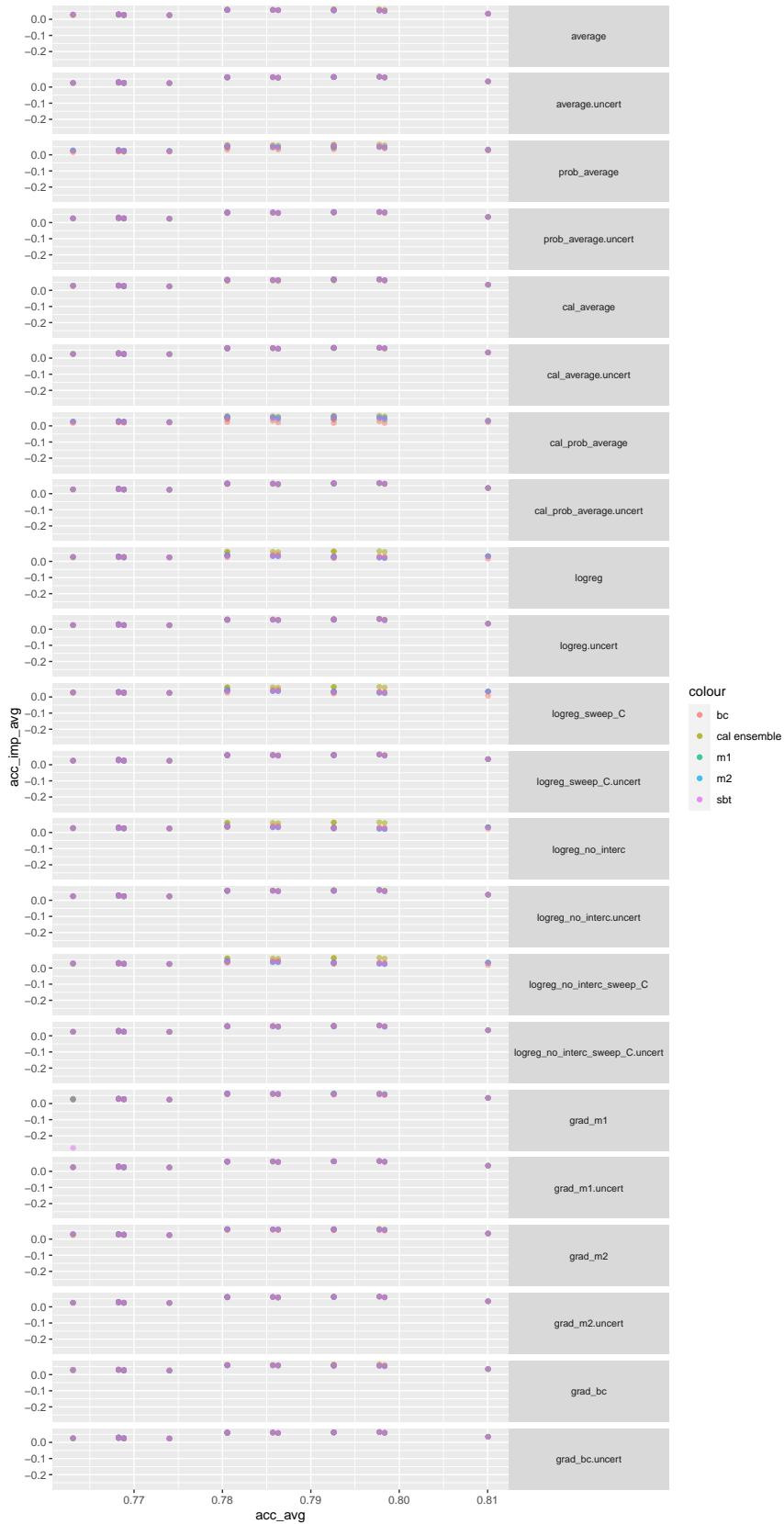
Accuracy improvement of ensemble over
the average of networks vs acc_max.
Ensemble size 2



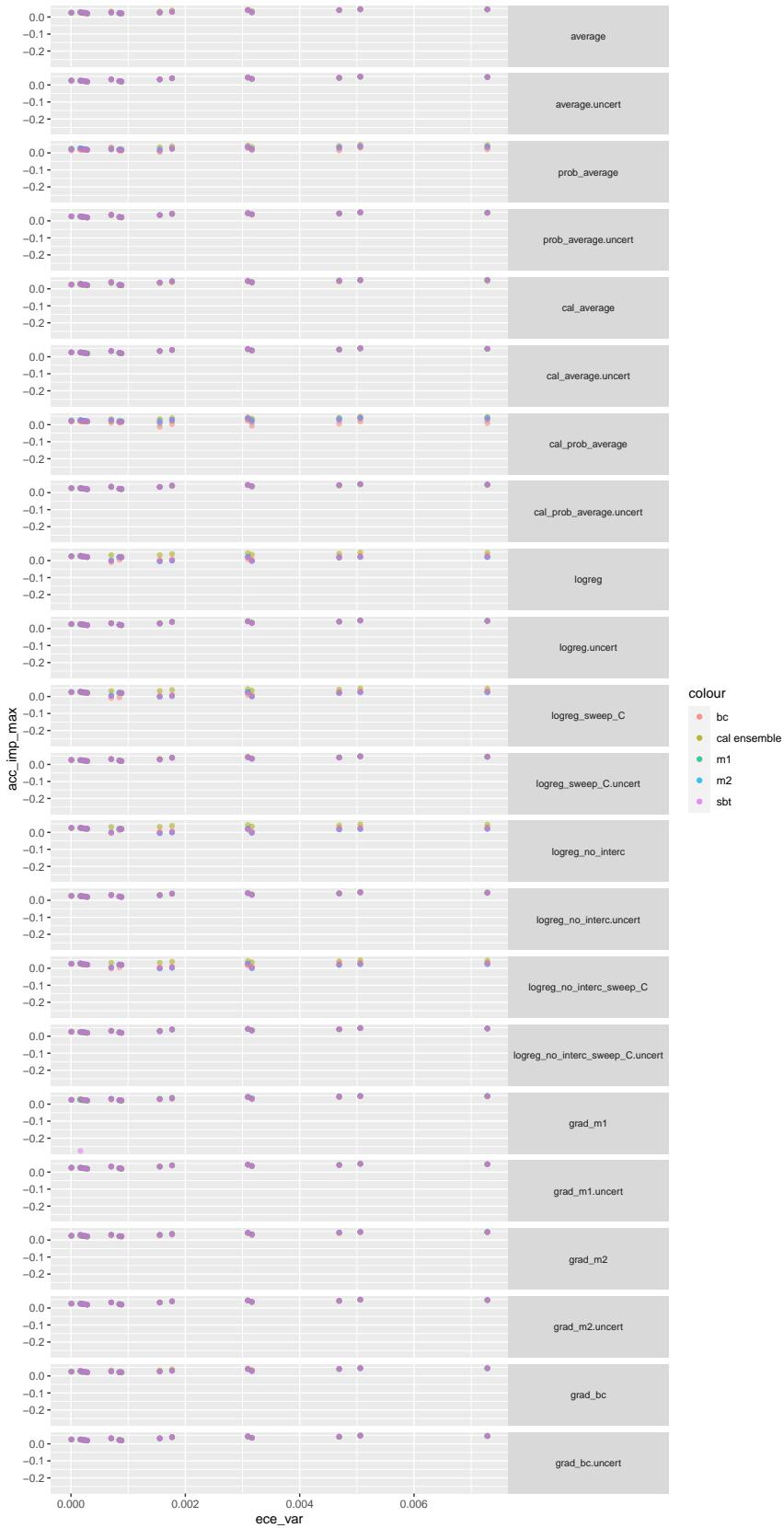
Accuracy improvement of ensemble over
the best of networks vs acc_avg.
Ensemble size 2



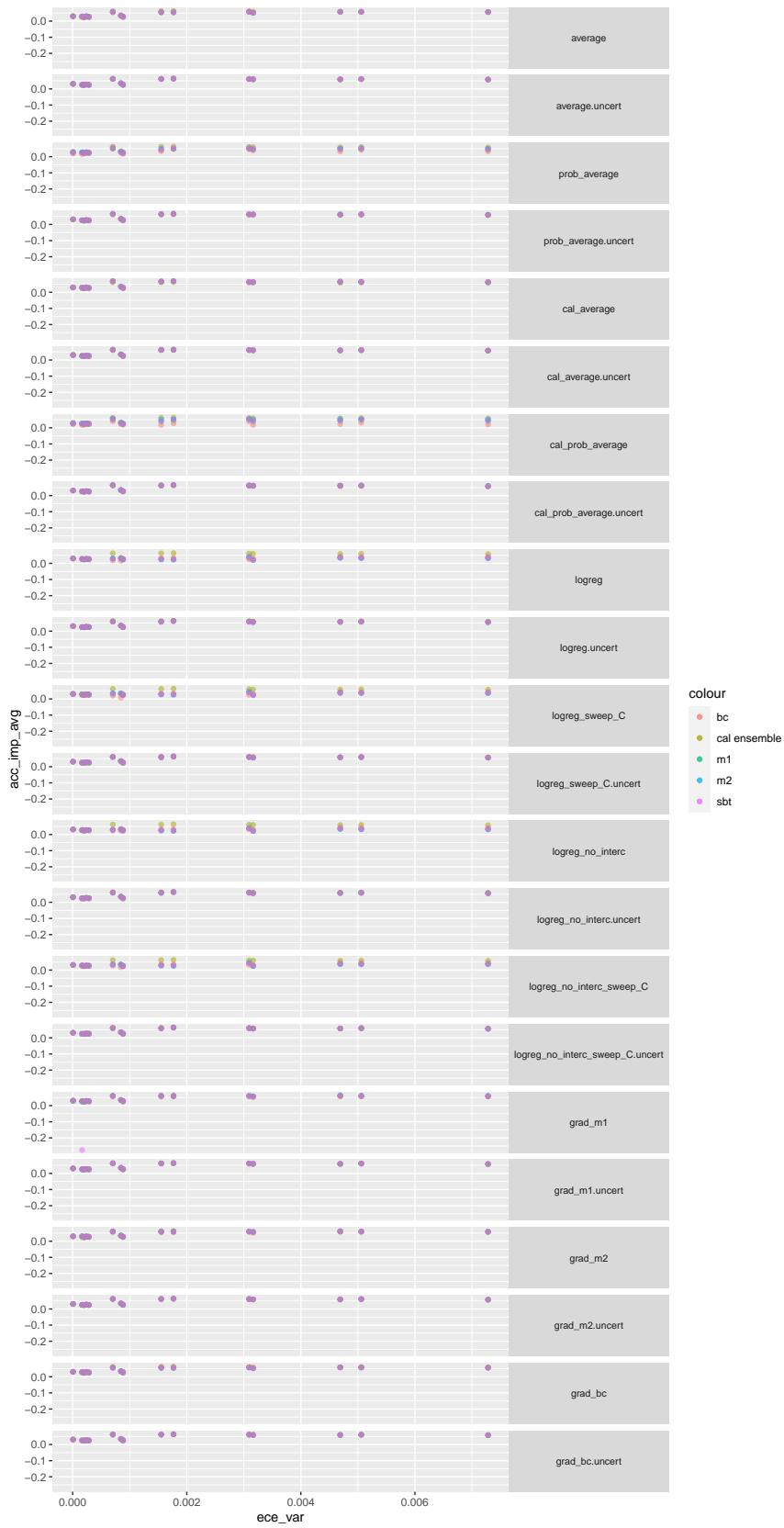
Accuracy improvement of ensemble over
the average of networks vs acc_avg.
Ensemble size 2



Accuracy improvement of ensemble over
the best of networks vs ece_var.
Ensemble size 2



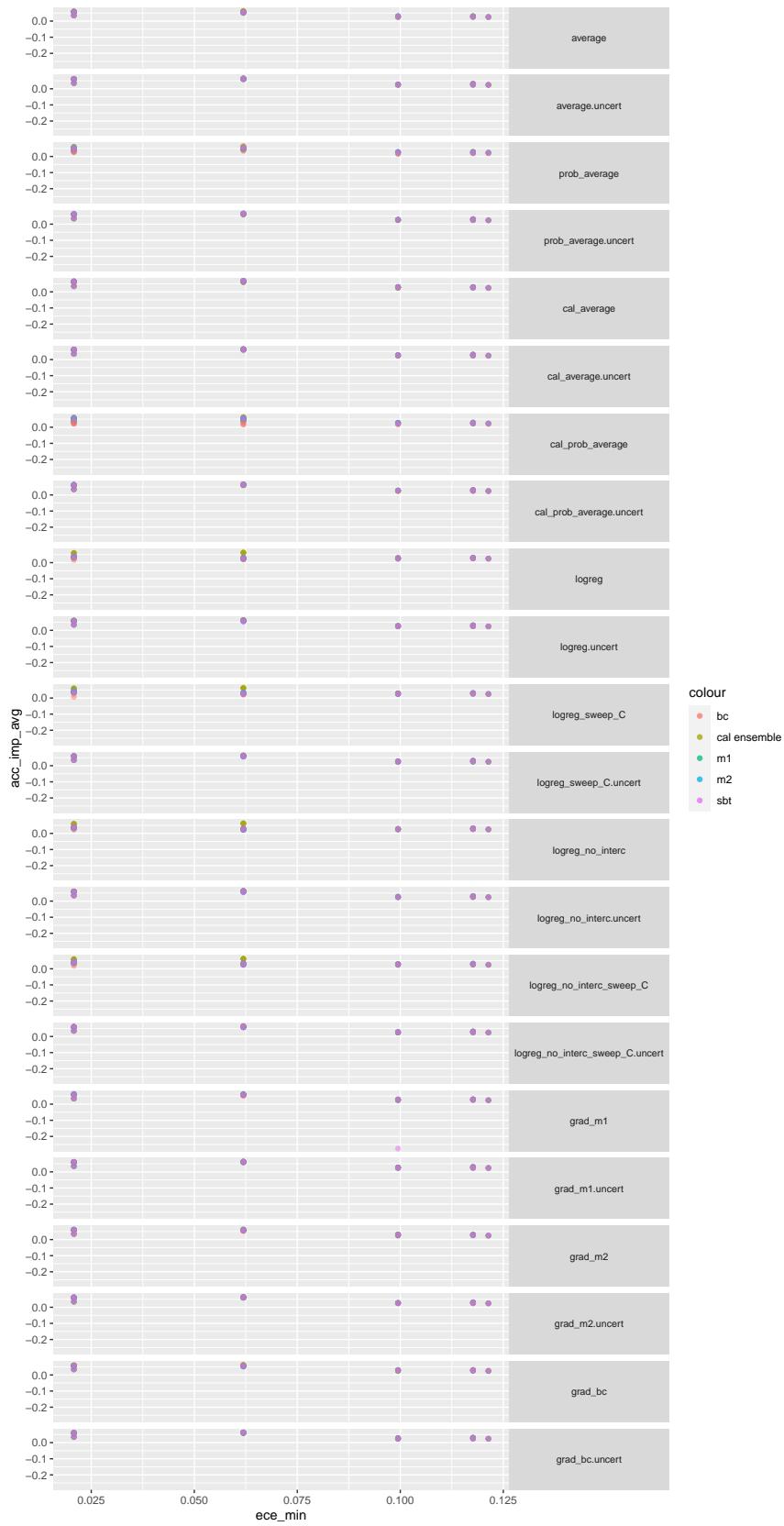
Accuracy improvement of ensemble over
the average of networks vs ece_var.
Ensemble size 2



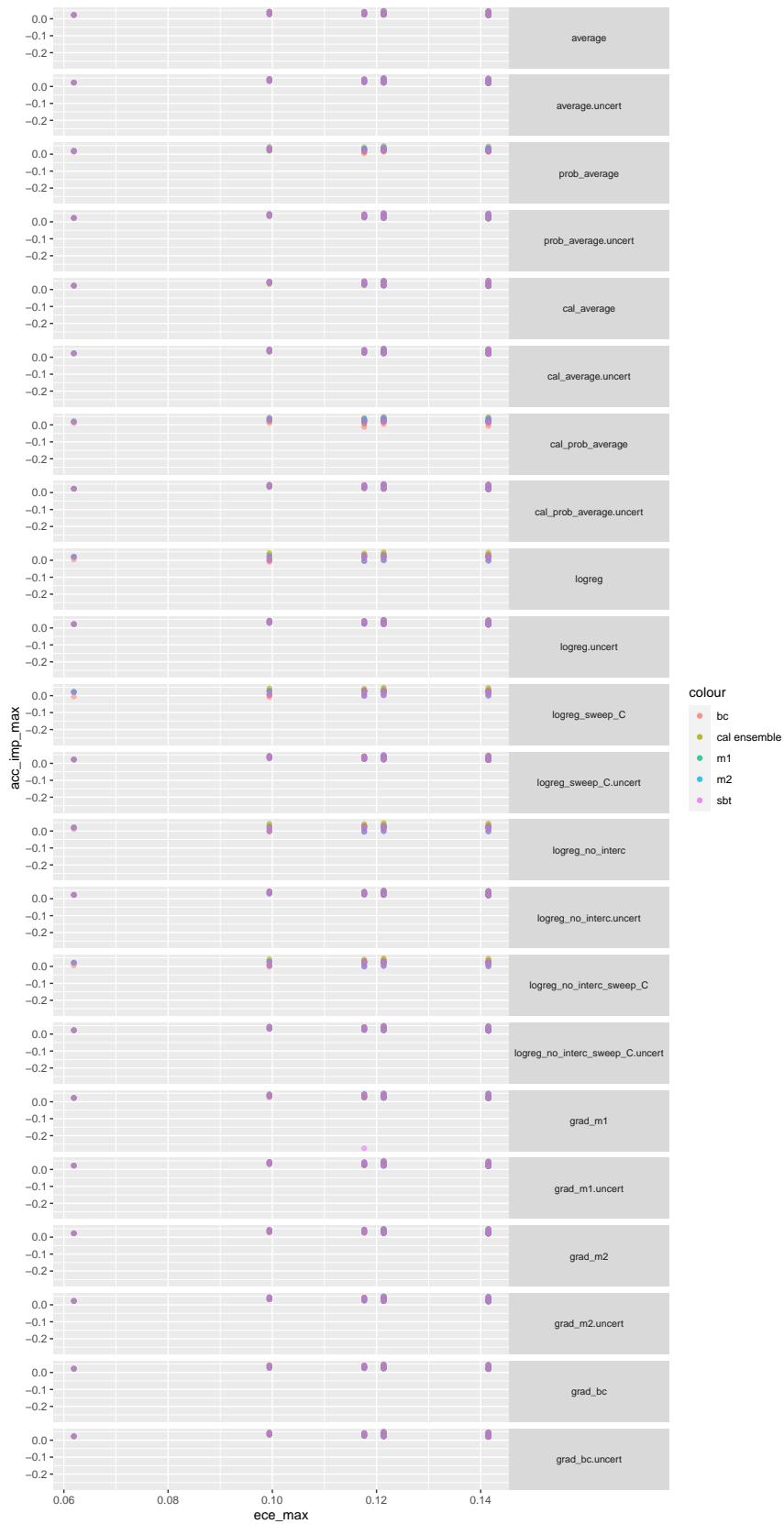
Accuracy improvement of ensemble over
the best of networks vs ece_min.
Ensemble size 2



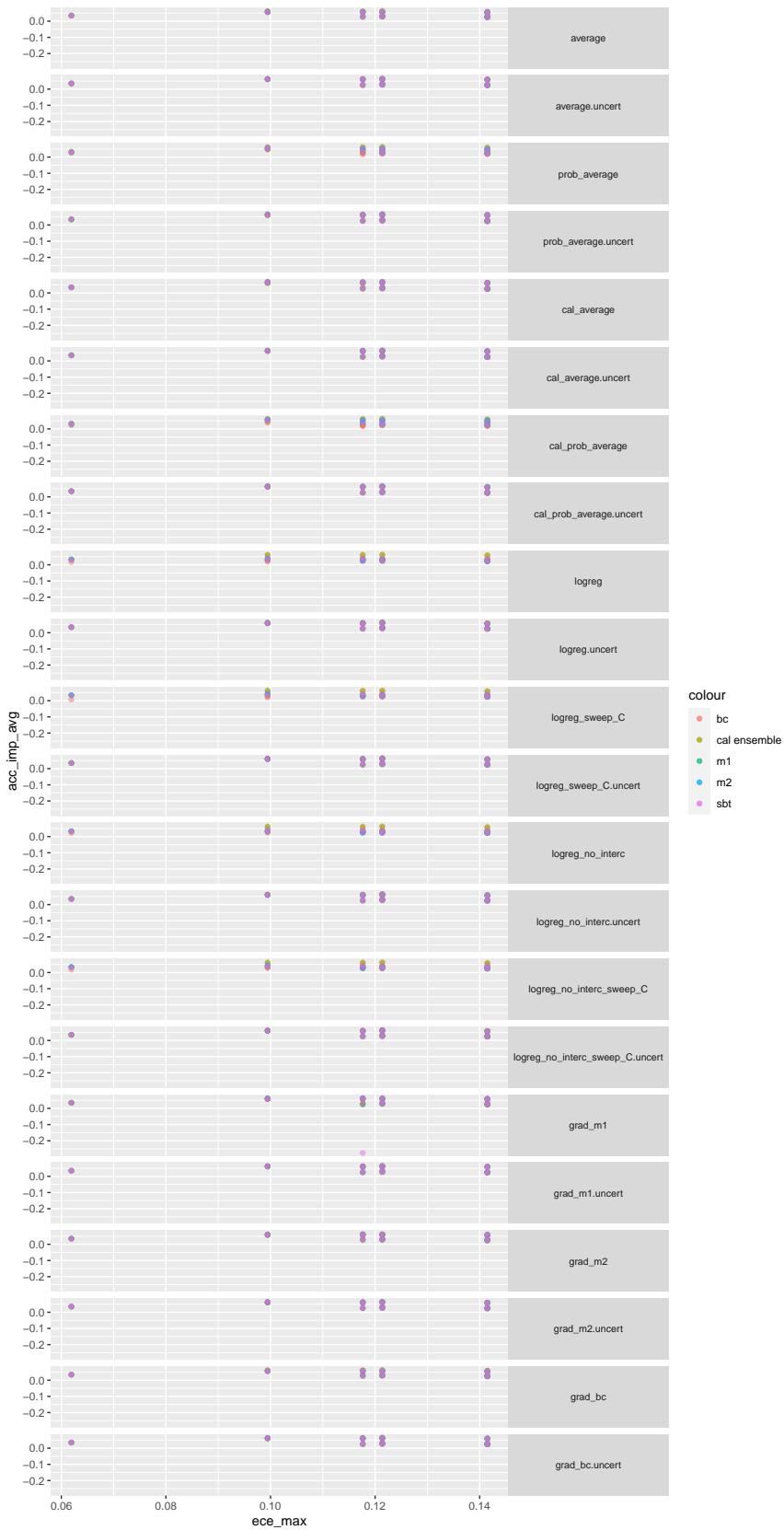
Accuracy improvement of ensemble over
the average of networks vs ece_min.
Ensemble size 2



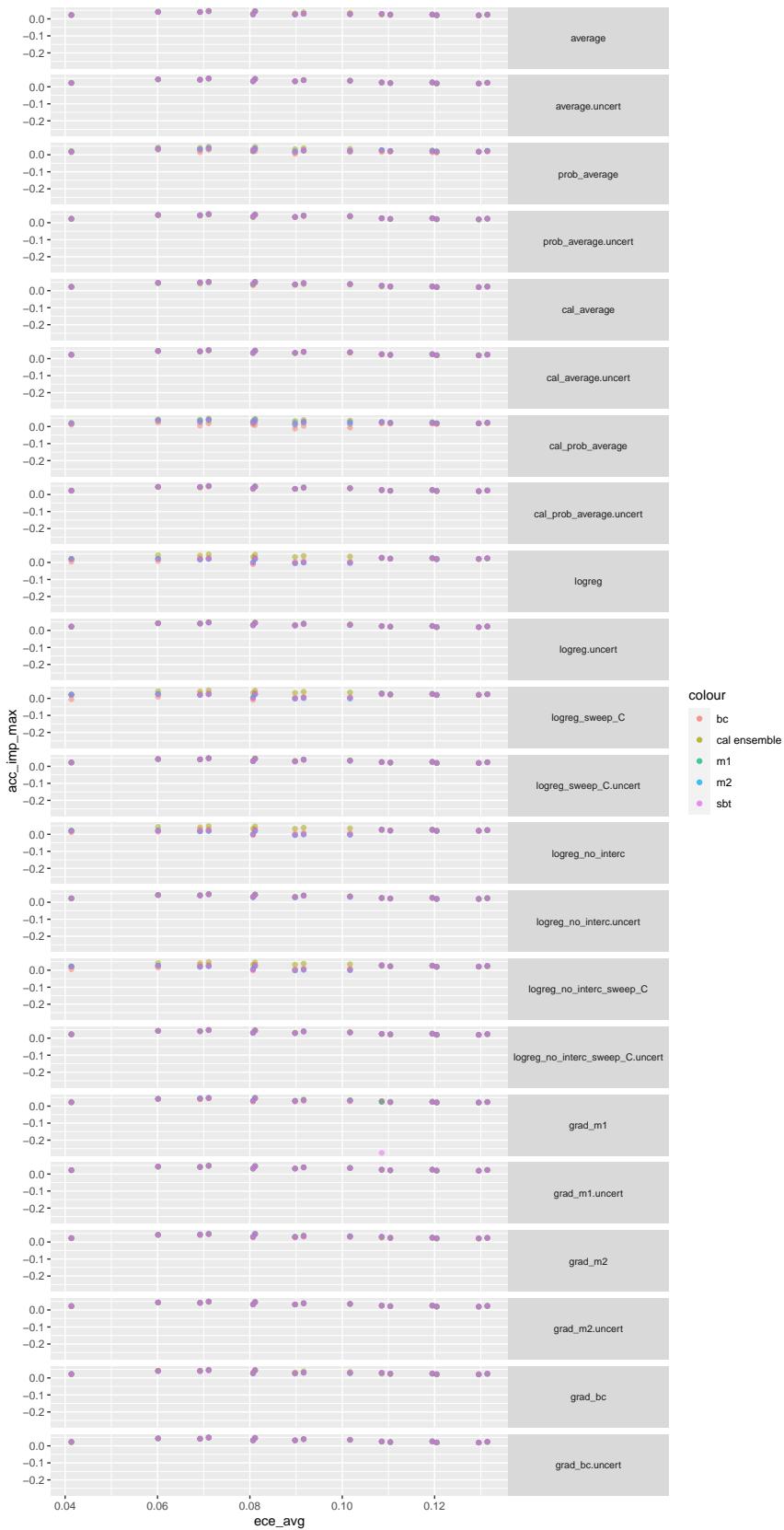
Accuracy improvement of ensemble over
the best of networks vs ece_max.
Ensemble size 2



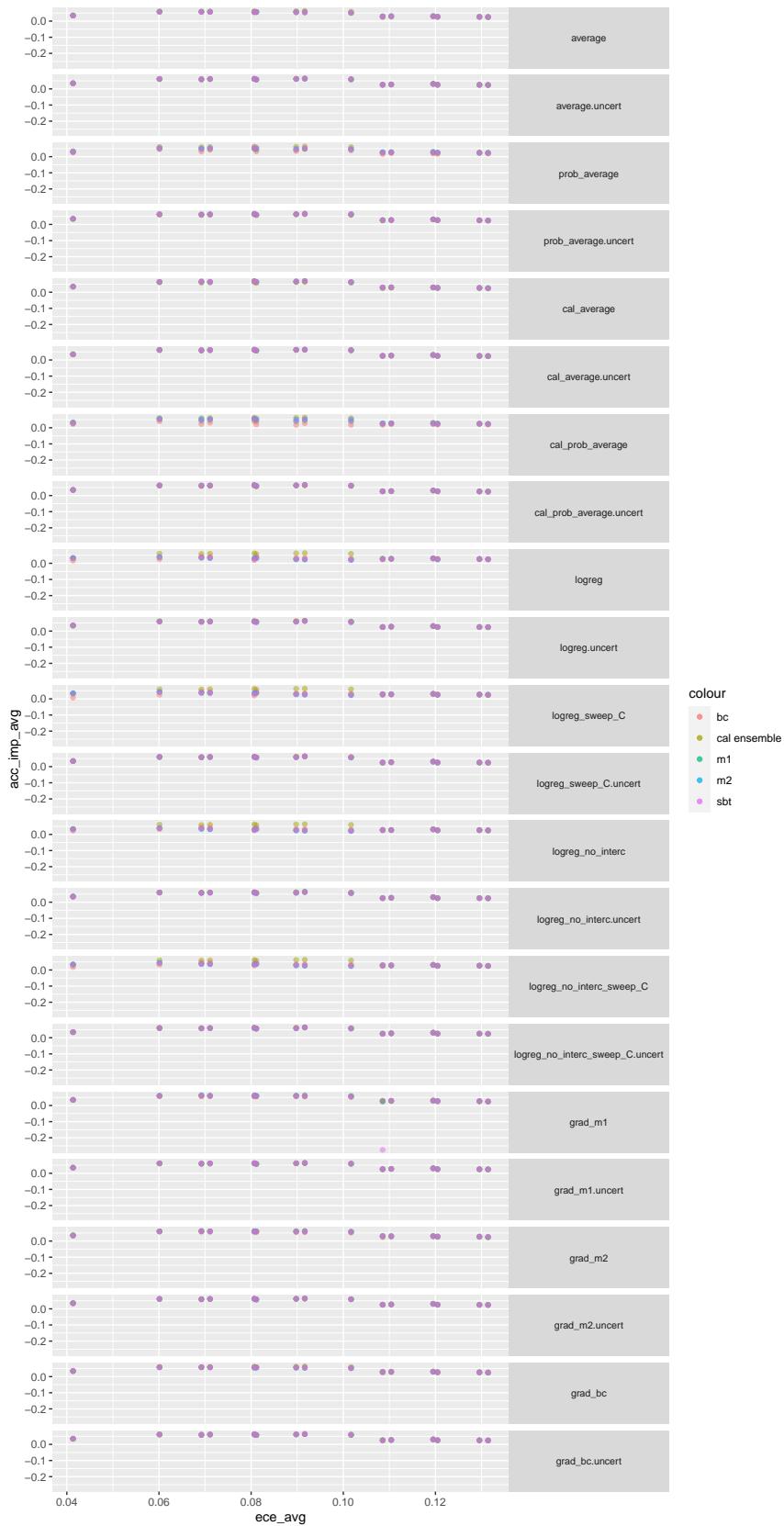
Accuracy improvement of ensemble over
the average of networks vs ece_max.
Ensemble size 2



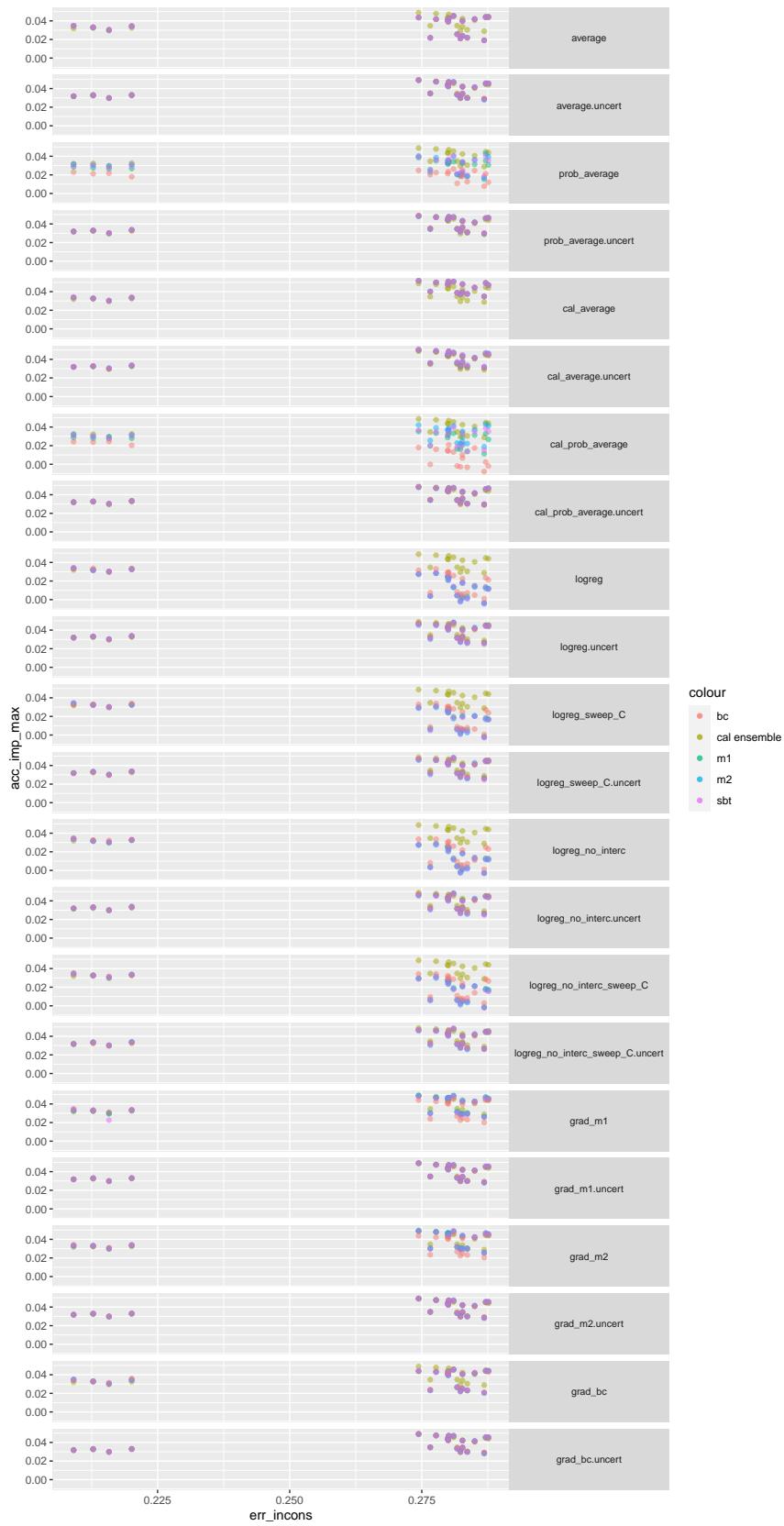
Accuracy improvement of ensemble over
the best of networks vs ece_avg.
Ensemble size 2



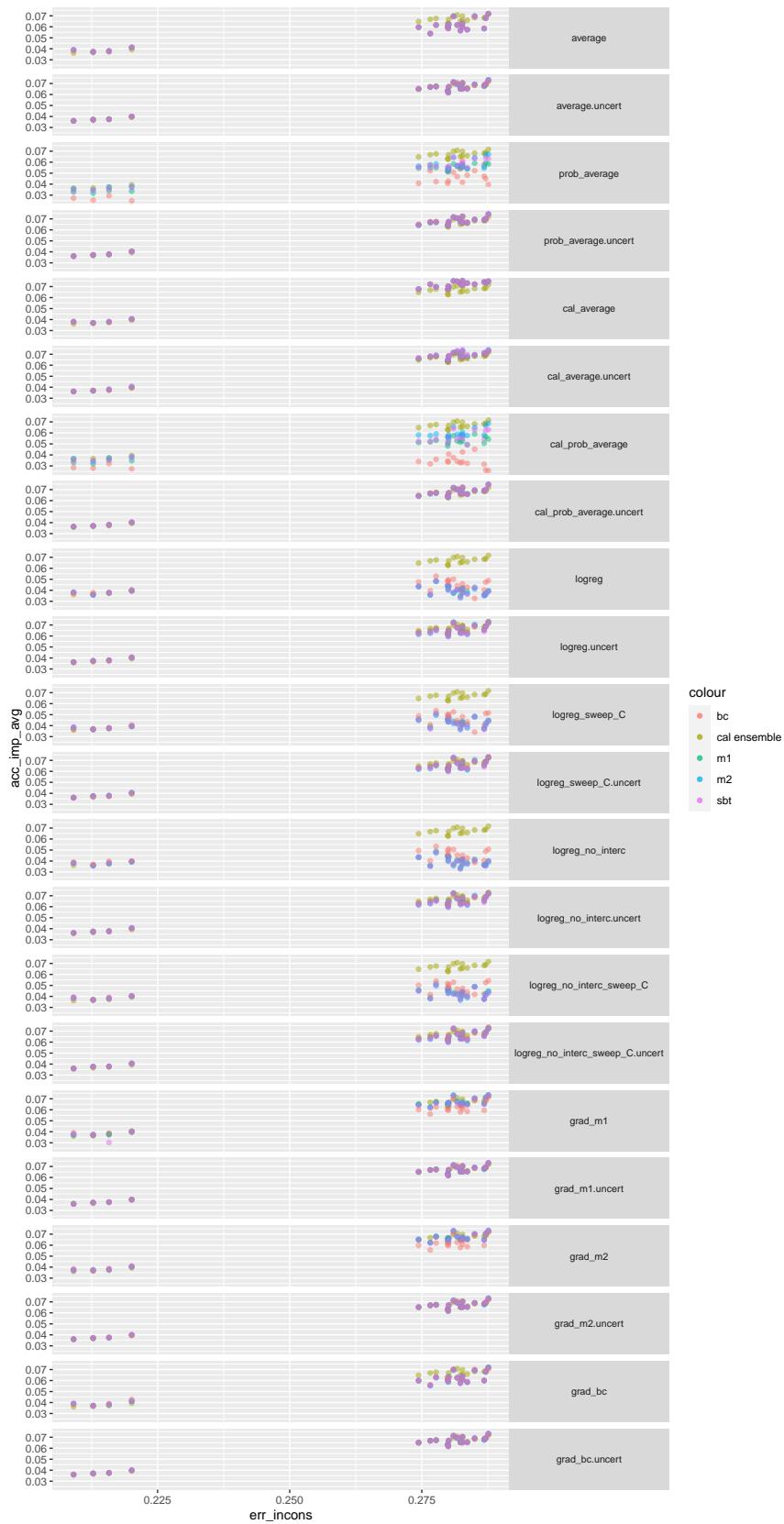
Accuracy improvement of ensemble over
the average of networks vs ece_avg.
Ensemble size 2



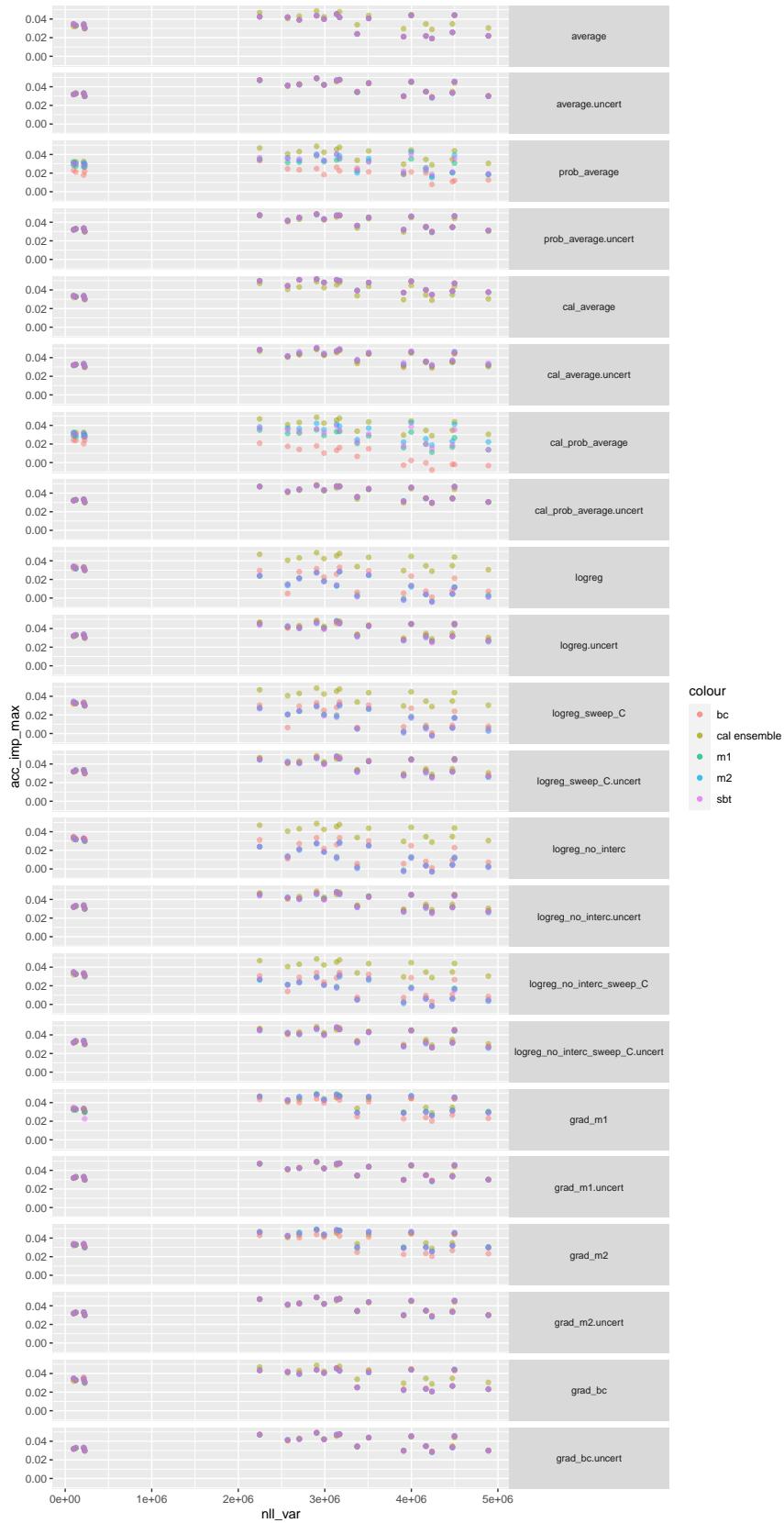
Accuracy improvement of ensemble over
the best of networks vs err_incons.
Ensemble size 3



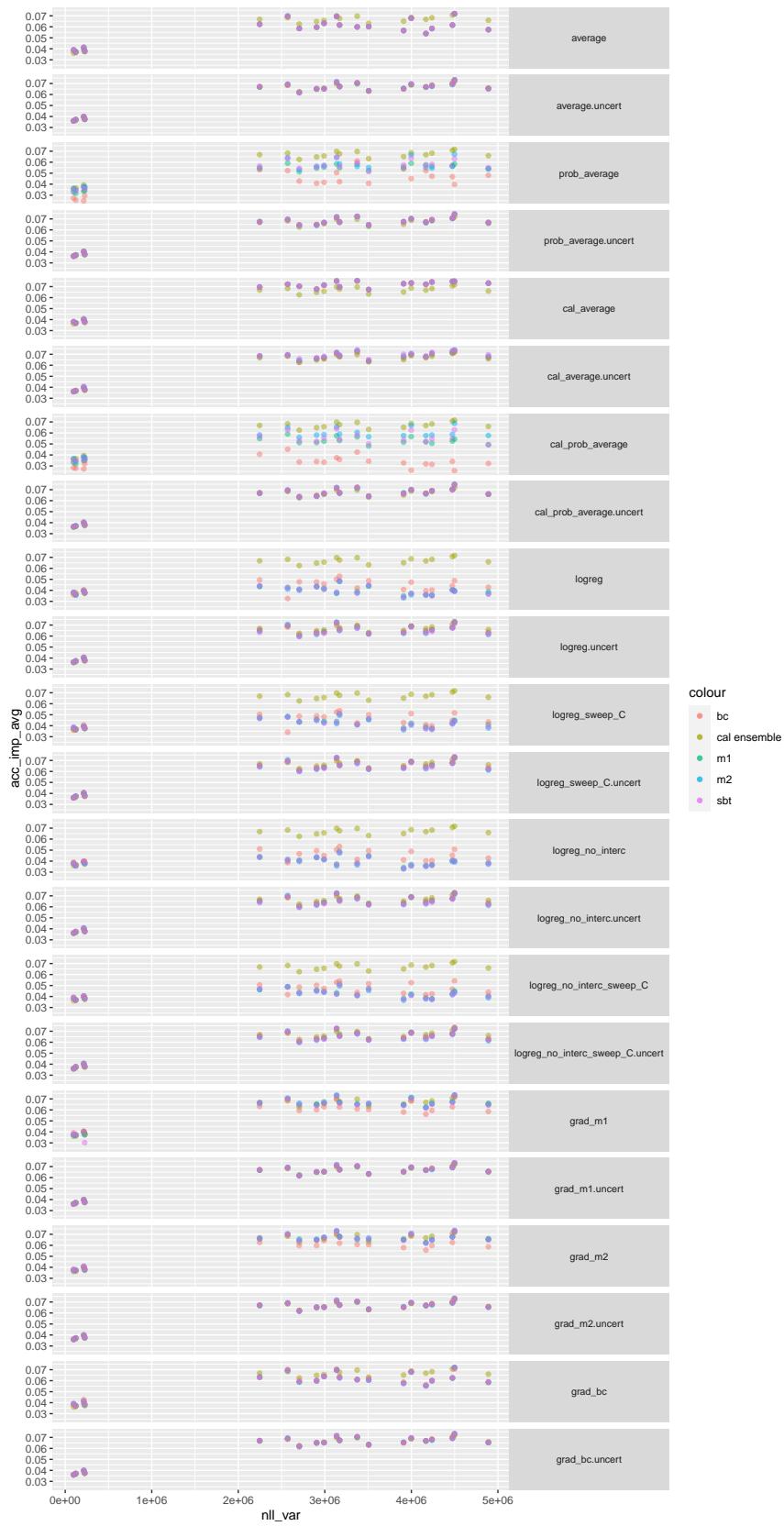
Accuracy improvement of ensemble over
the average of networks vs err_incons.
Ensemble size 3



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



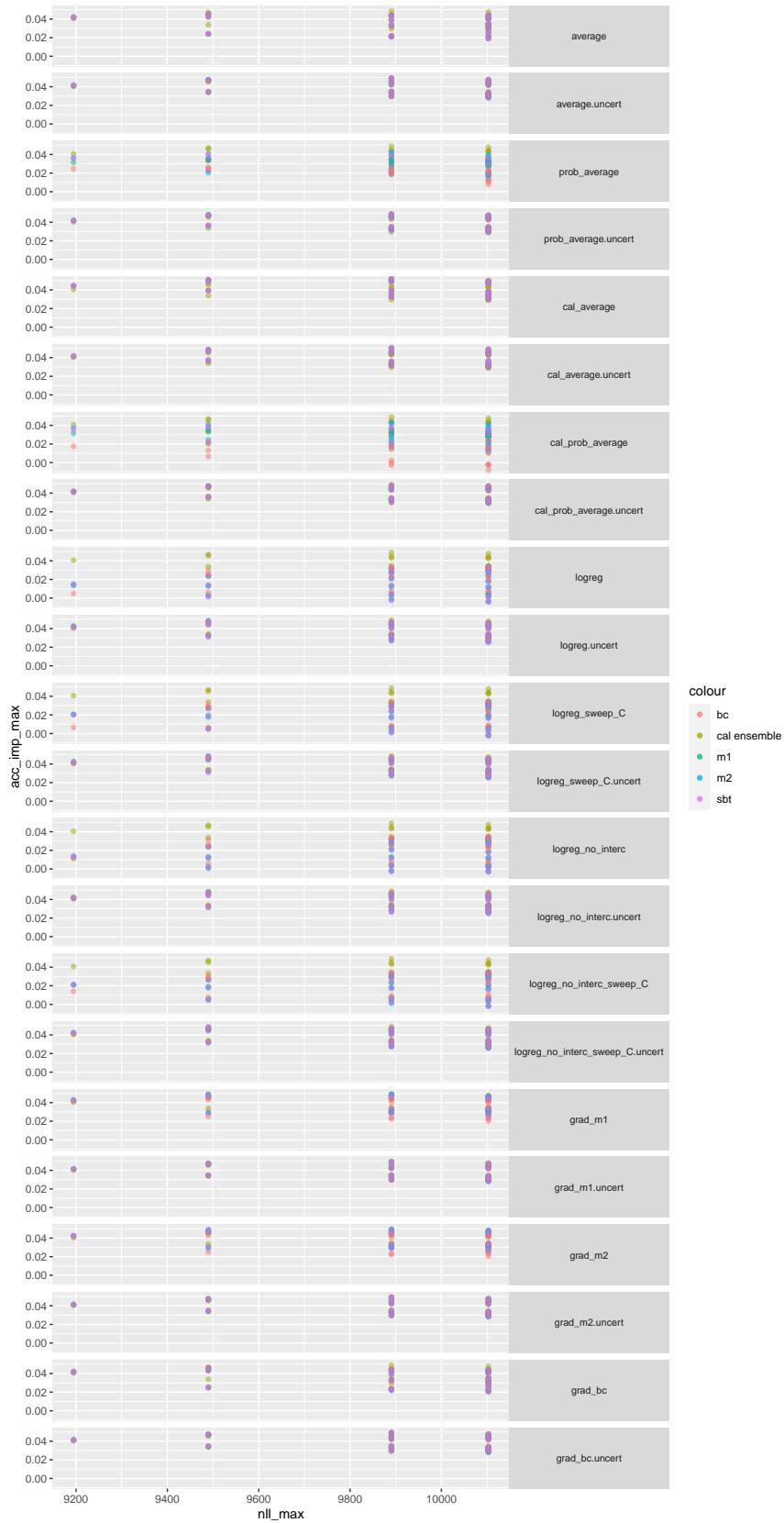
Accuracy improvement of ensemble over
the best of networks vs nll_min.
Ensemble size 3



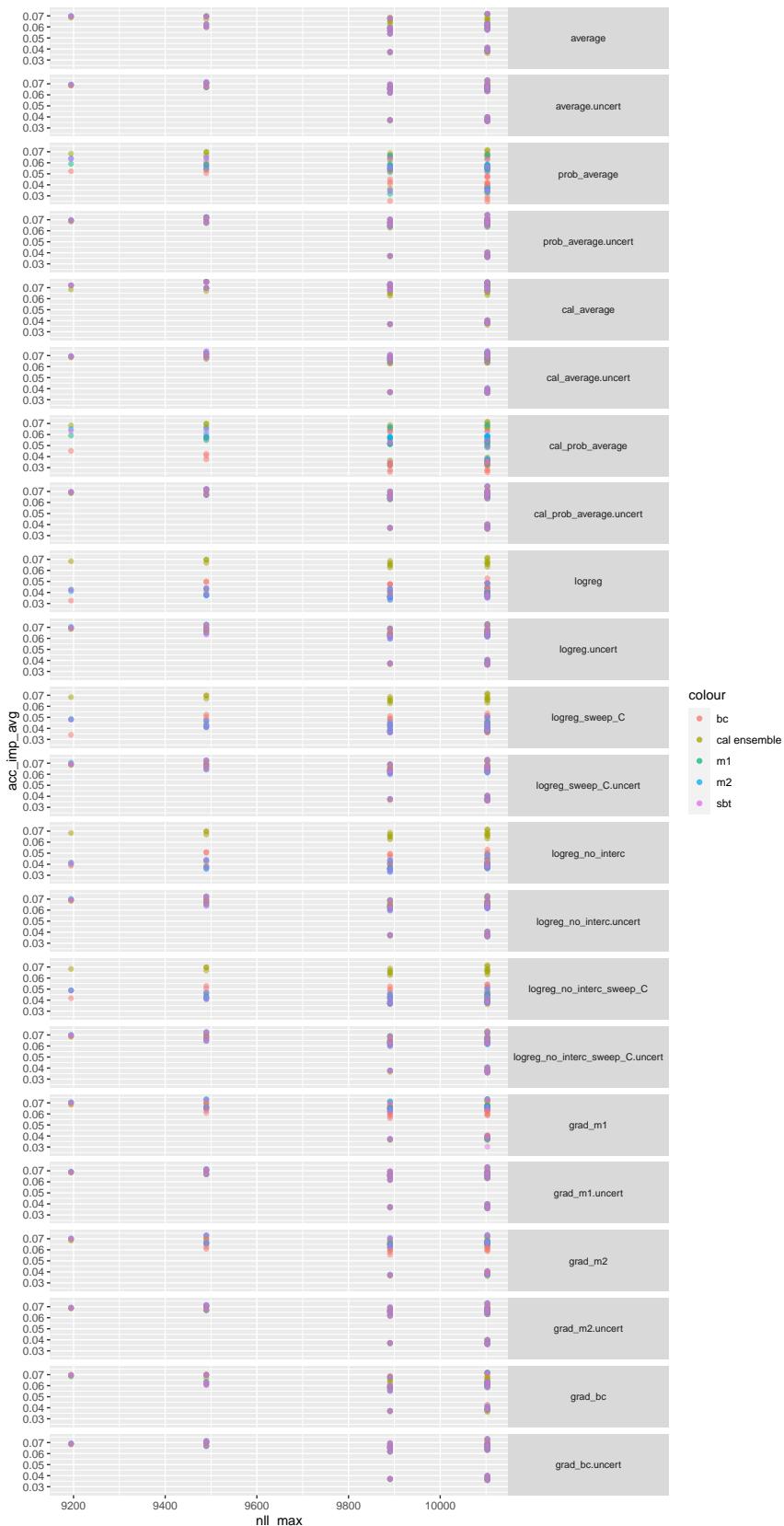
Accuracy improvement of ensemble over
the average of networks vs nll_min.
Ensemble size 3



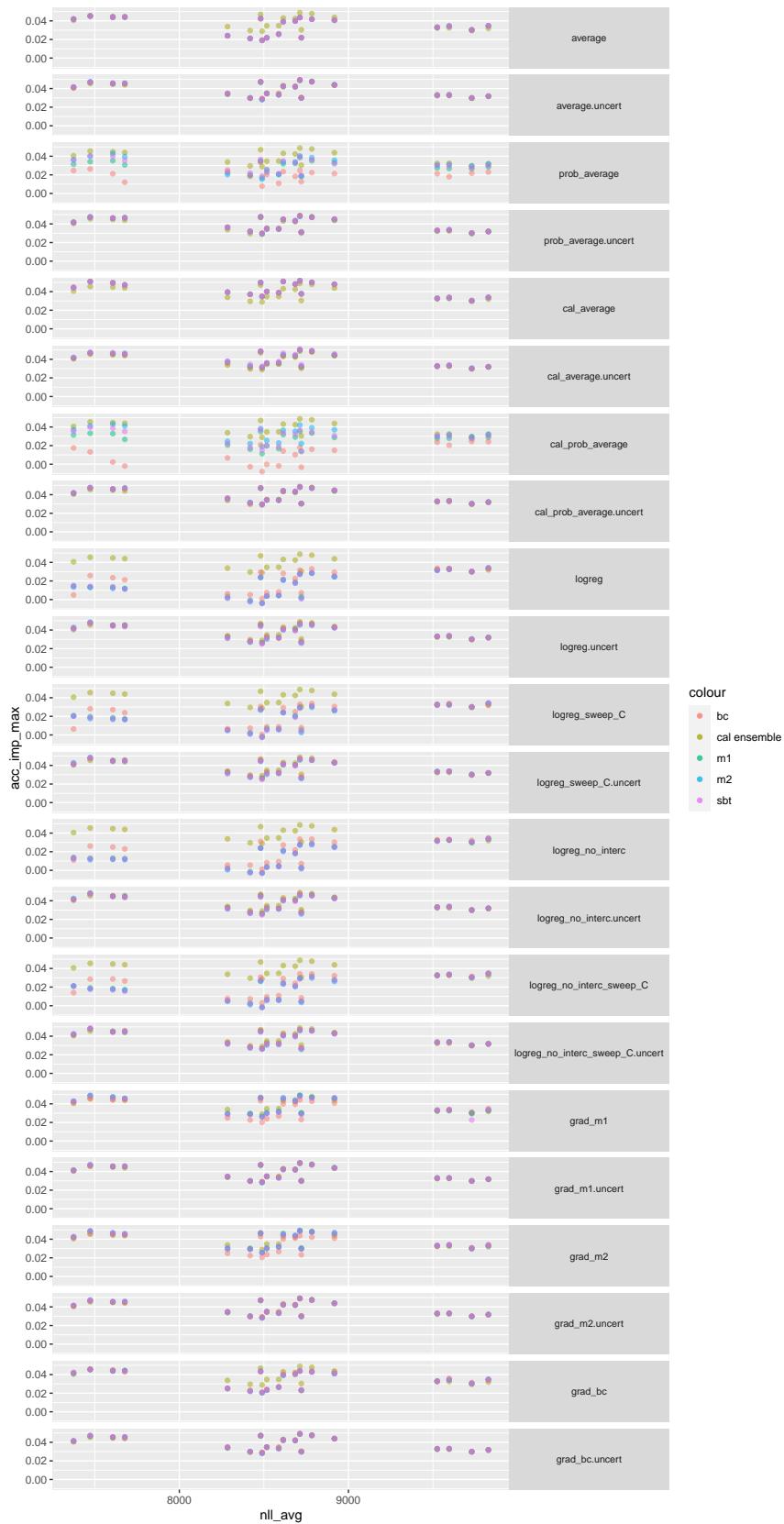
Accuracy improvement of ensemble over
the best of networks vs nll_max.
Ensemble size 3



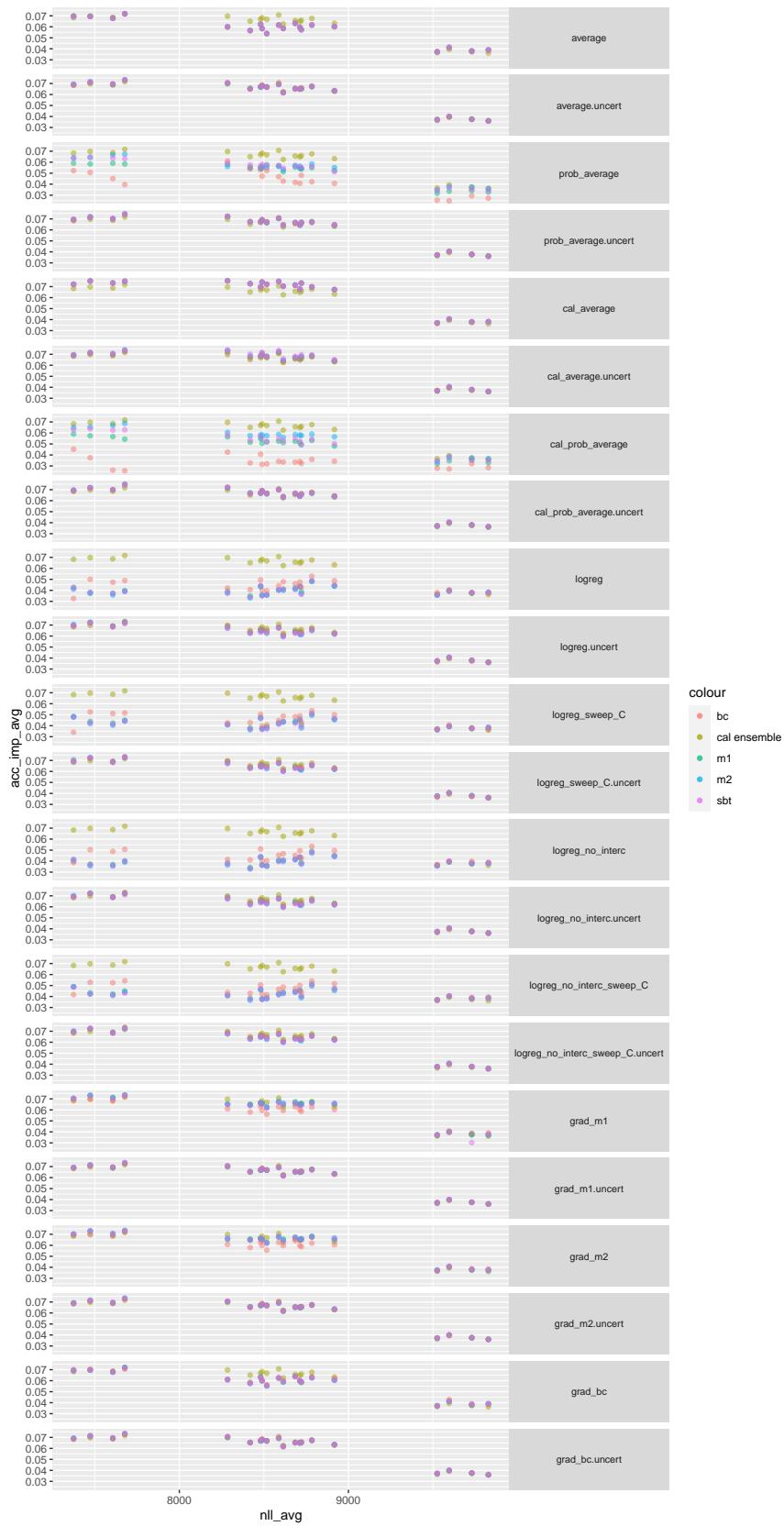
Accuracy improvement of ensemble over
the average of networks vs nll_max.
Ensemble size 3



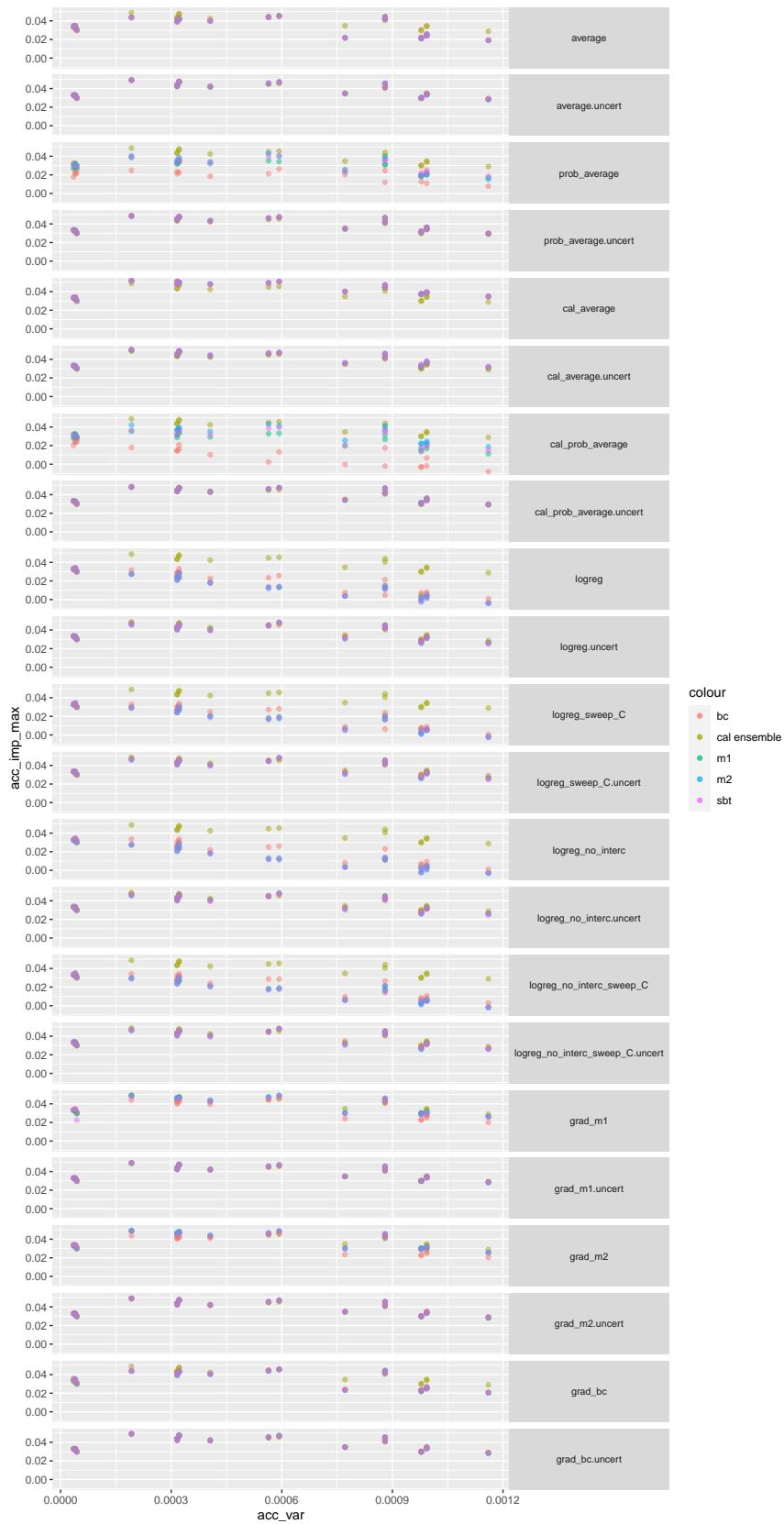
Accuracy improvement of ensemble over
the best of networks vs nll_avg.
Ensemble size 3



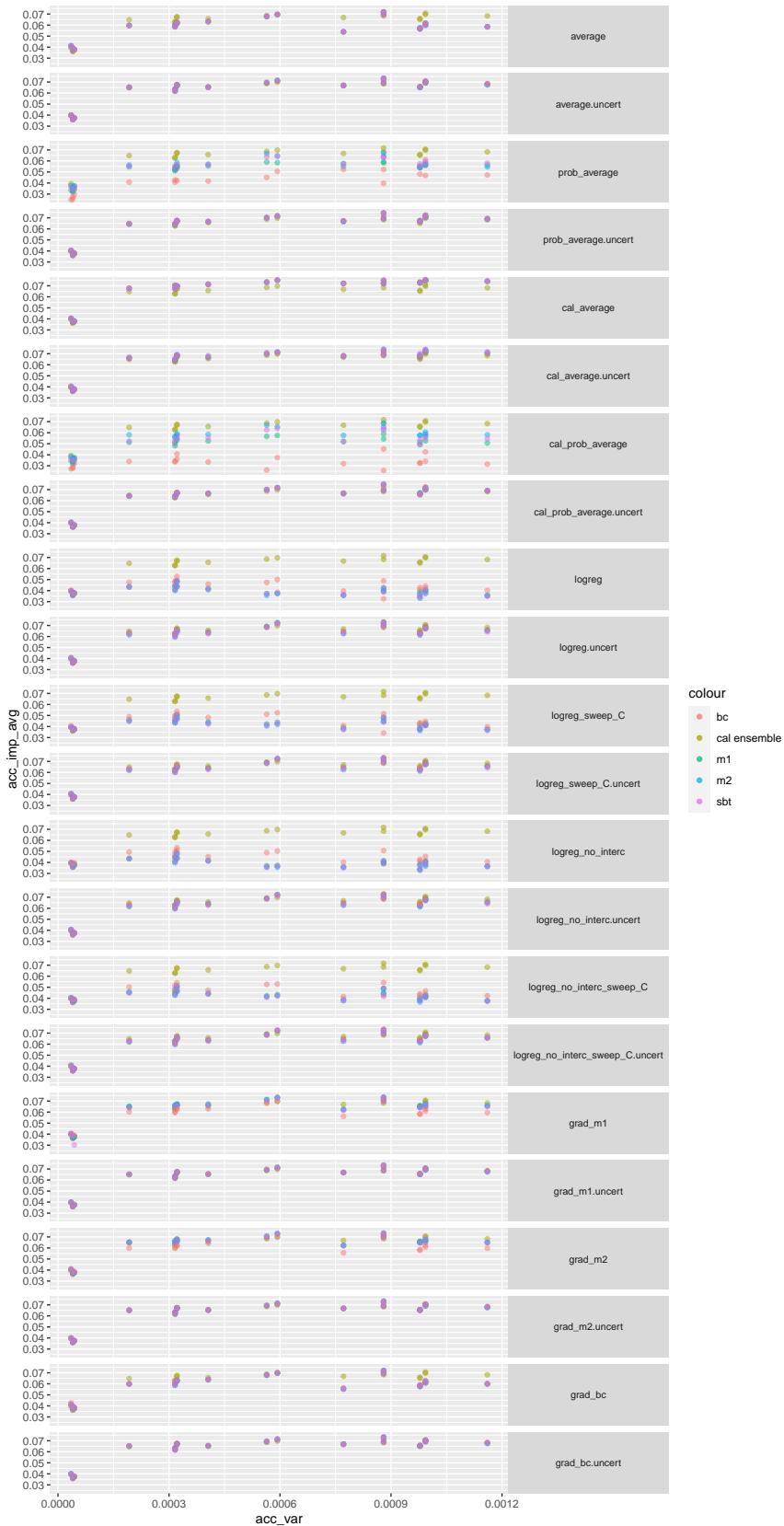
Accuracy improvement of ensemble over
the average of networks vs nll_avg.
Ensemble size 3



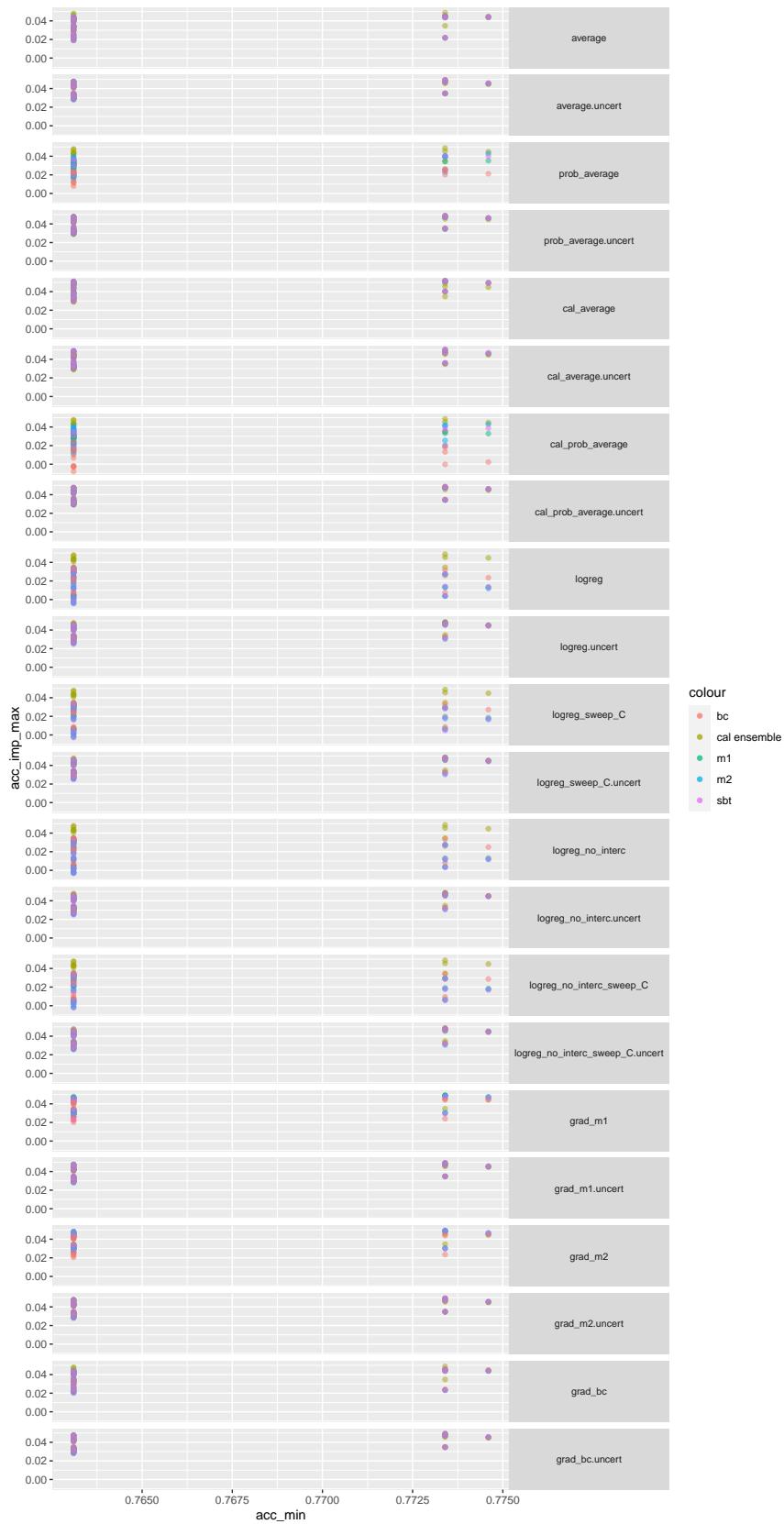
Accuracy improvement of ensemble over
the best of networks vs acc_var.
Ensemble size 3



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



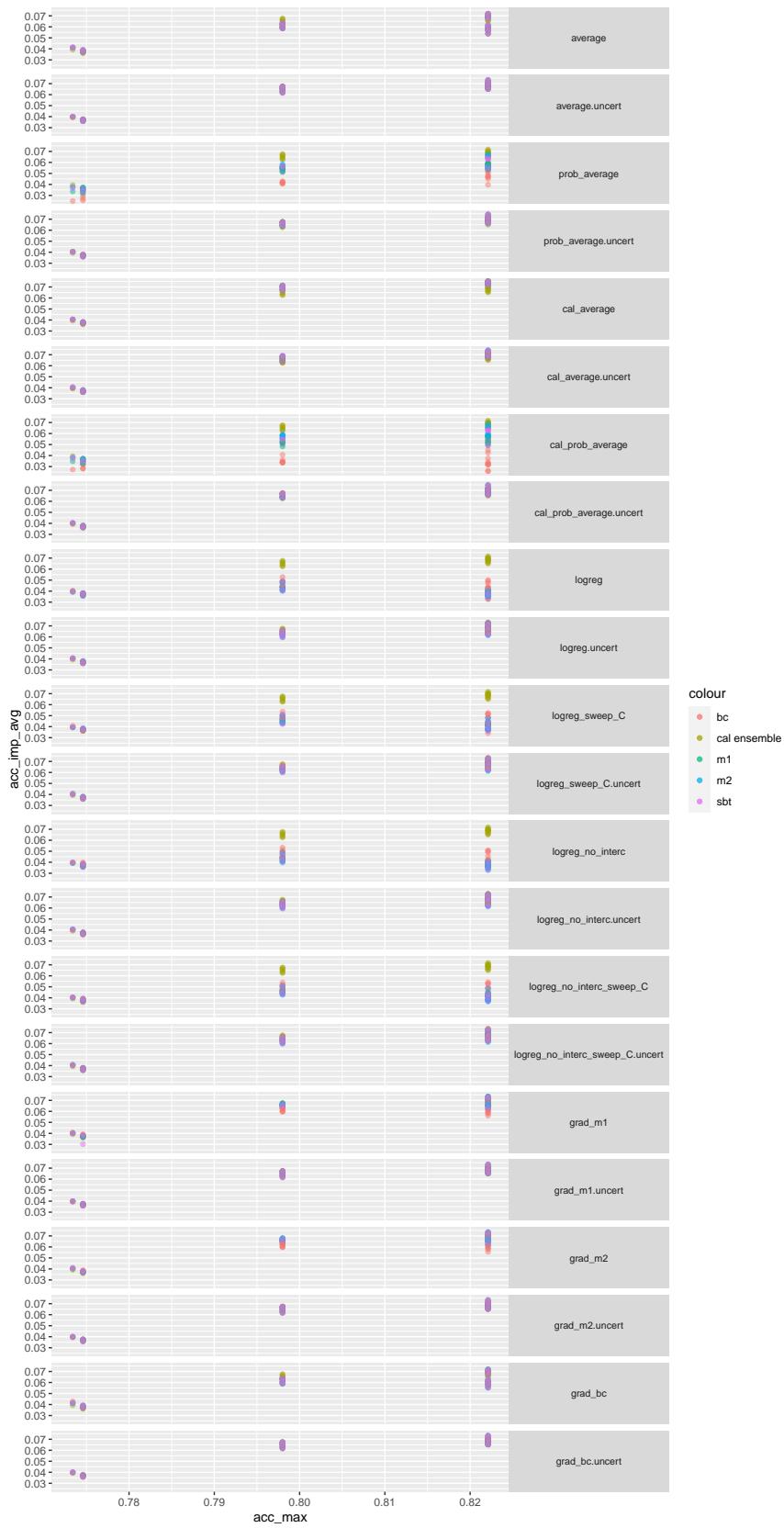
Accuracy improvement of ensemble over
the average of networks vs acc_min.
Ensemble size 3



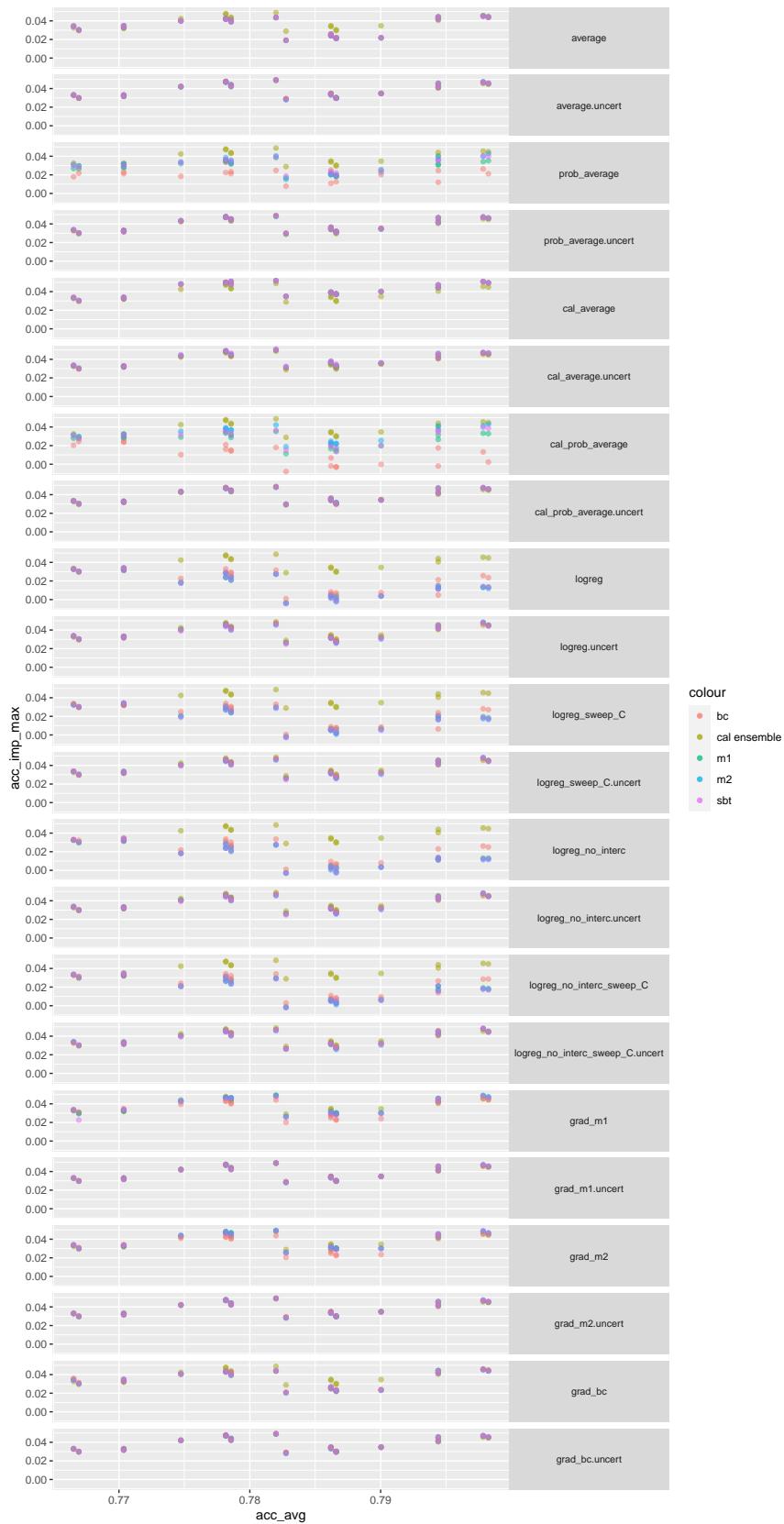
Accuracy improvement of ensemble over
the best of networks vs acc_max.
Ensemble size 3



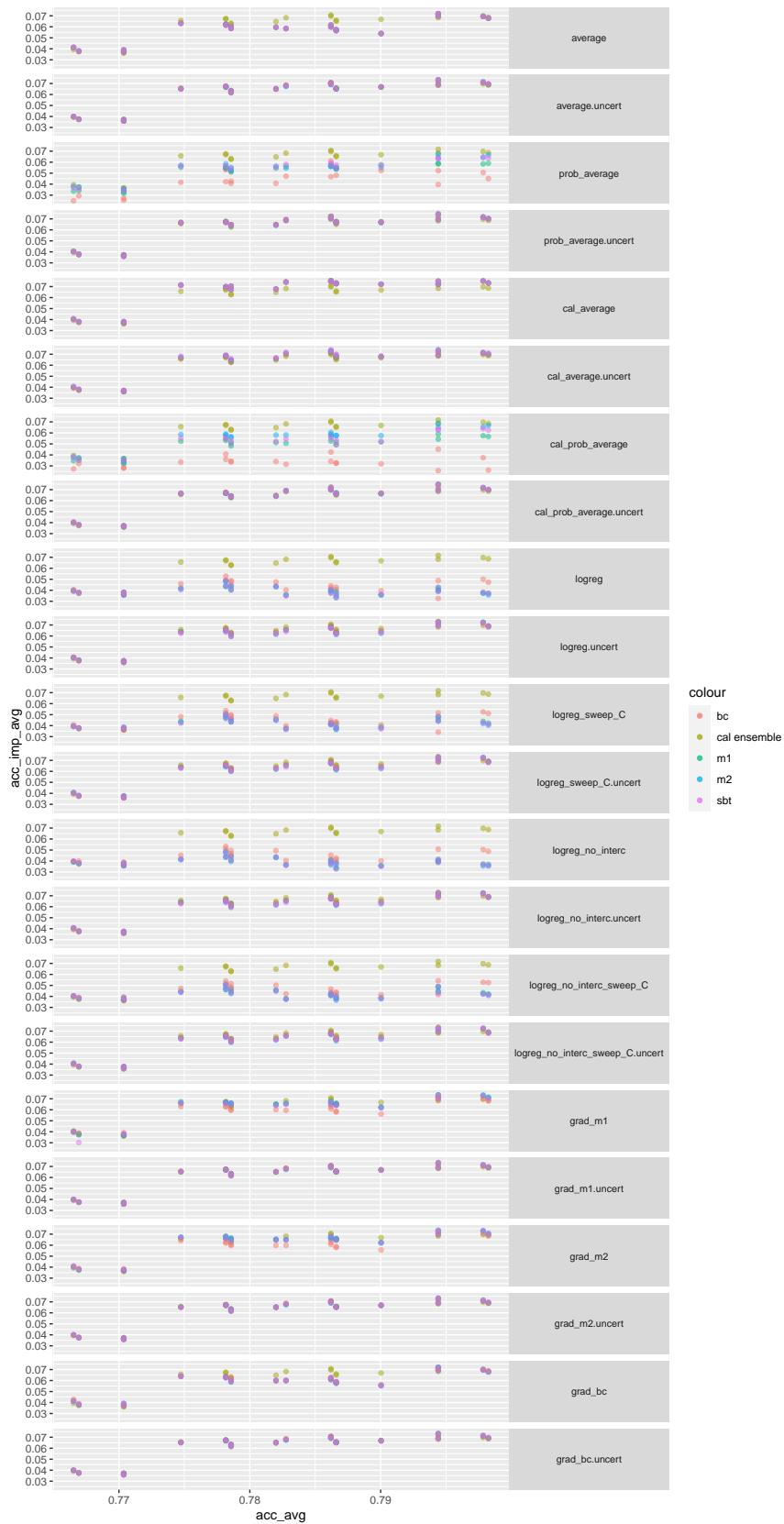
Accuracy improvement of ensemble over
the average of networks vs acc_max.
Ensemble size 3



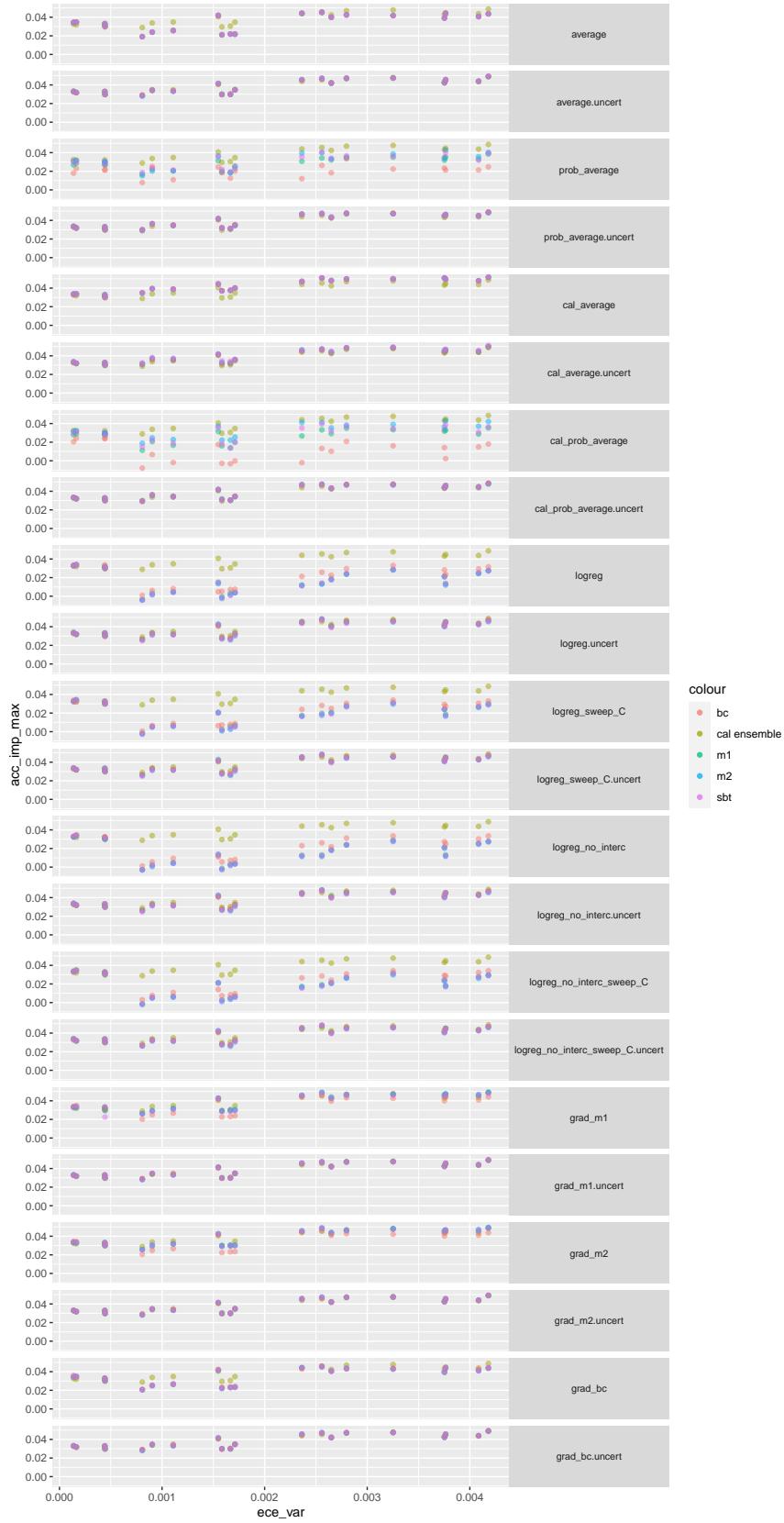
Accuracy improvement of ensemble over
the best of networks vs acc_avg.
Ensemble size 3



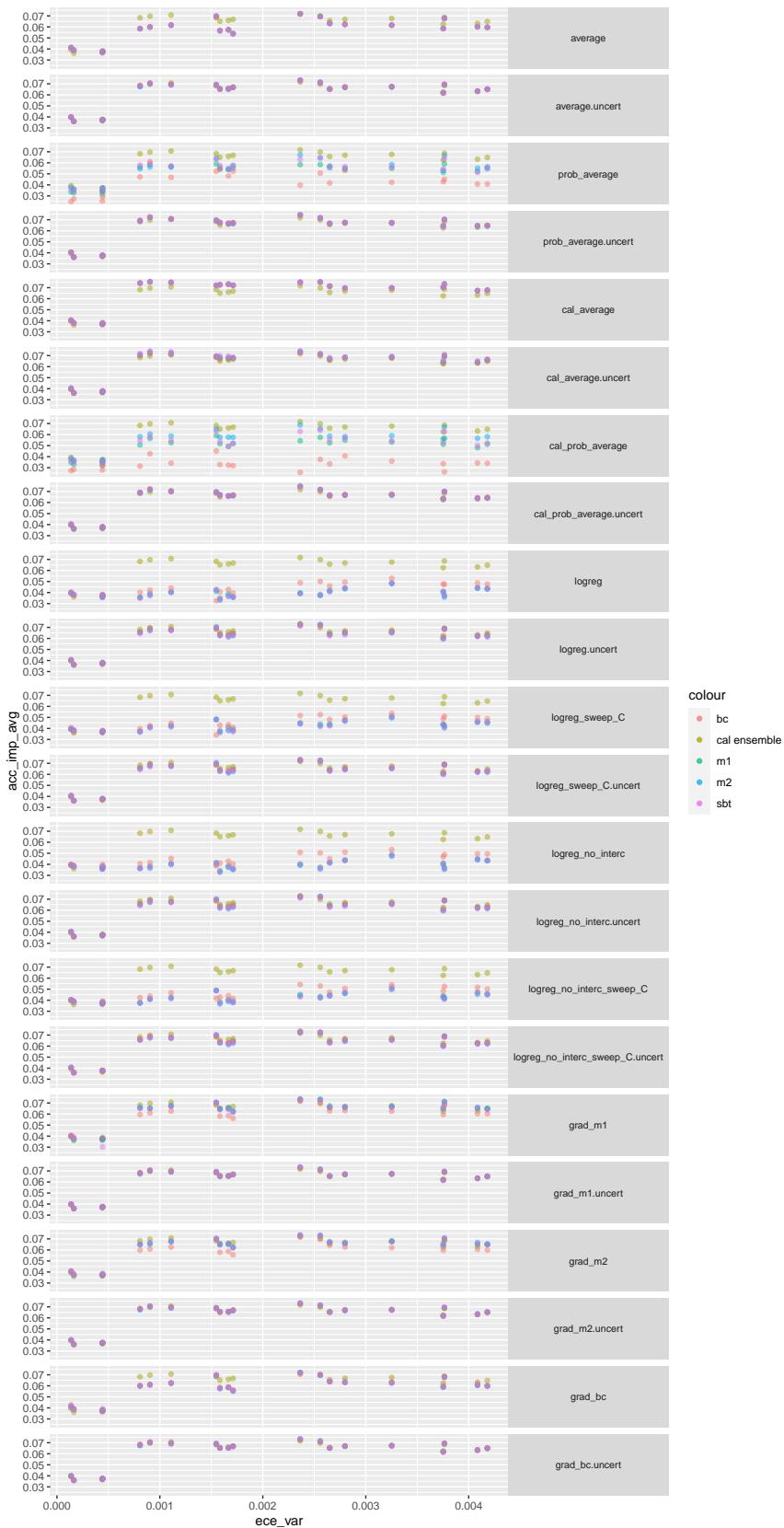
Accuracy improvement of ensemble over
the average of networks vs acc_avg.
Ensemble size 3



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



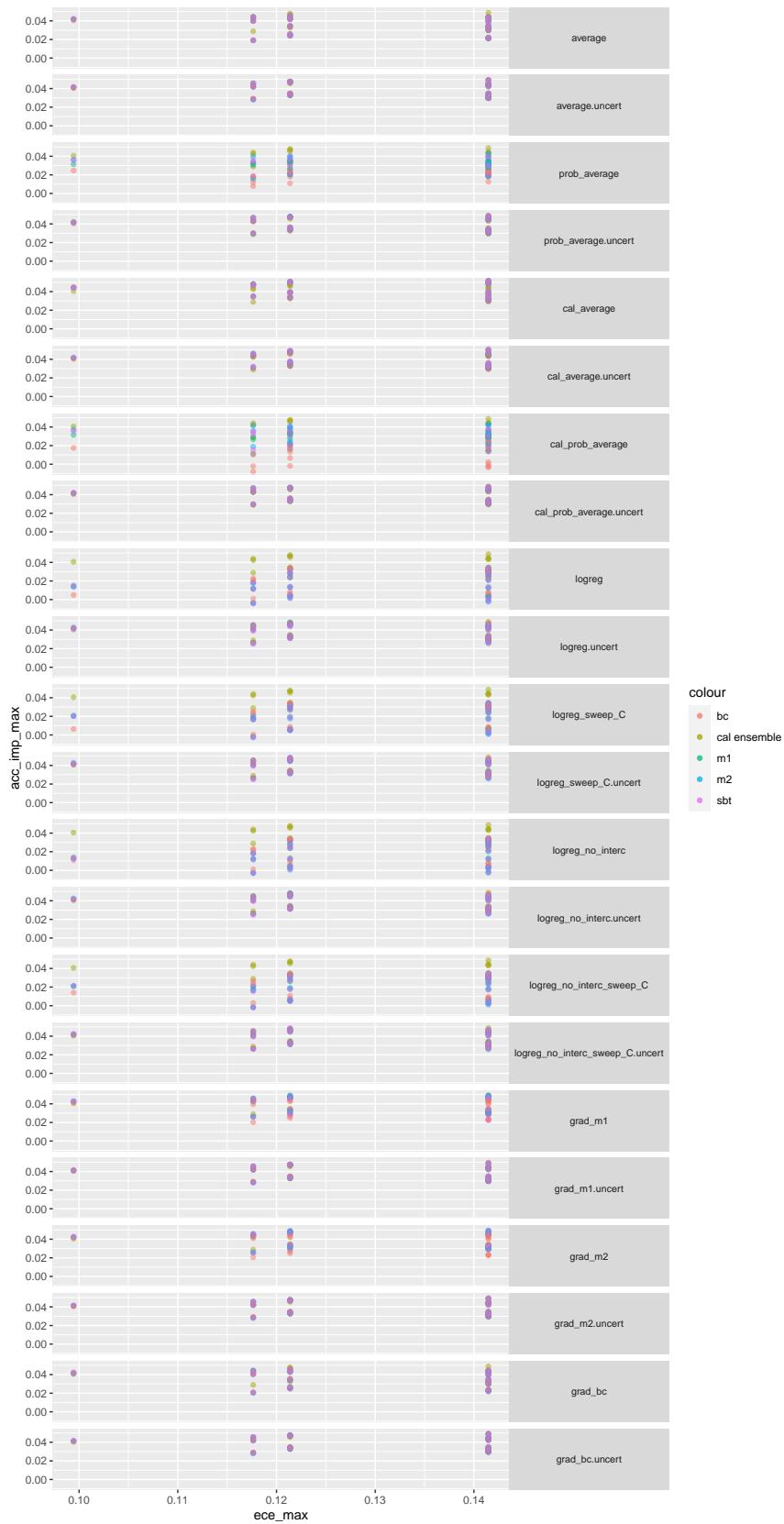
Accuracy improvement of ensemble over
the best of networks vs ece_min.
Ensemble size 3



Accuracy improvement of ensemble over
the average of networks vs ece_min.
Ensemble size 3



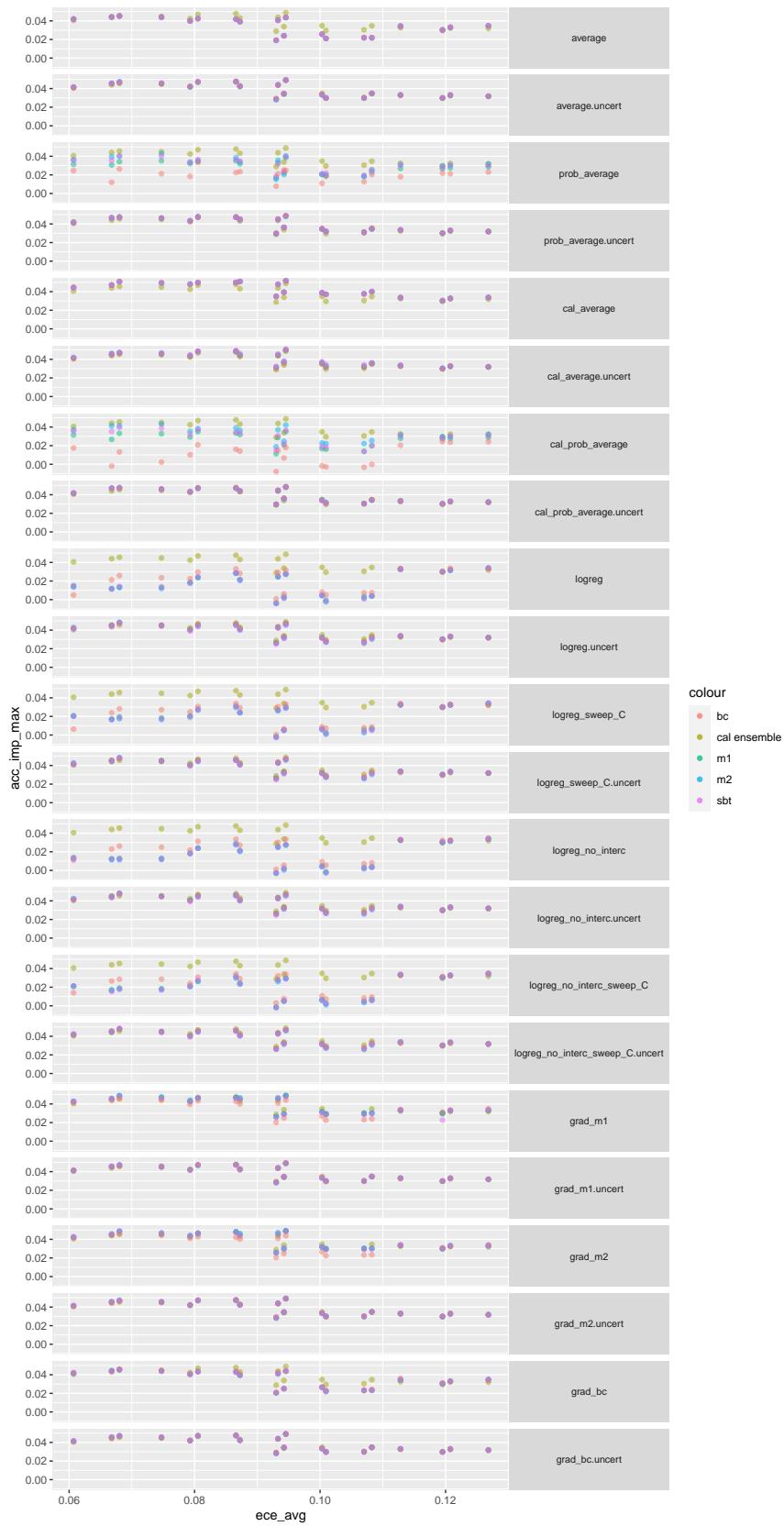
Accuracy improvement of ensemble over
the best of networks vs ece_max.
Ensemble size 3



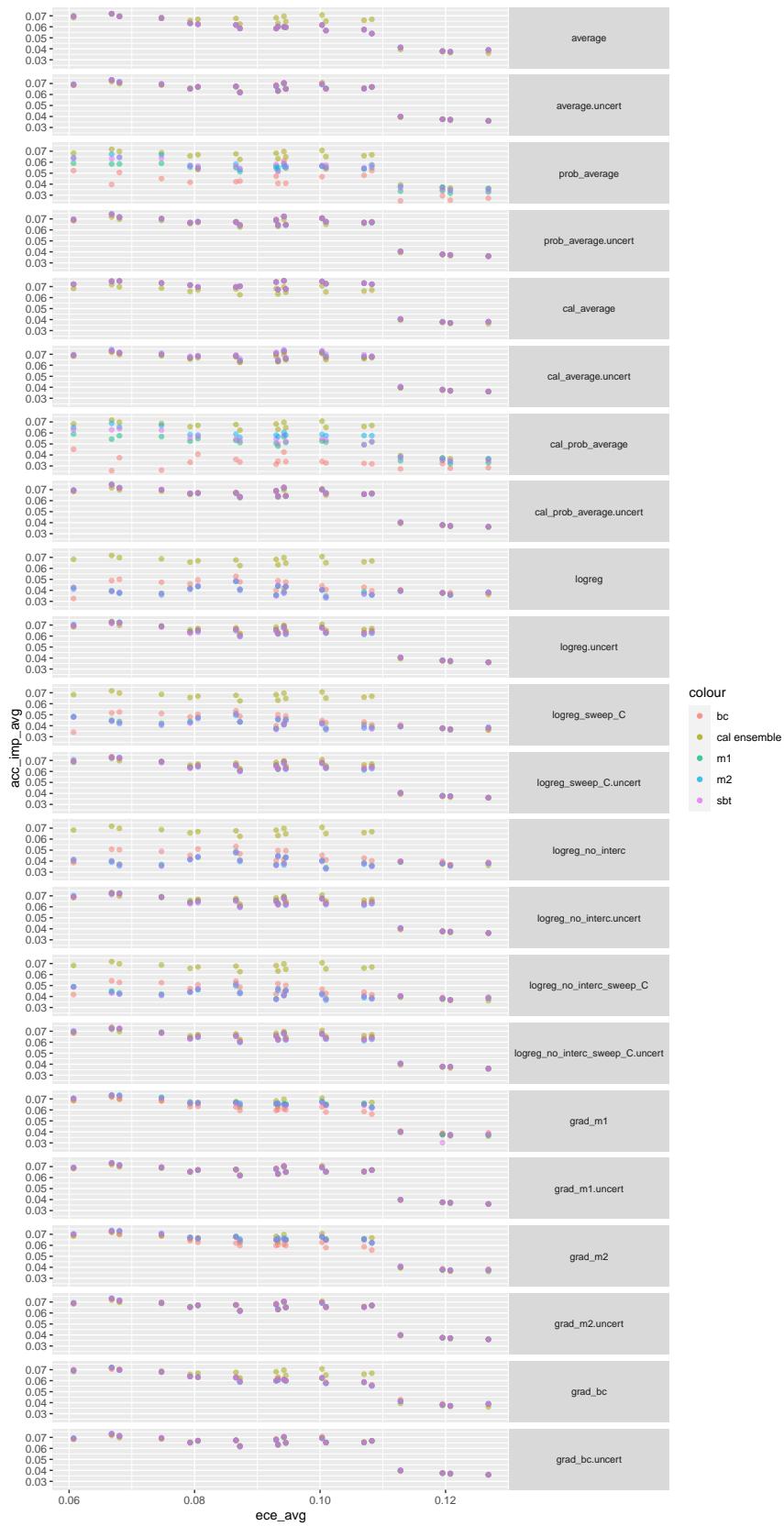
Accuracy improvement of ensemble over
the average of networks vs ece_max.
Ensemble size 3



Accuracy improvement of ensemble over
the best of networks vs ece_avg.
Ensemble size 3



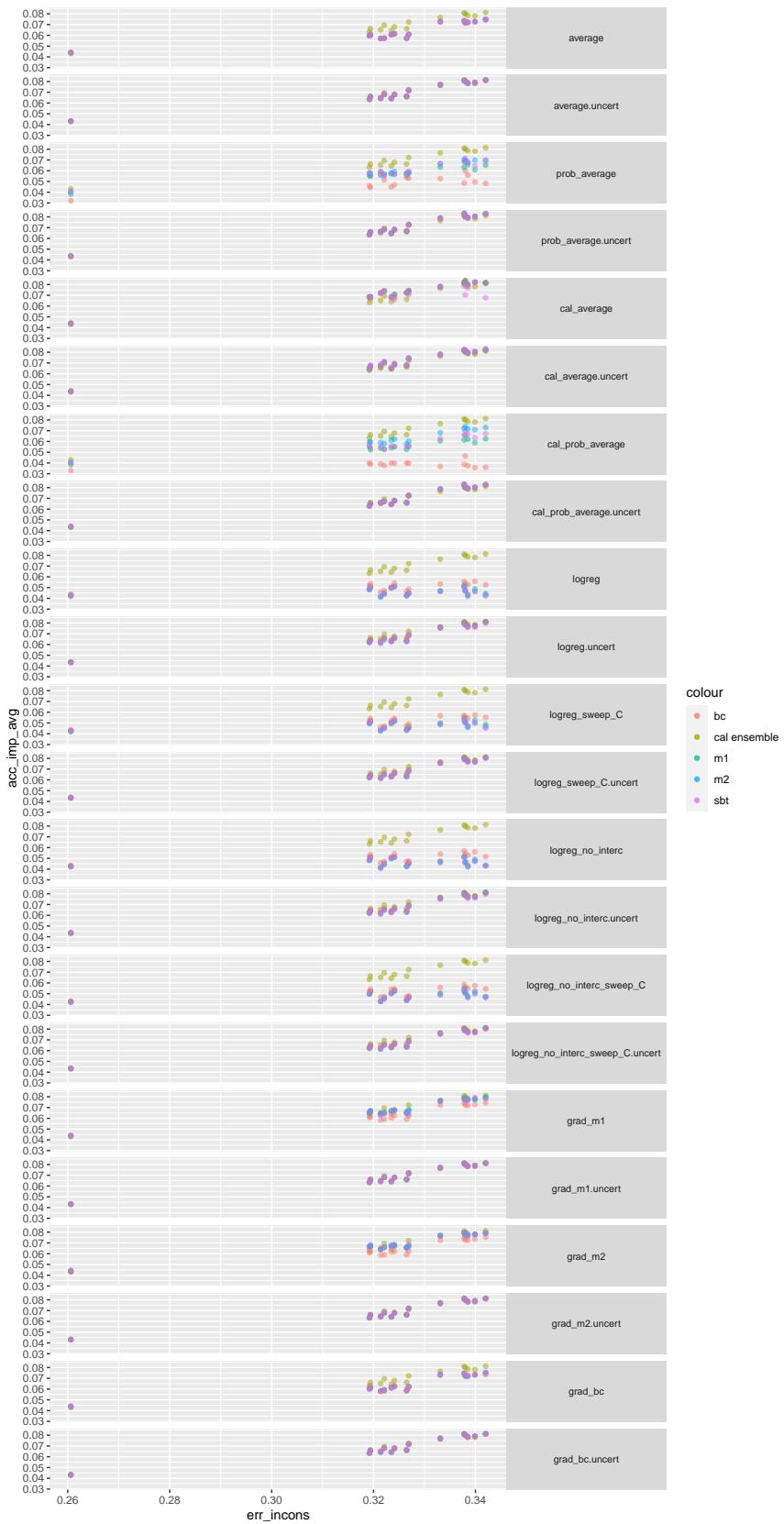
Accuracy improvement of ensemble over
the average of networks vs ece_avg.
Ensemble size 3



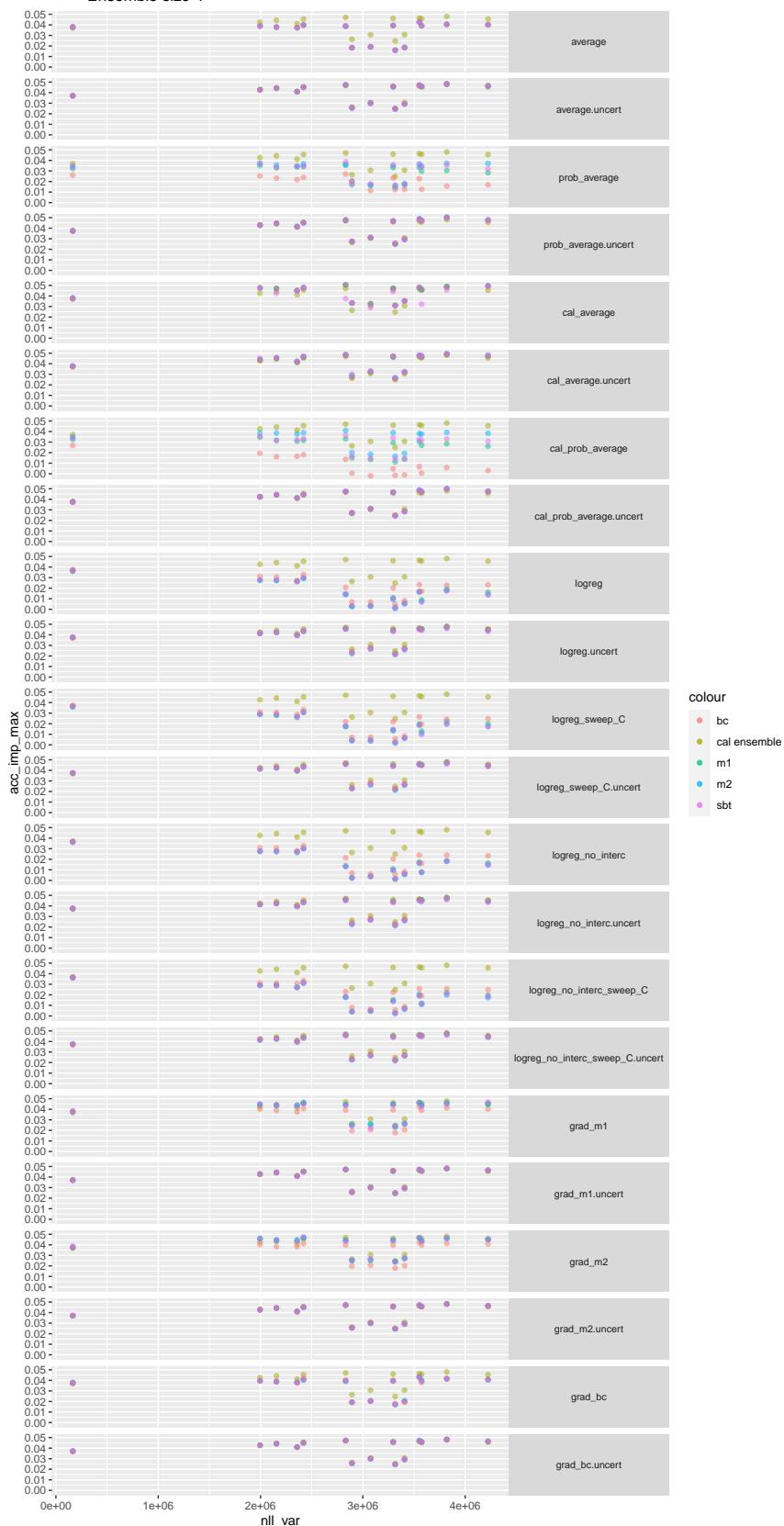
Accuracy improvement of ensemble over
the best of networks vs err_incons.
Ensemble size 4



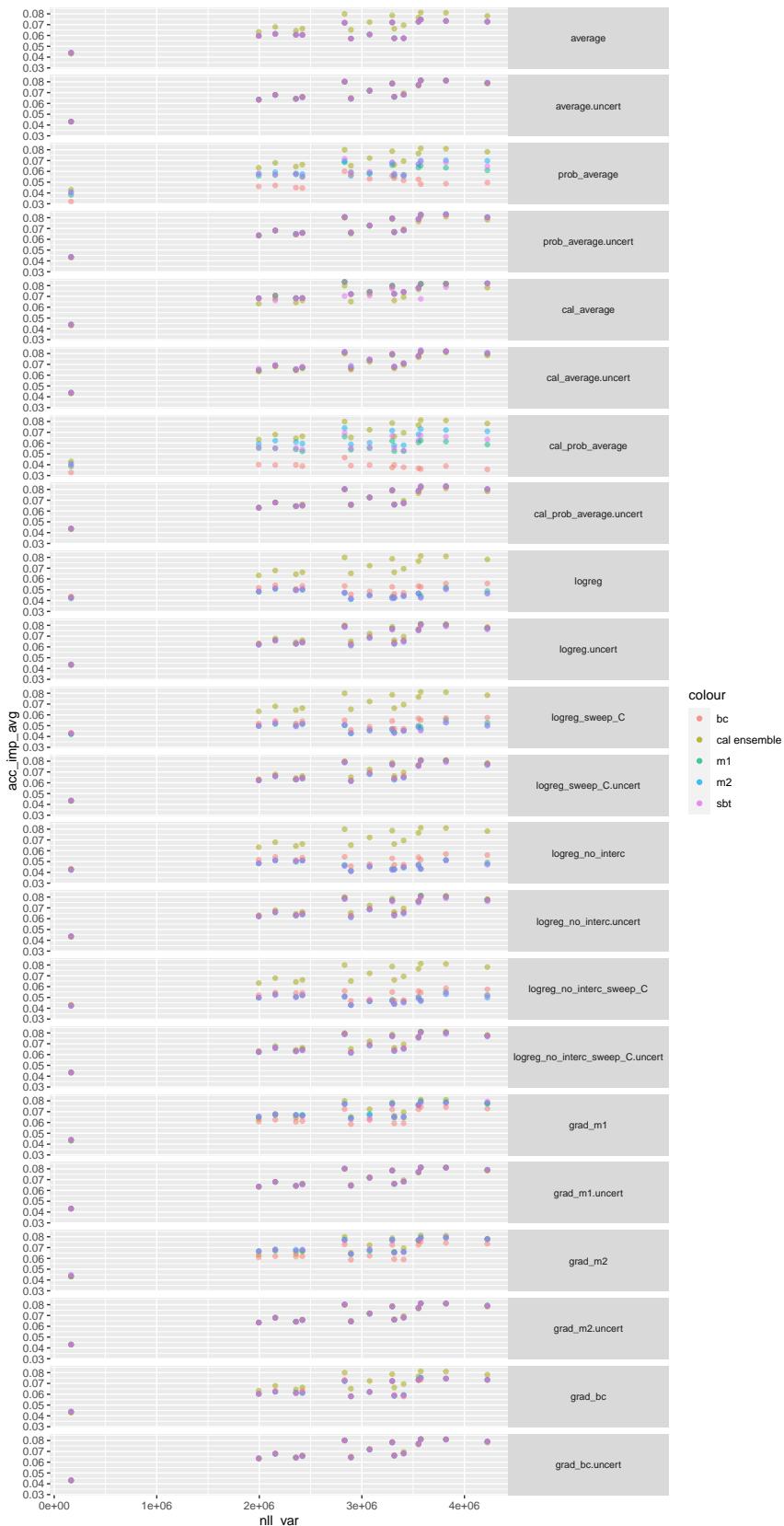
Accuracy improvement of ensemble over
the average of networks vs err_incons.
Ensemble size 4



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



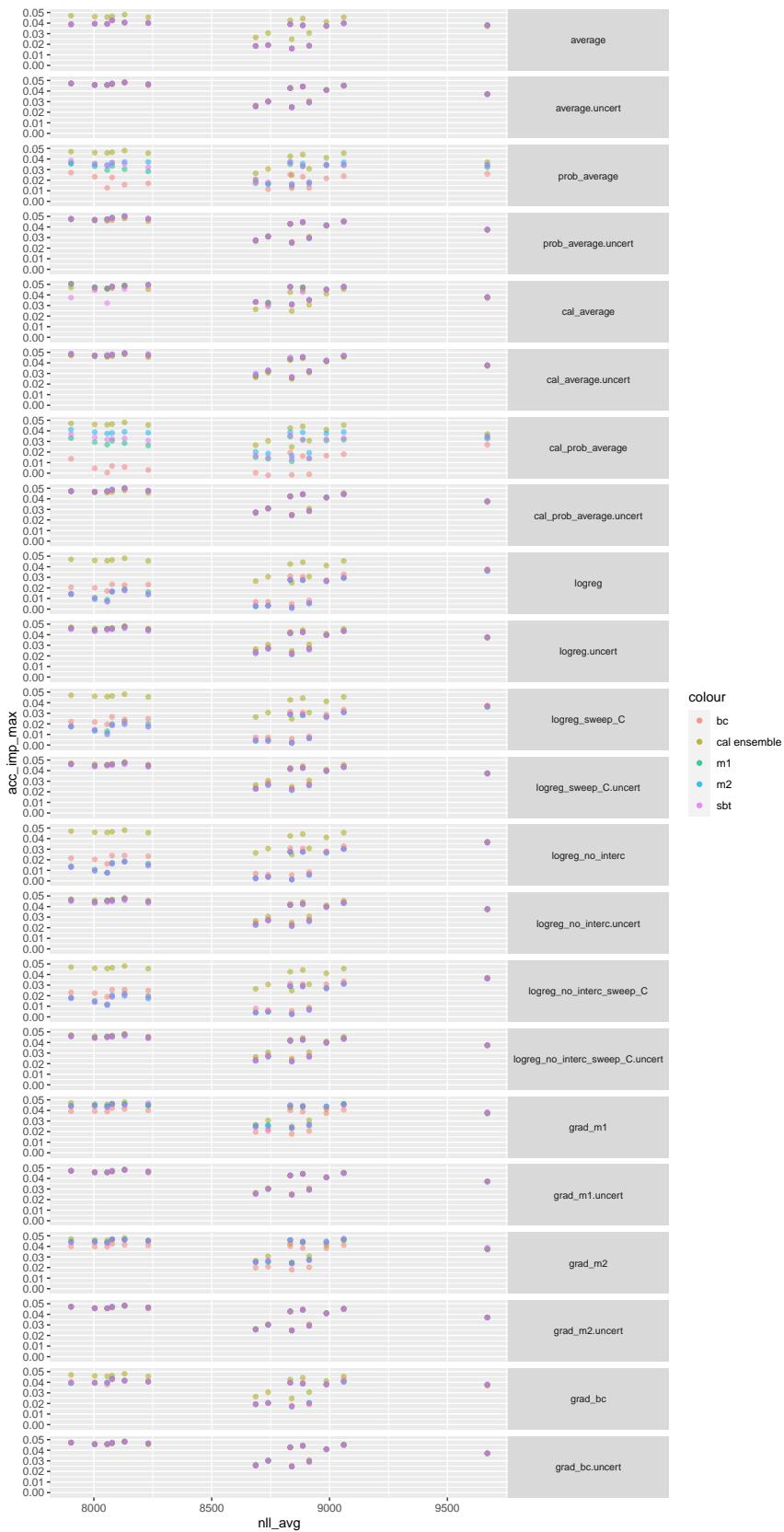
Accuracy improvement of ensemble over
the best of networks vs nll_max.
Ensemble size 4



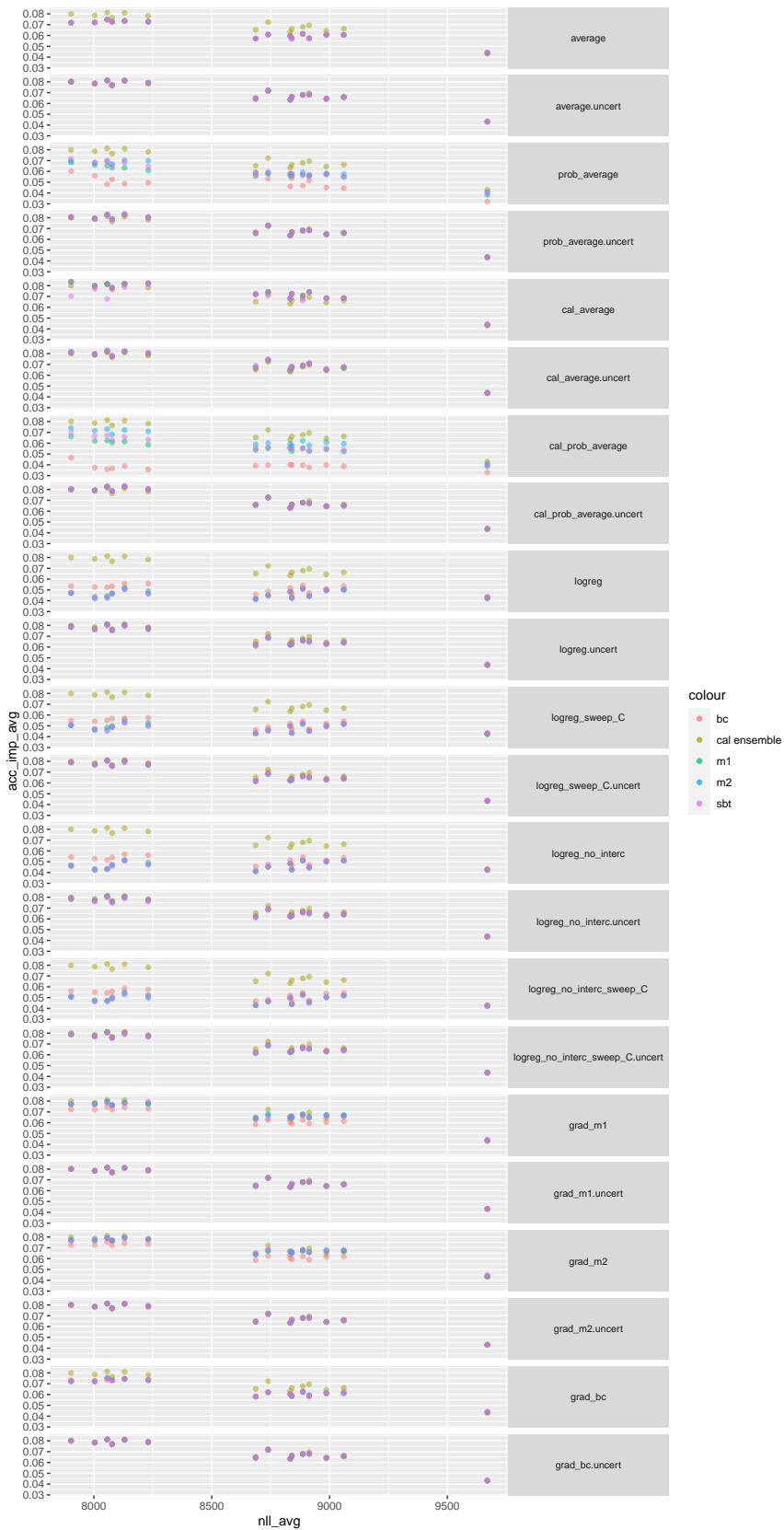
Accuracy improvement of ensemble over
the average of networks vs nll_max.
Ensemble size 4



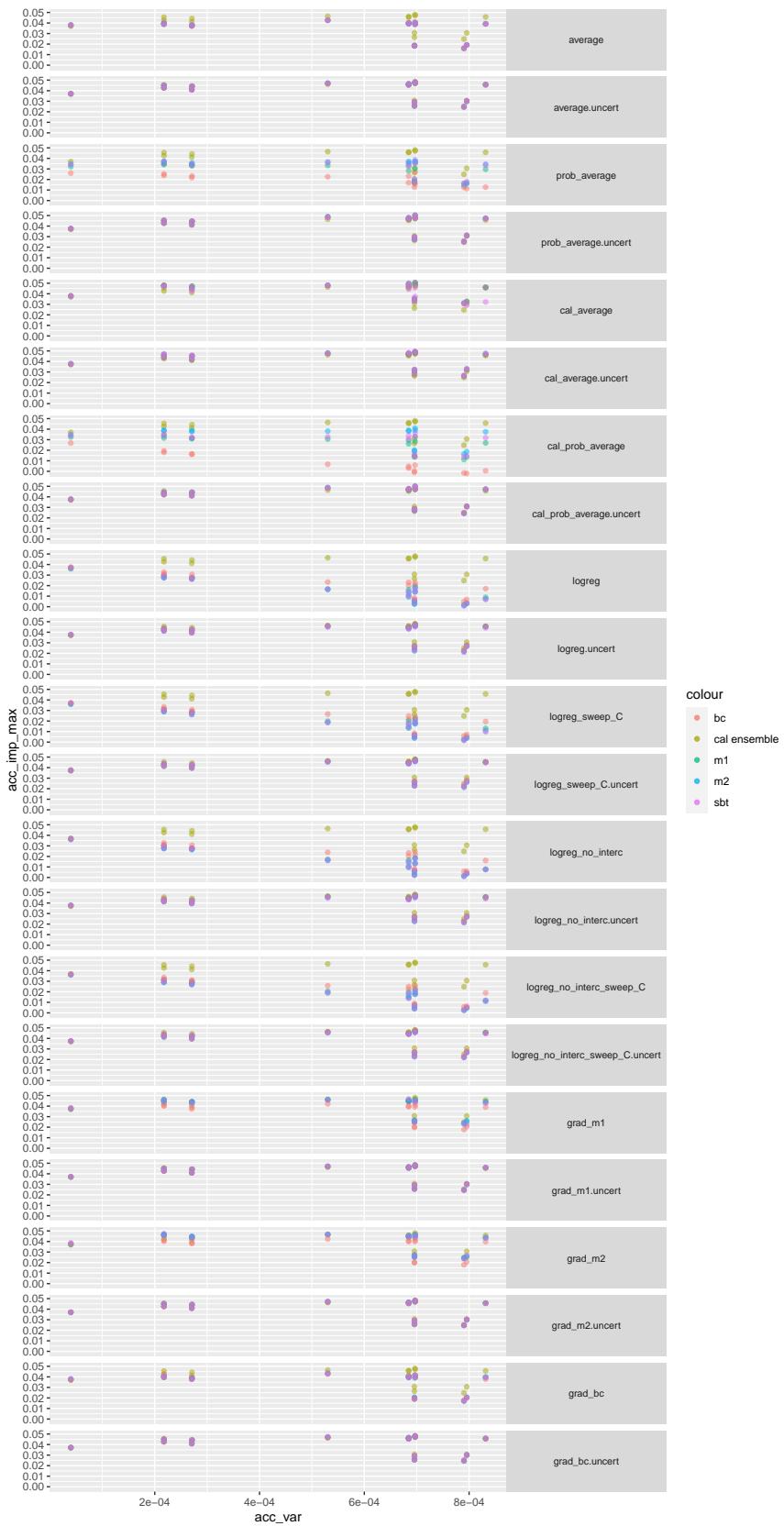
Accuracy improvement of ensemble over
the best of networks vs nll_avg.
Ensemble size 4



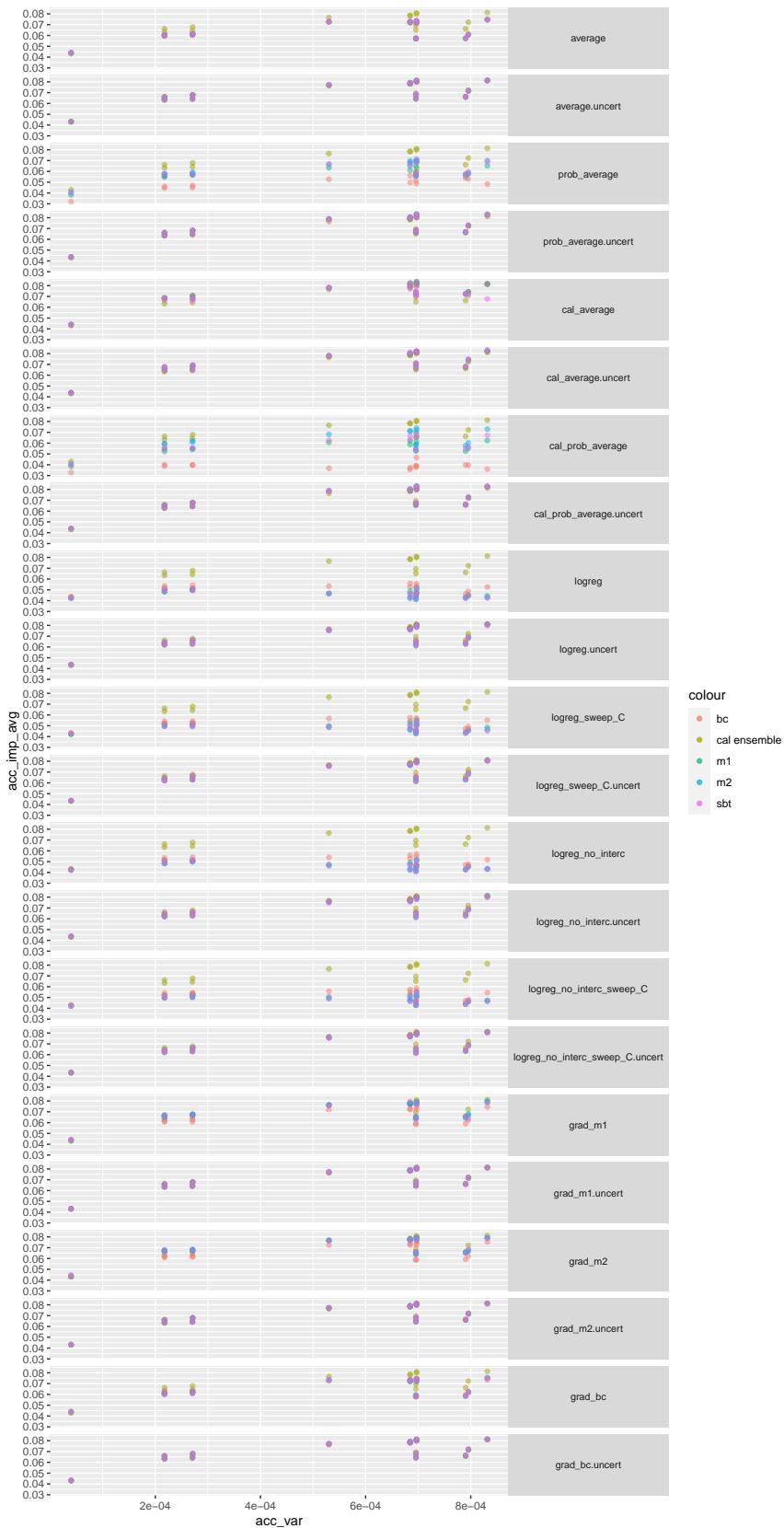
Accuracy improvement of ensemble over
the average of networks vs nll_avg.
Ensemble size 4



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



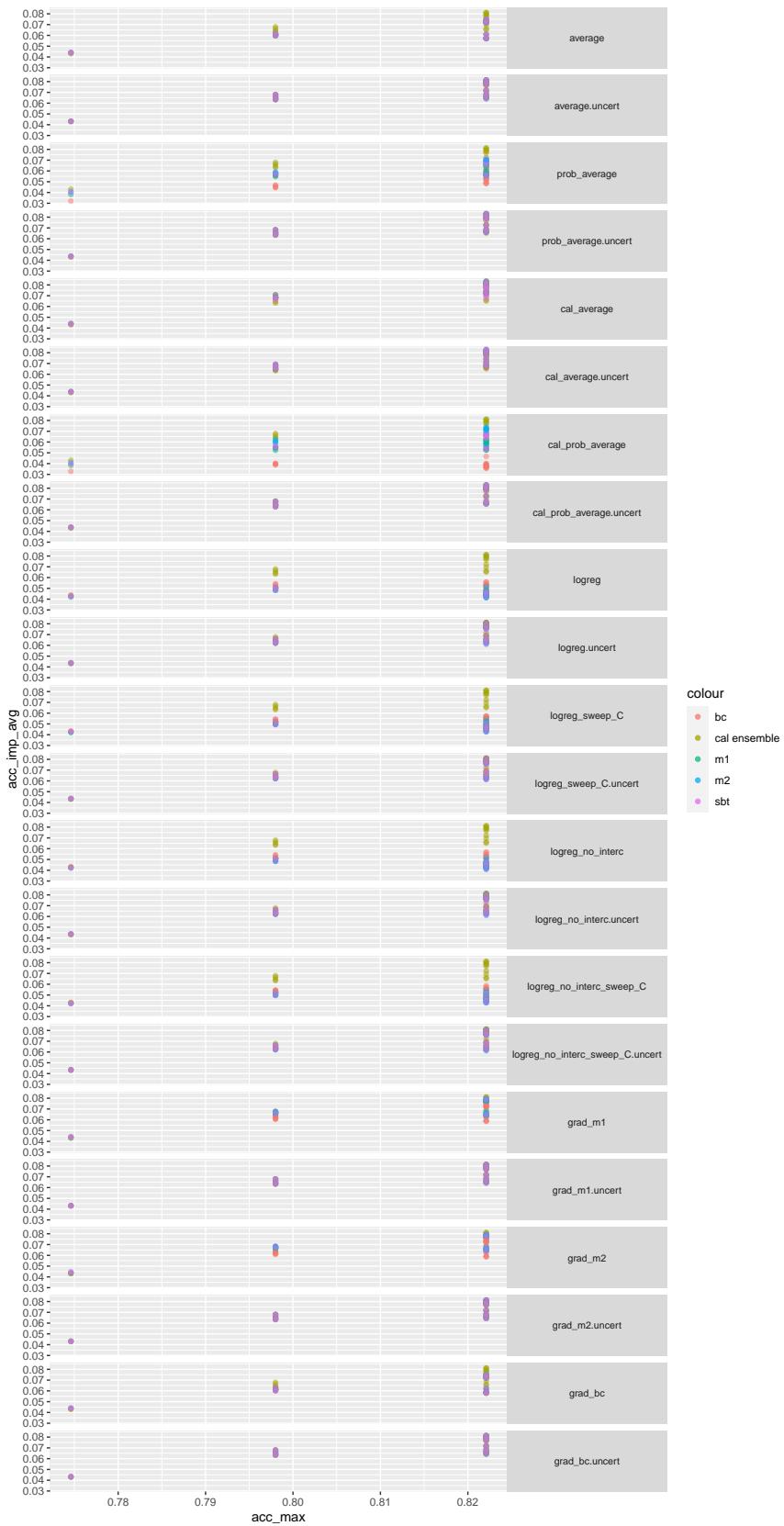
Accuracy improvement of ensemble over
the average of networks vs acc_min.
Ensemble size 4



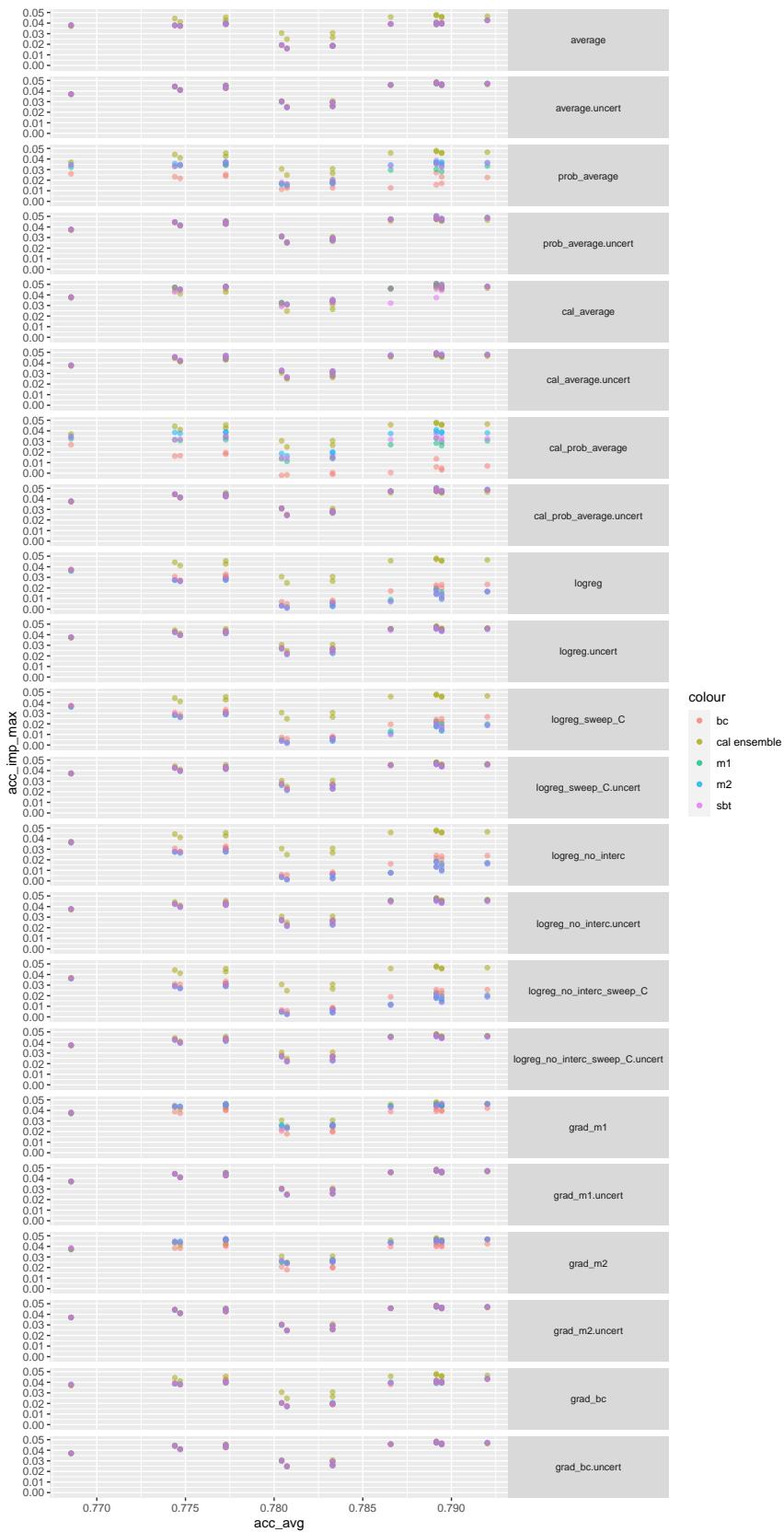
Accuracy improvement of ensemble over
the best of networks vs acc_max.
Ensemble size 4



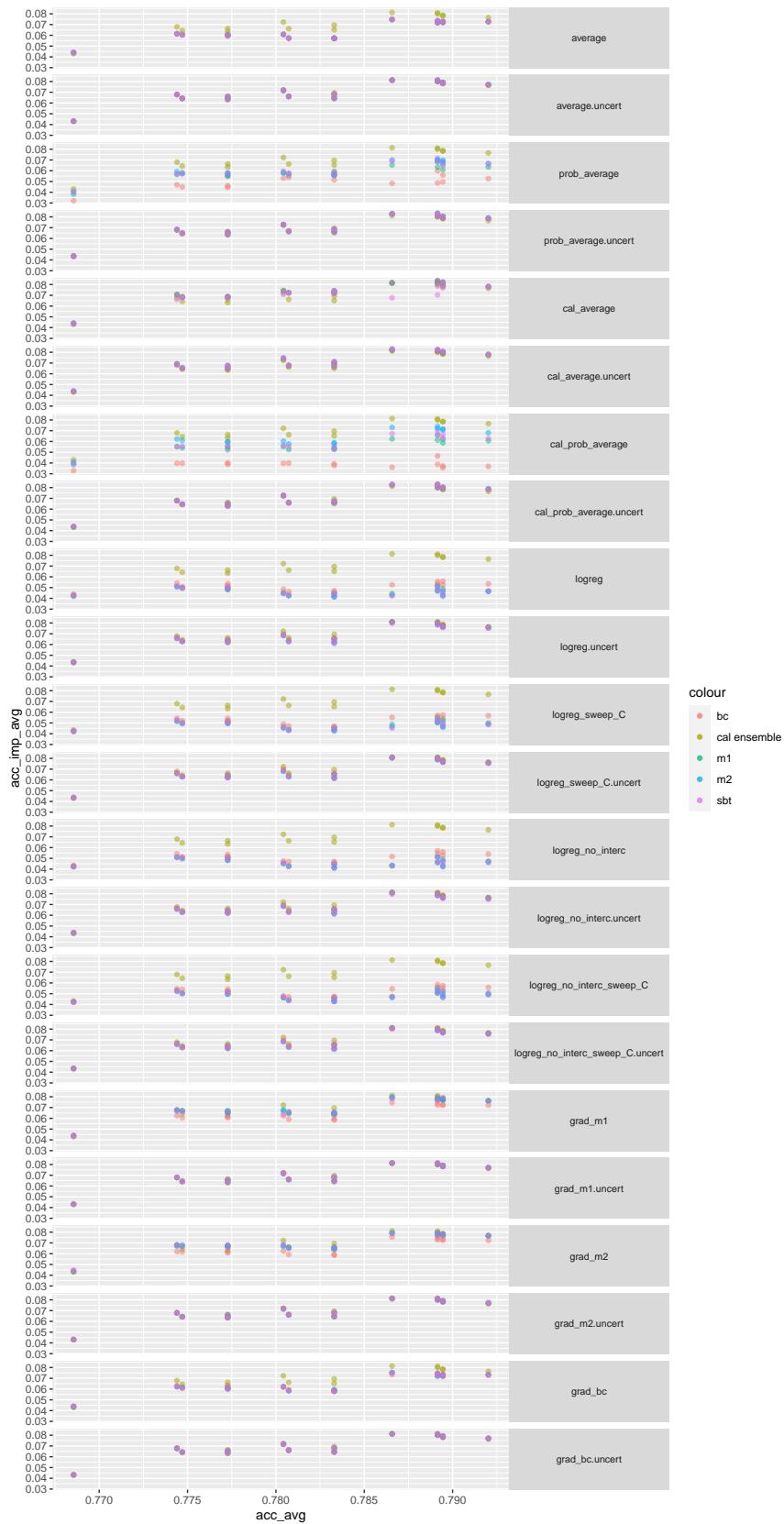
Accuracy improvement of ensemble over
the average of networks vs acc_max.
Ensemble size 4



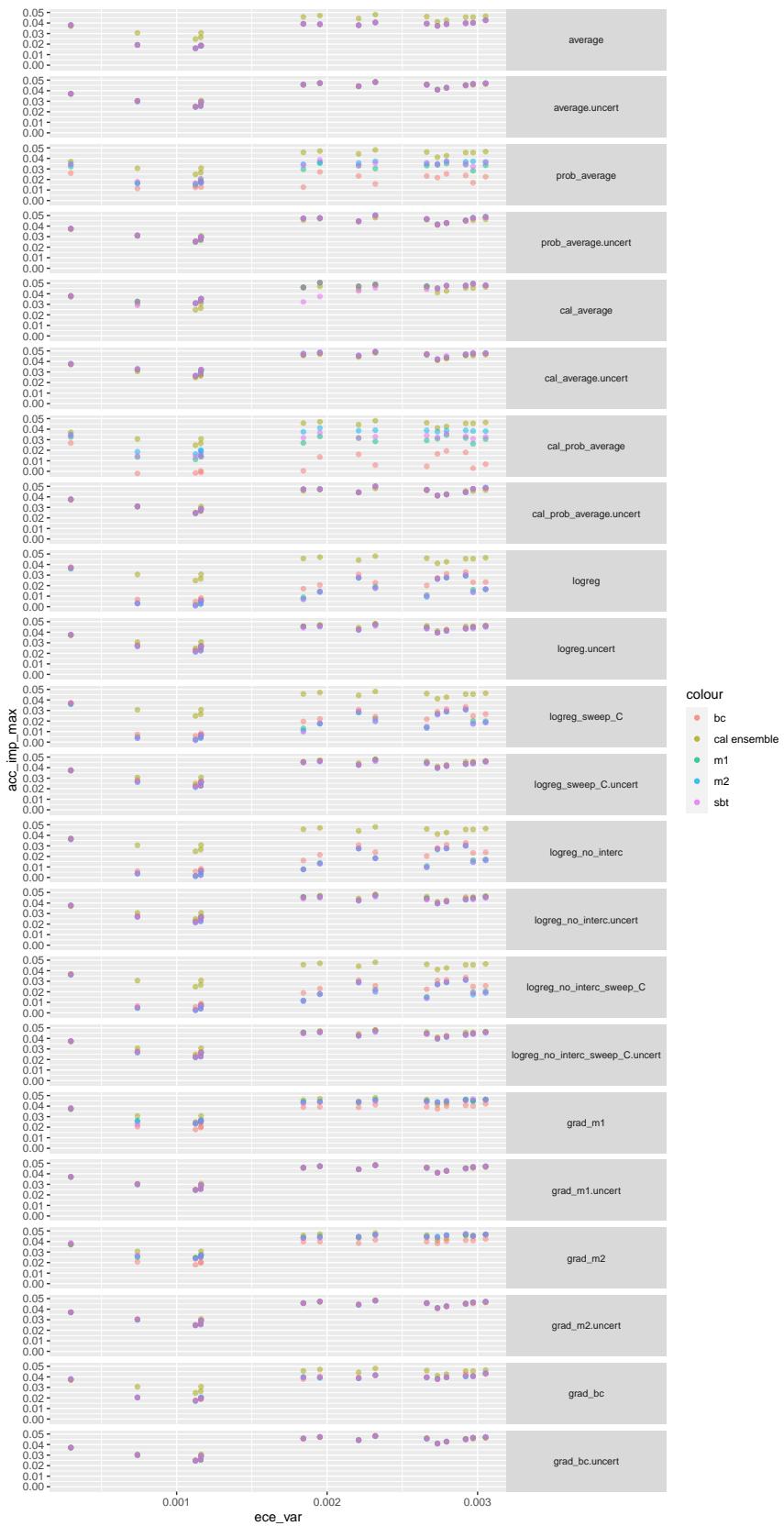
Accuracy improvement of ensemble over
the best of networks vs acc_avg.
Ensemble size 4



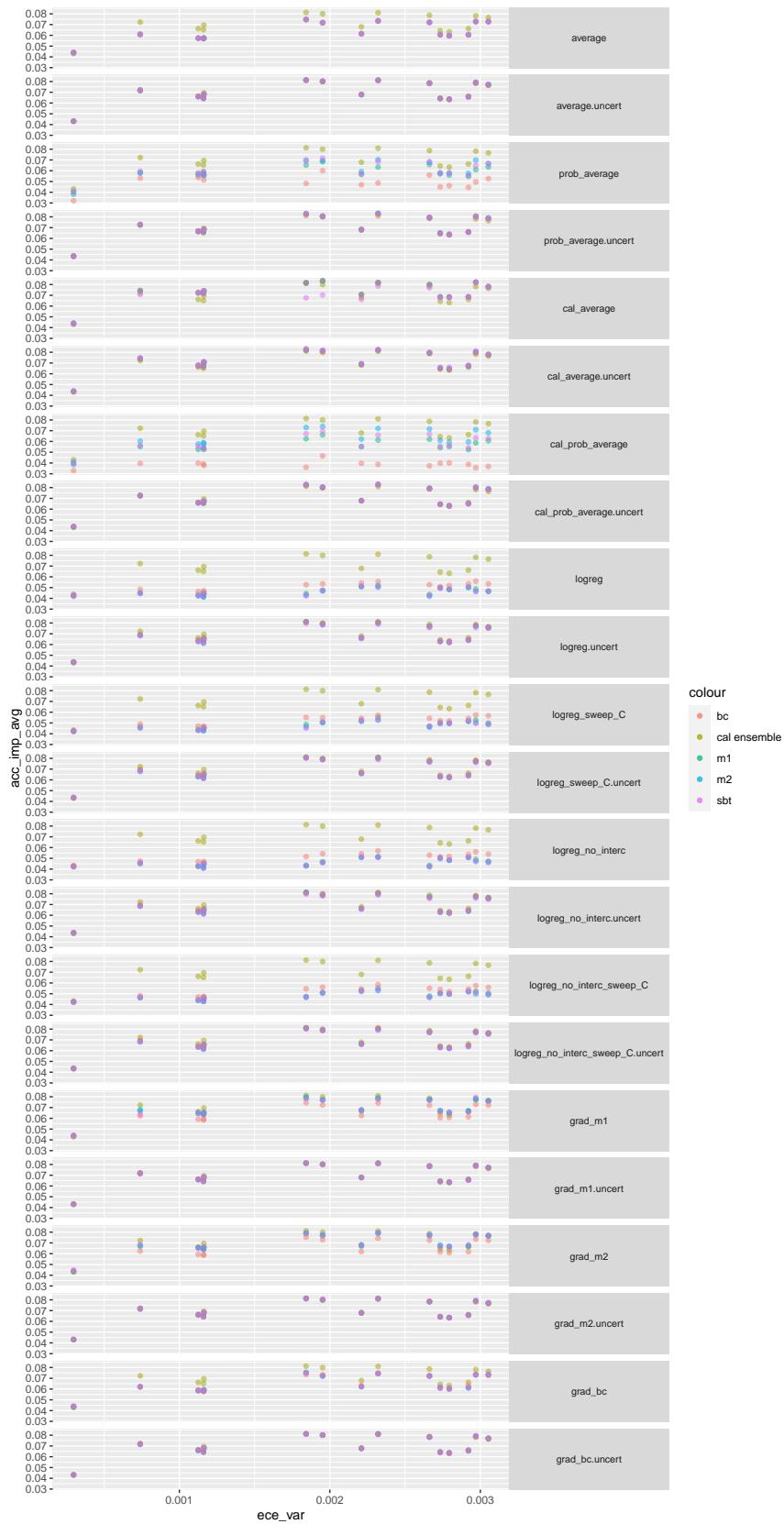
Accuracy improvement of ensemble over
the average of networks vs acc_avg.
Ensemble size 4



Accuracy improvement of ensemble over
the best of networks vs ece_var.
Ensemble size 4



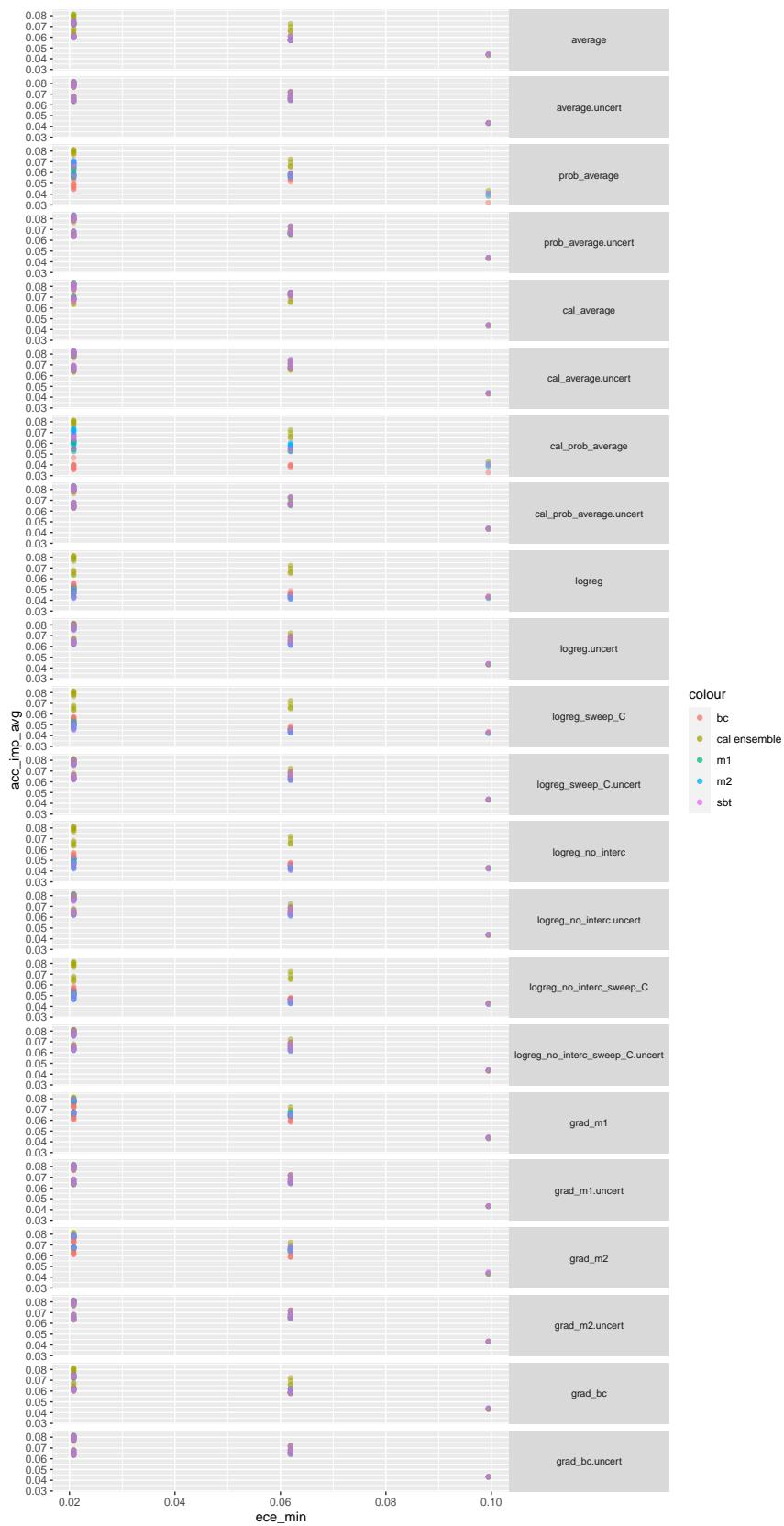
Accuracy improvement of ensemble over
the average of networks vs ece_var.
Ensemble size 4



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



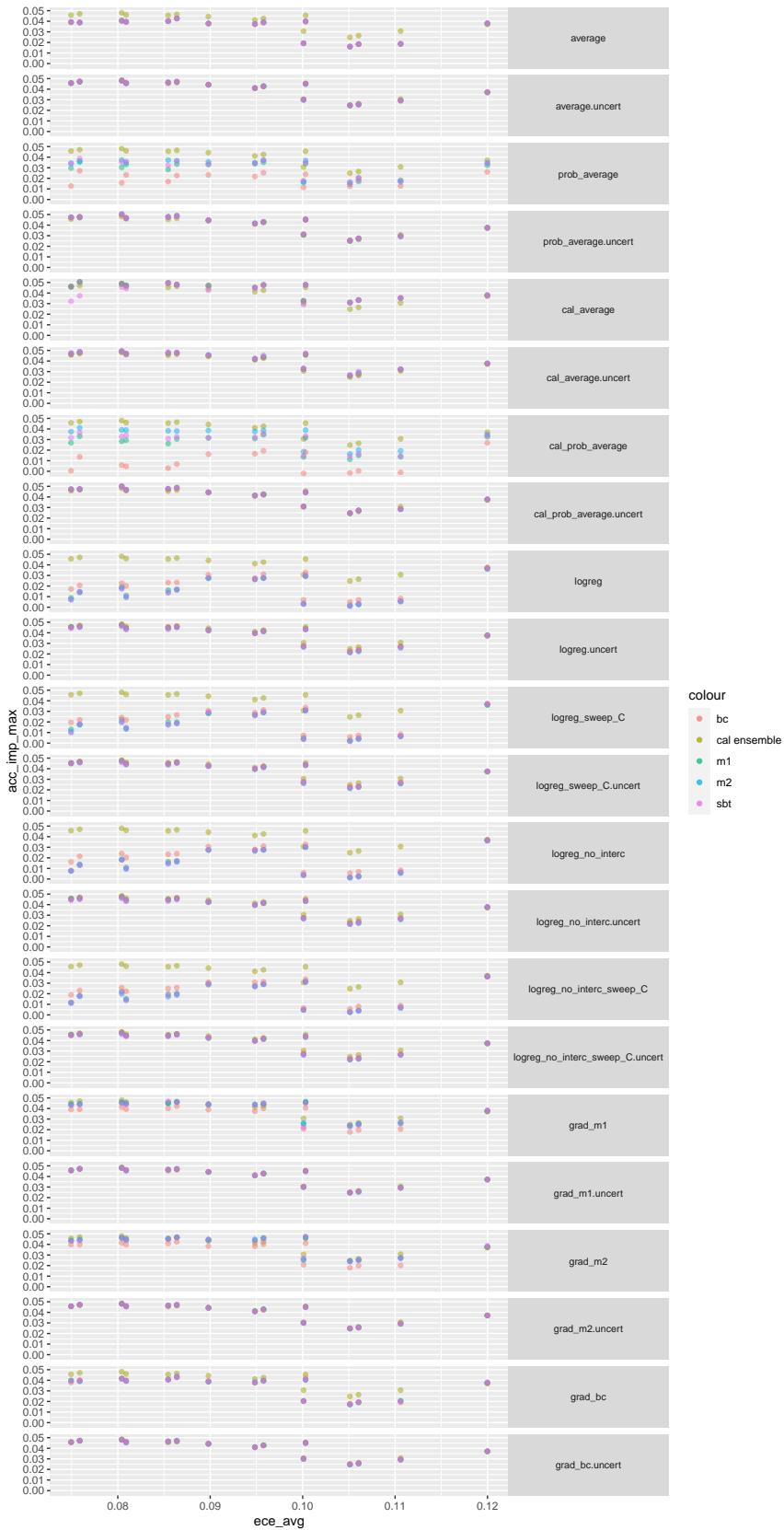
Accuracy improvement of ensemble over
the best of networks vs ece_max.
Ensemble size 4



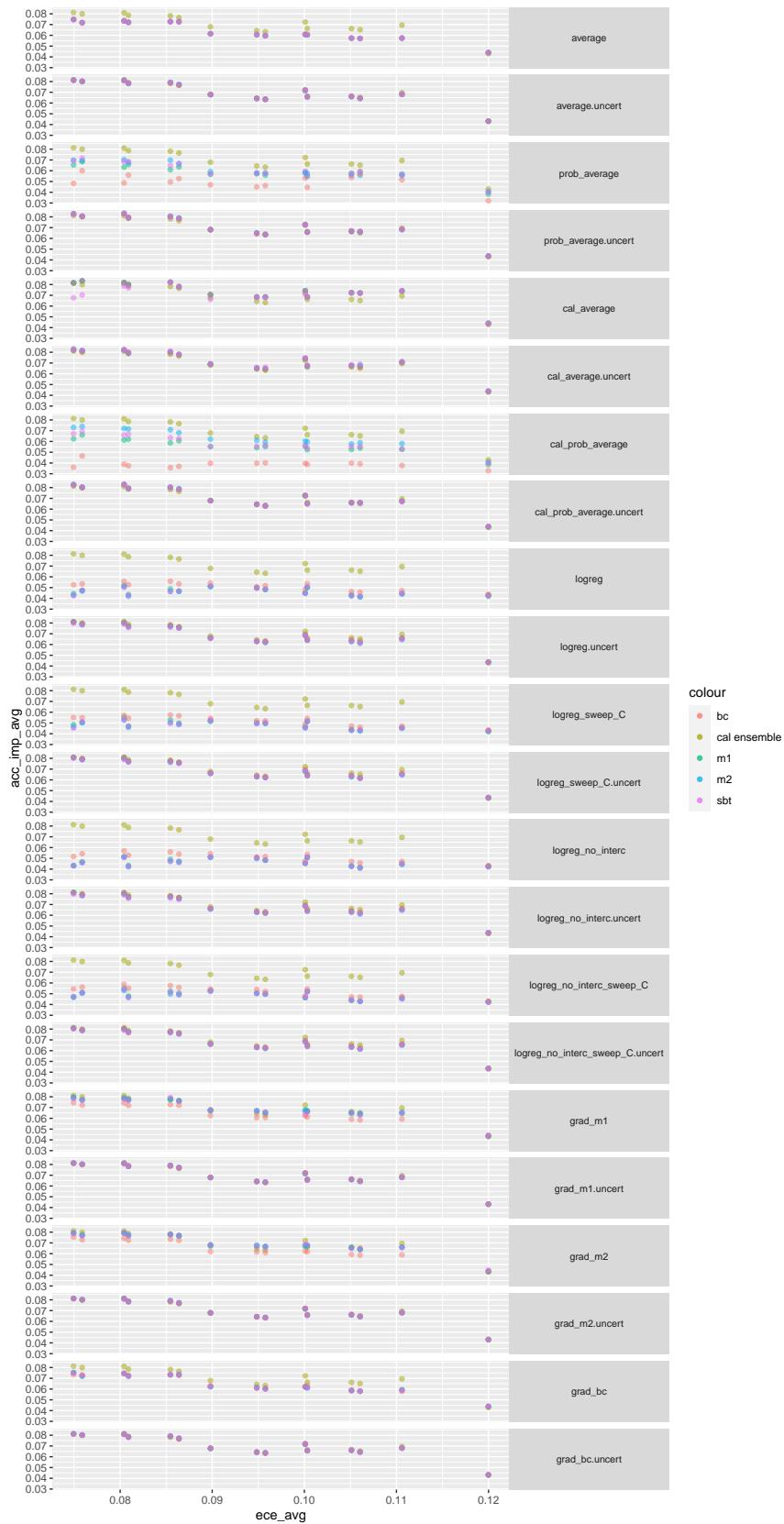
Accuracy improvement of ensemble over
the average of networks vs ece_max.
Ensemble size 4

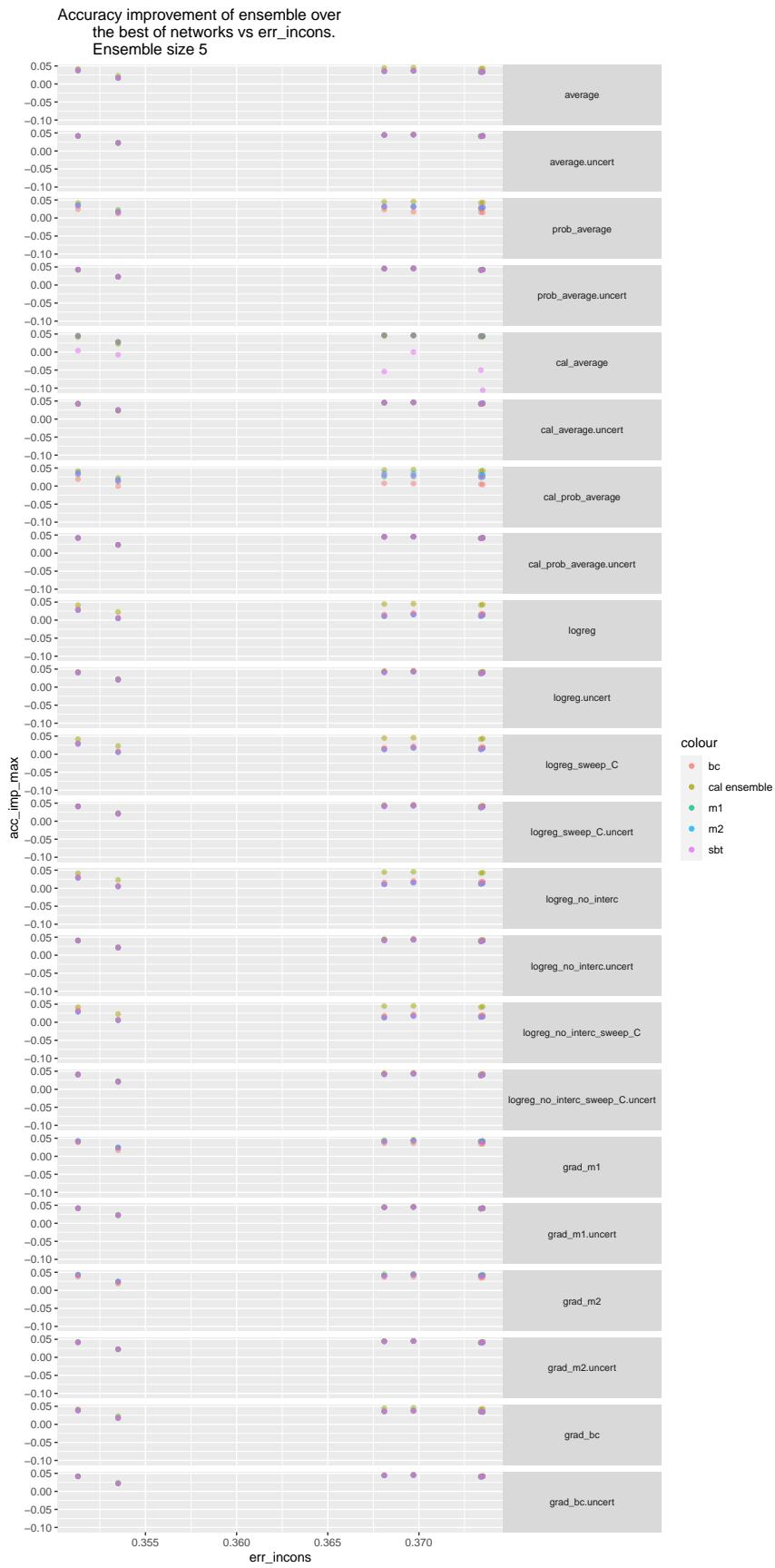


Accuracy improvement of ensemble over
the best of networks vs ece_avg.
Ensemble size 4

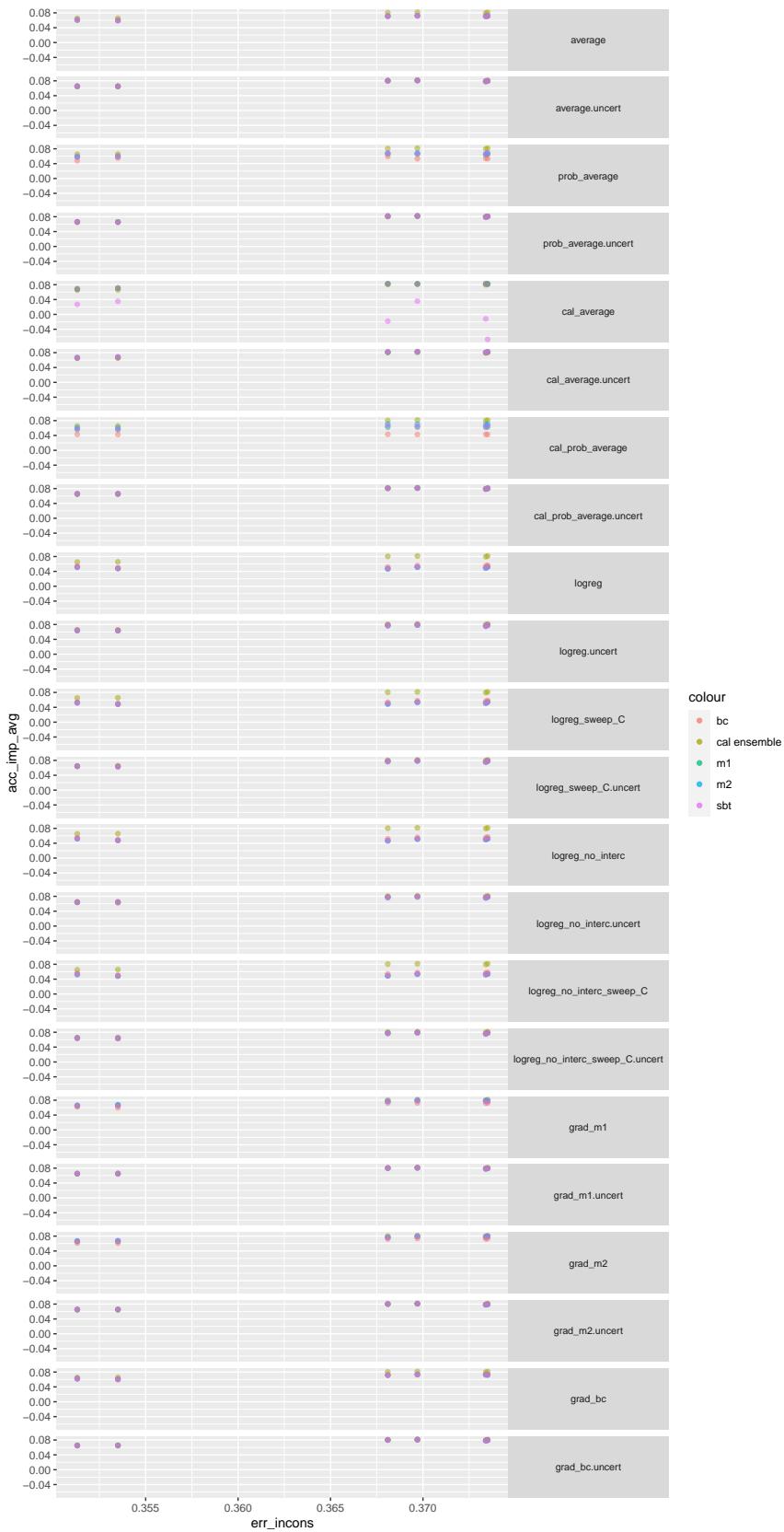


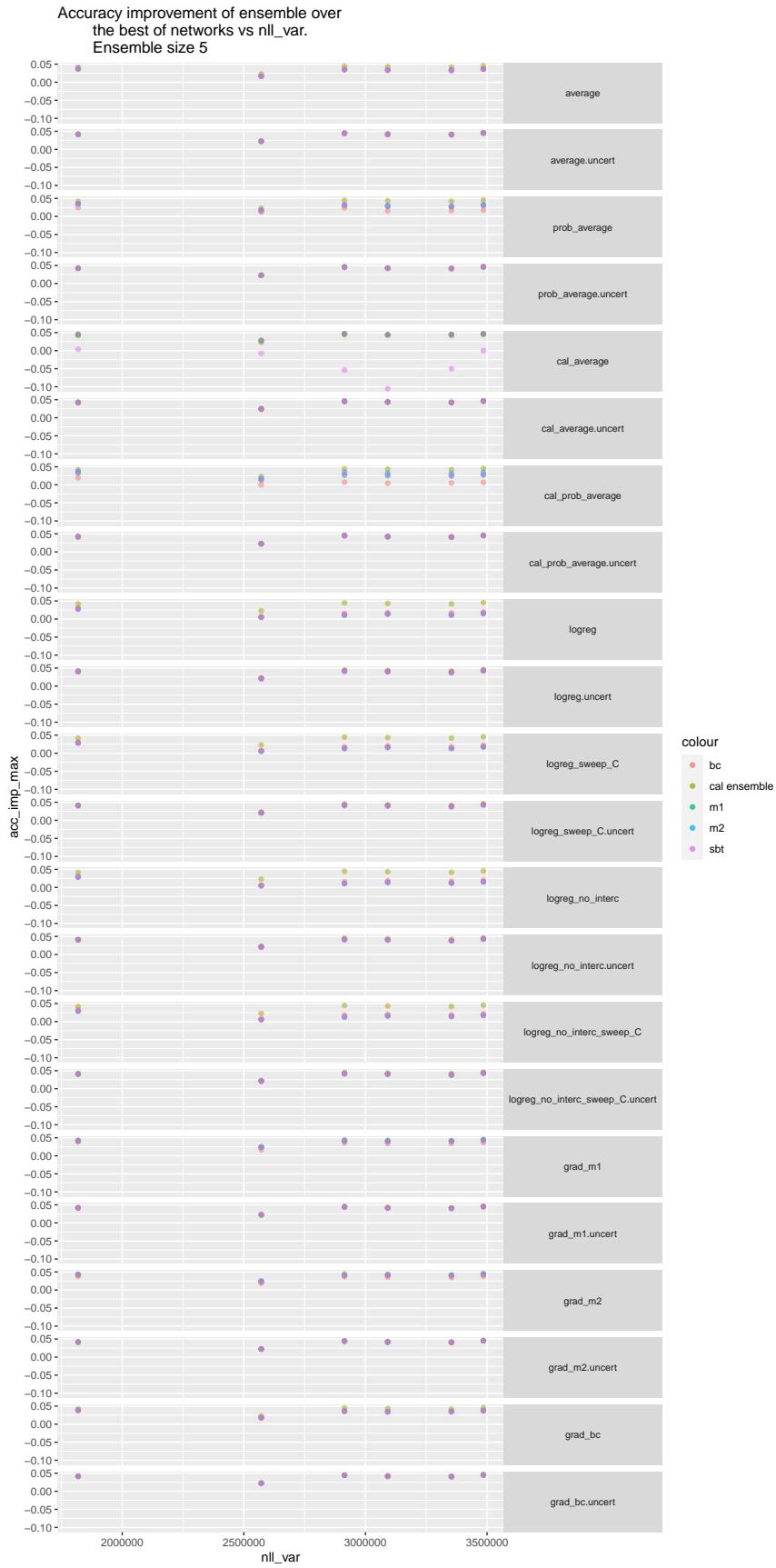
Accuracy improvement of ensemble over
the average of networks vs ece_avg.
Ensemble size 4

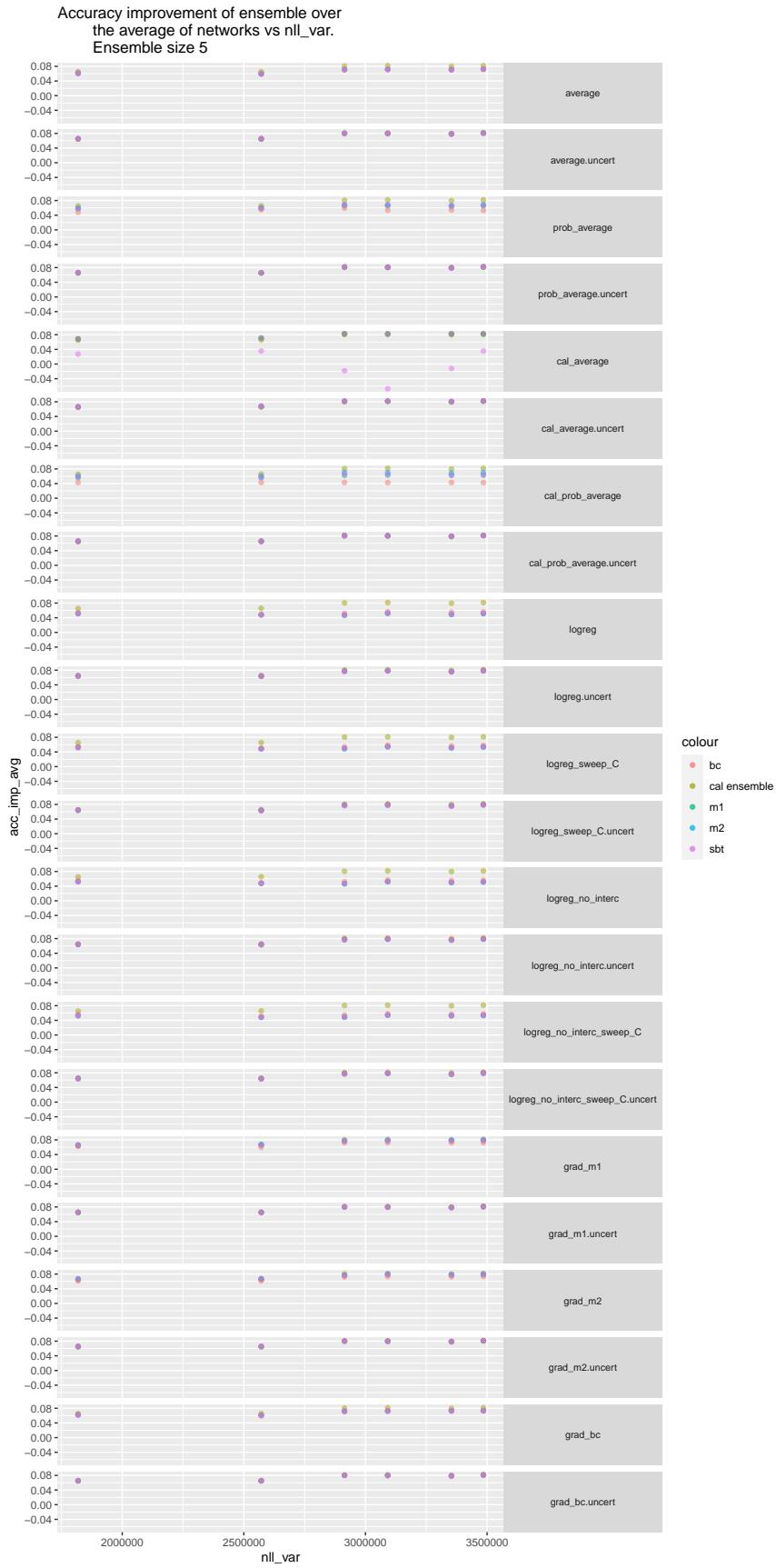


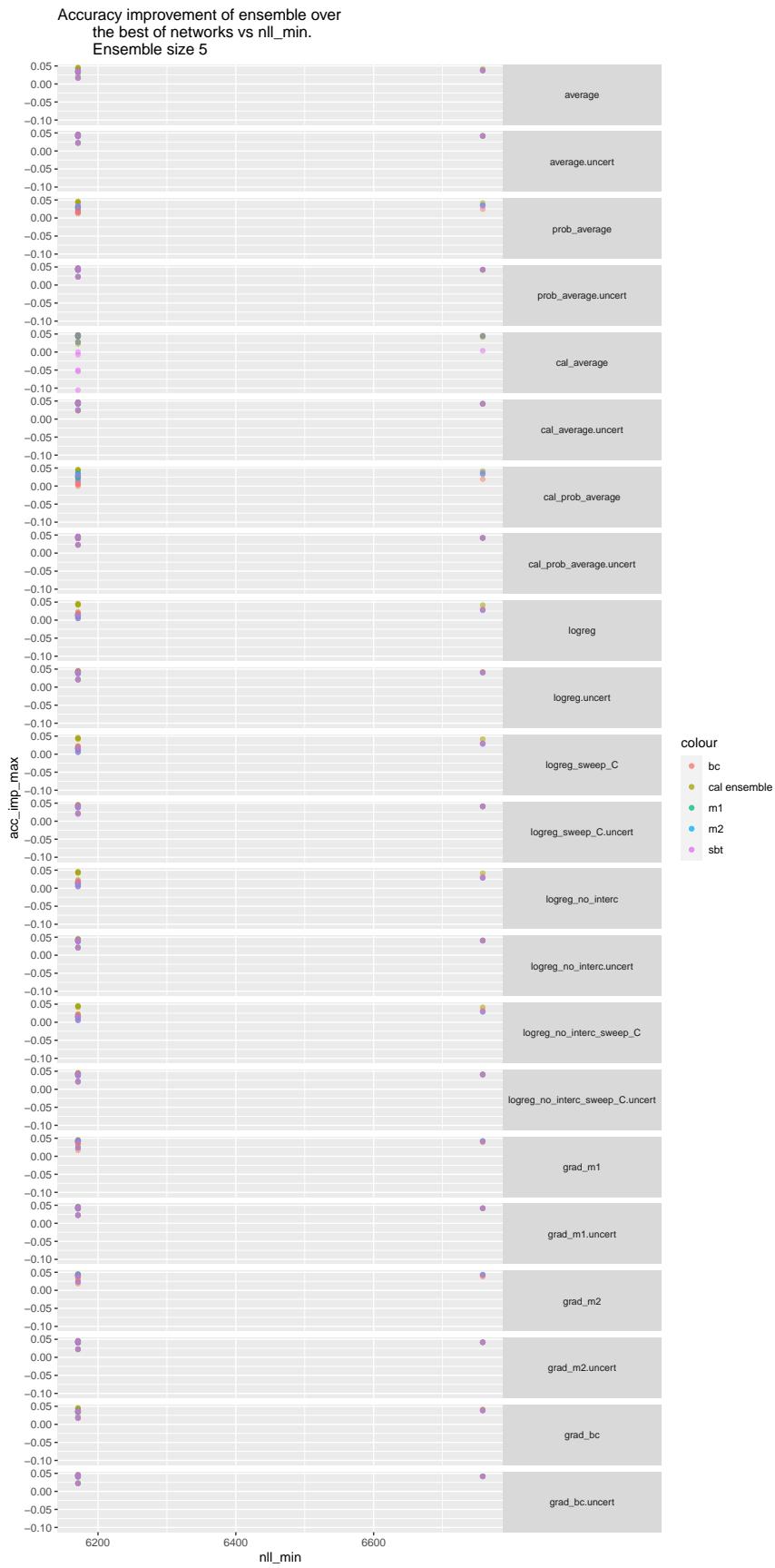


Accuracy improvement of ensemble over
the average of networks vs err_incons.
Ensemble size 5





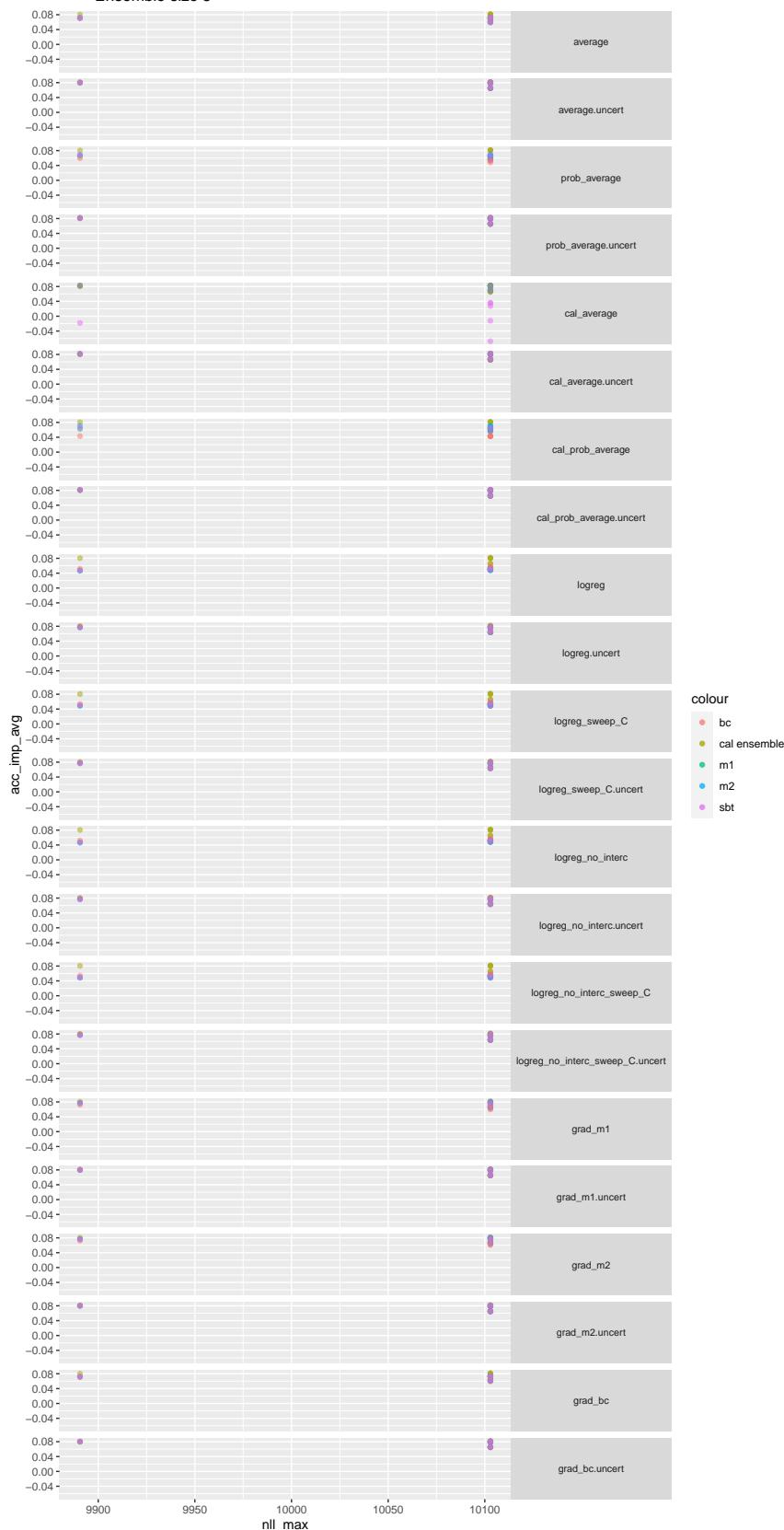


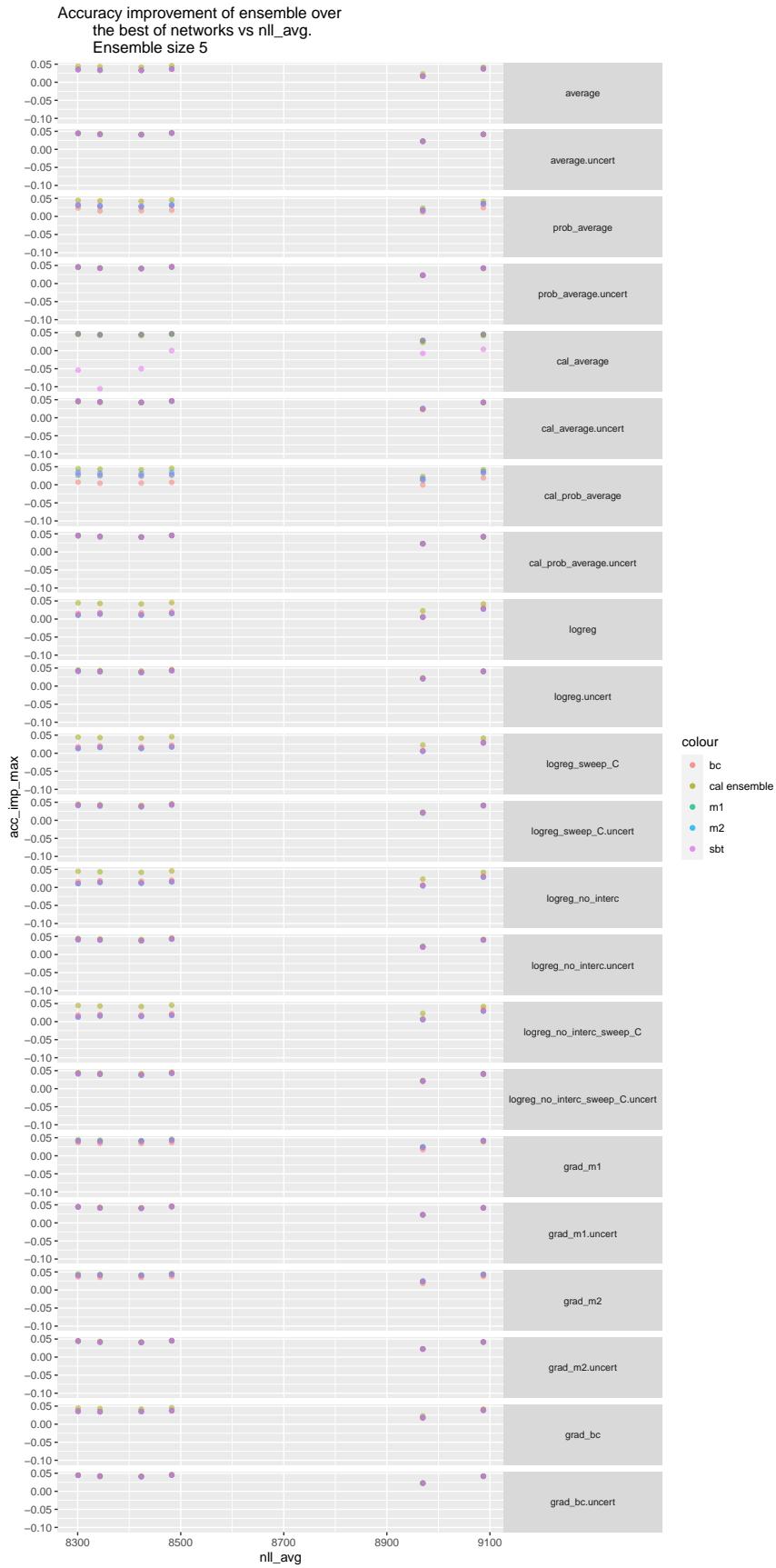


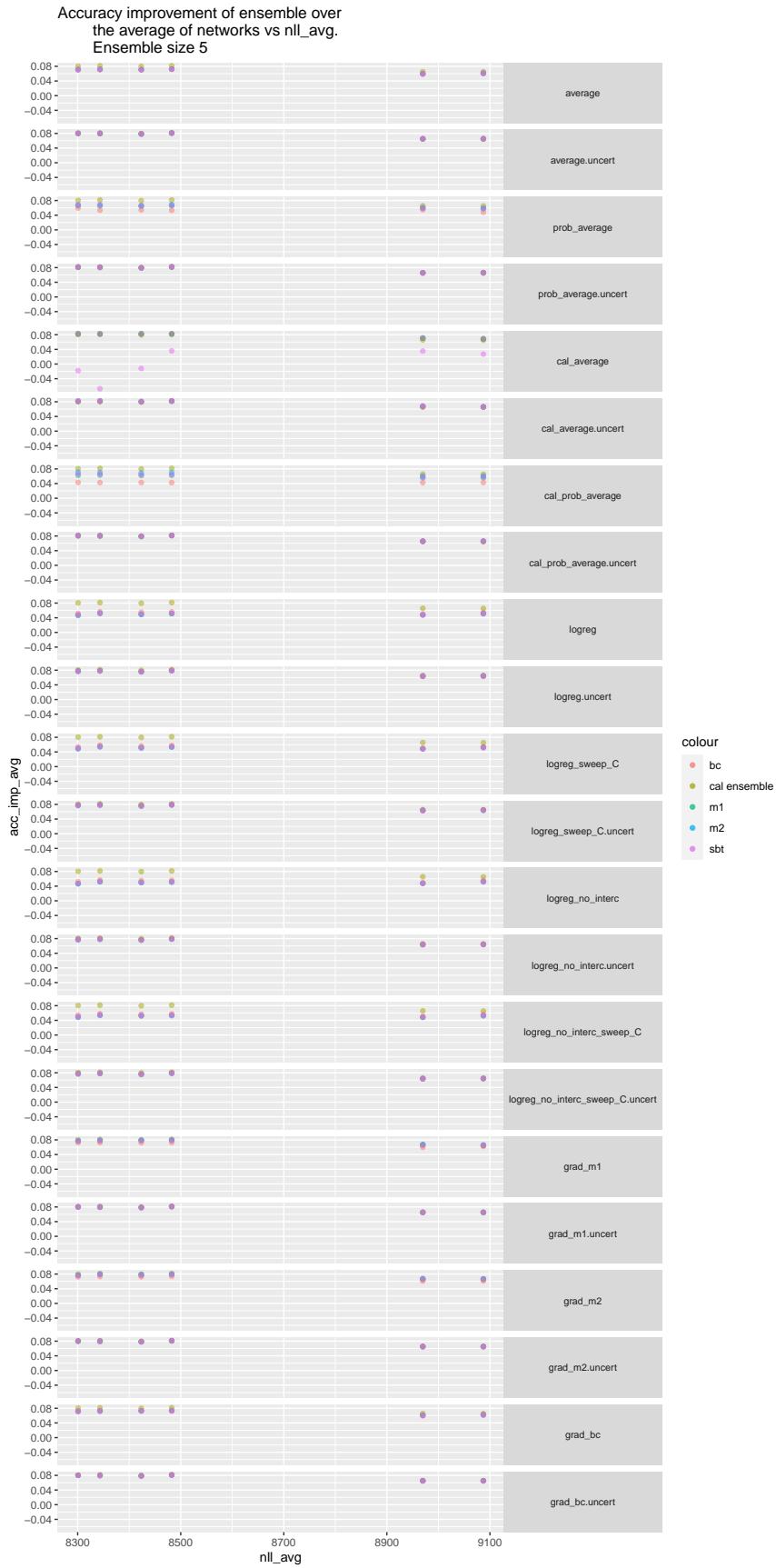


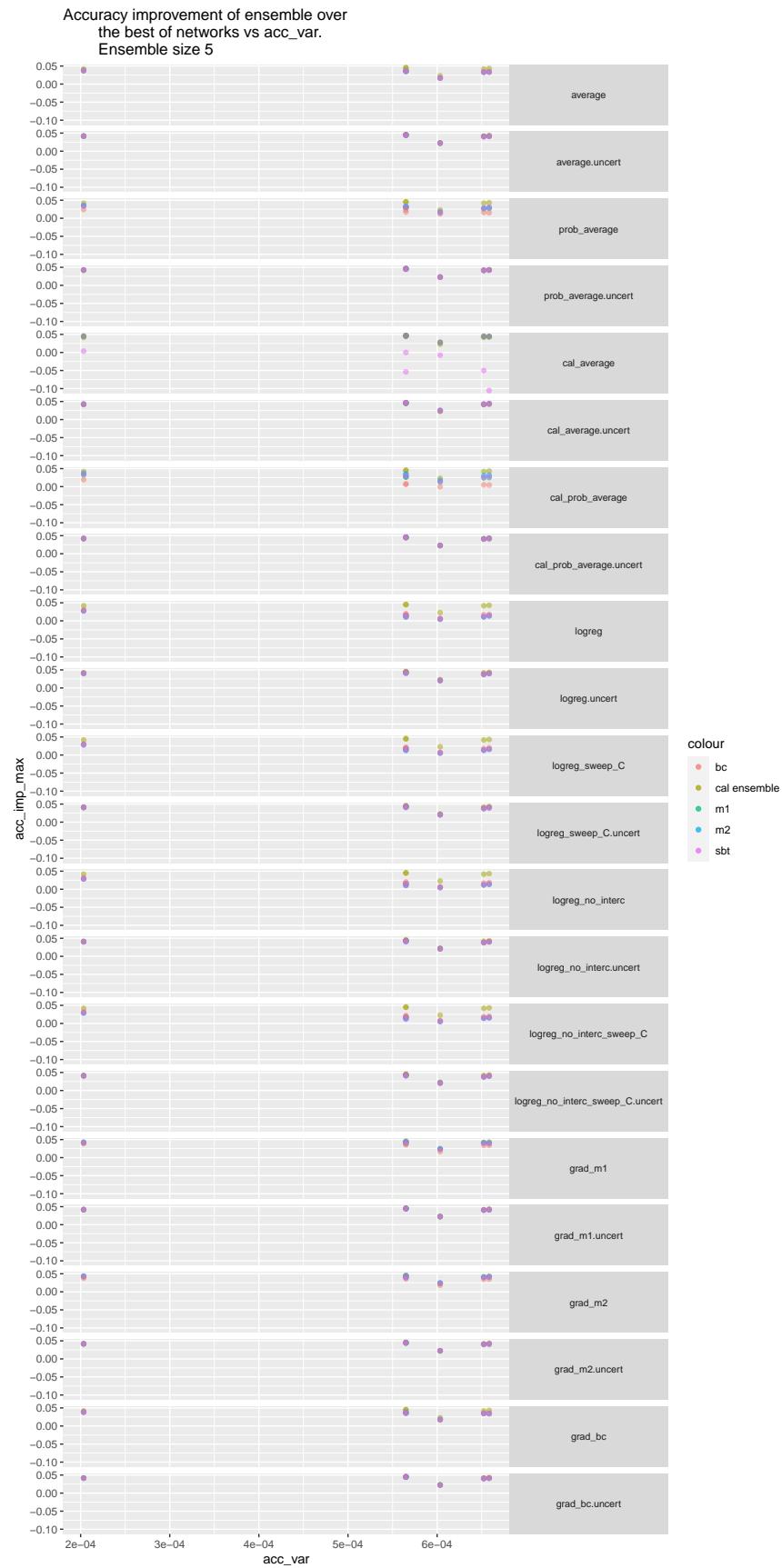


Accuracy improvement of ensemble over
the average of networks vs nll_max.
Ensemble size 5

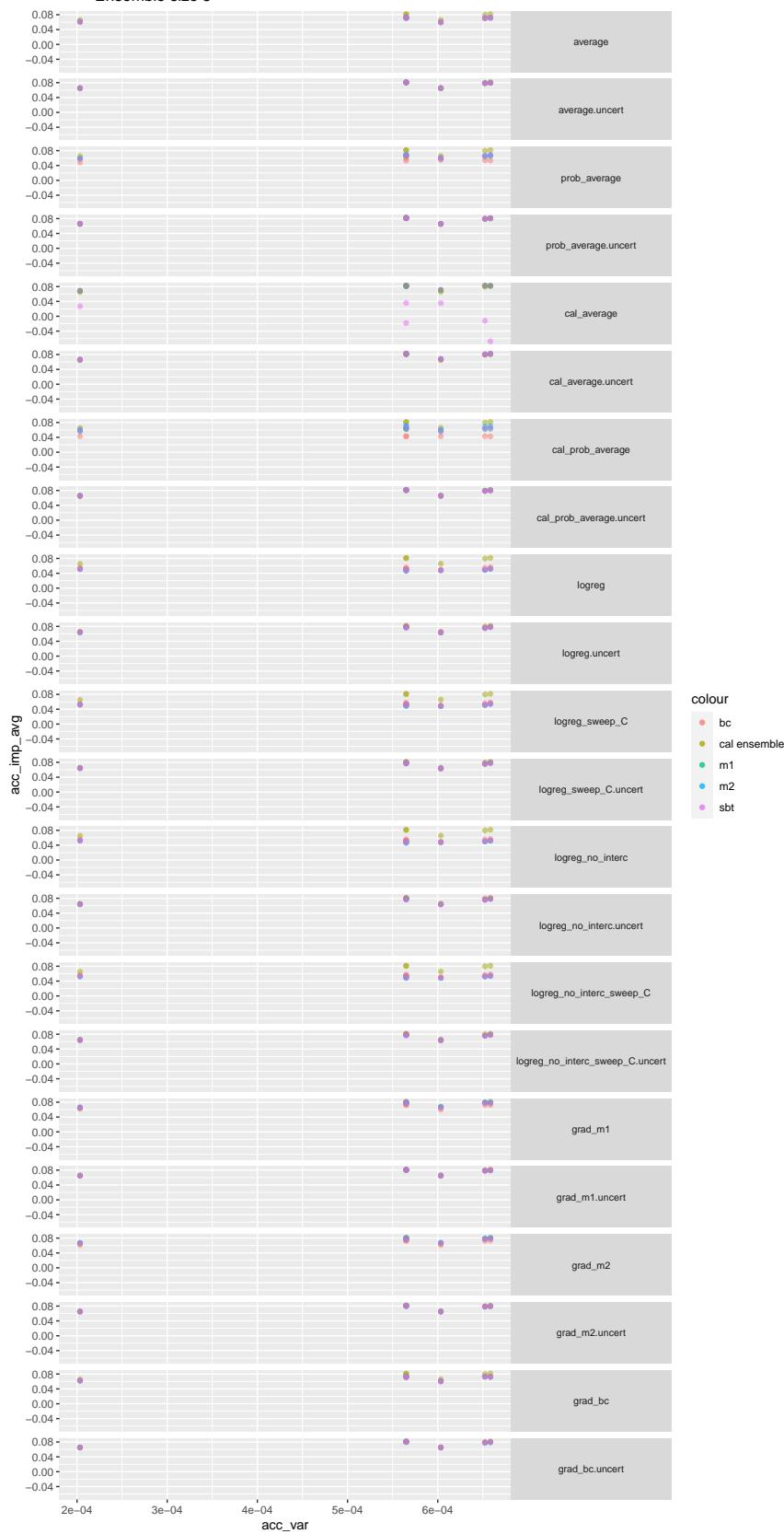


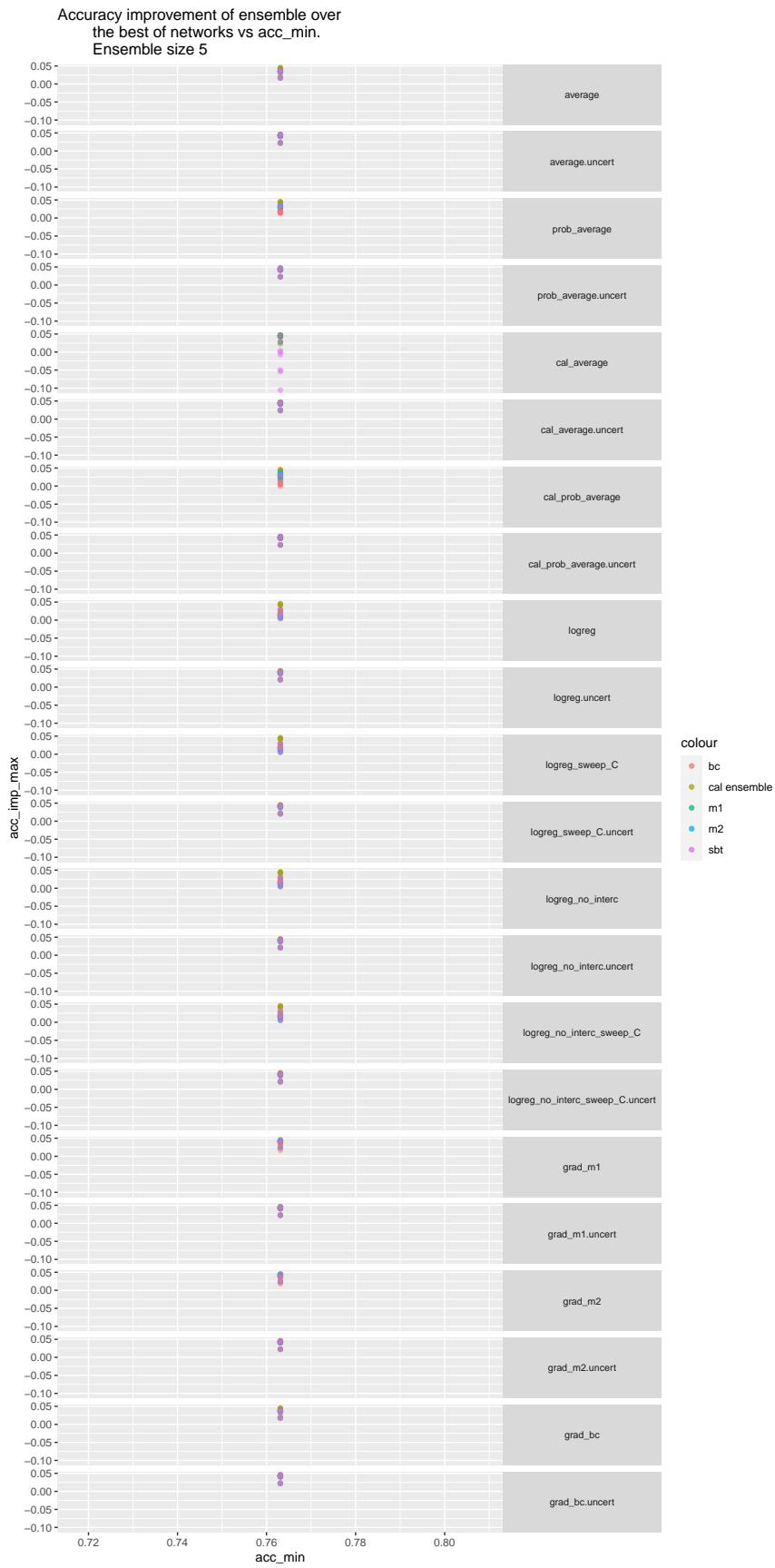


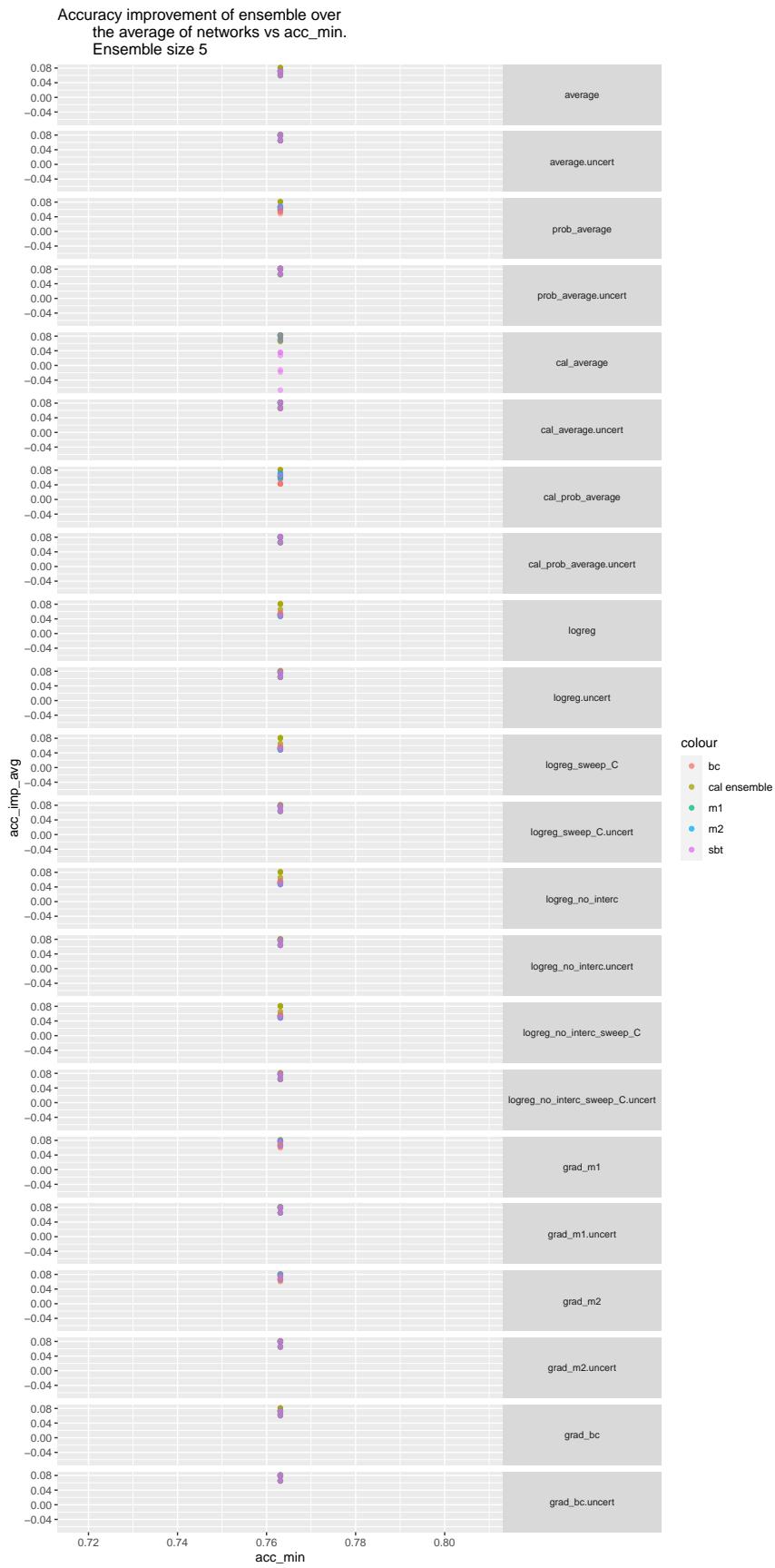




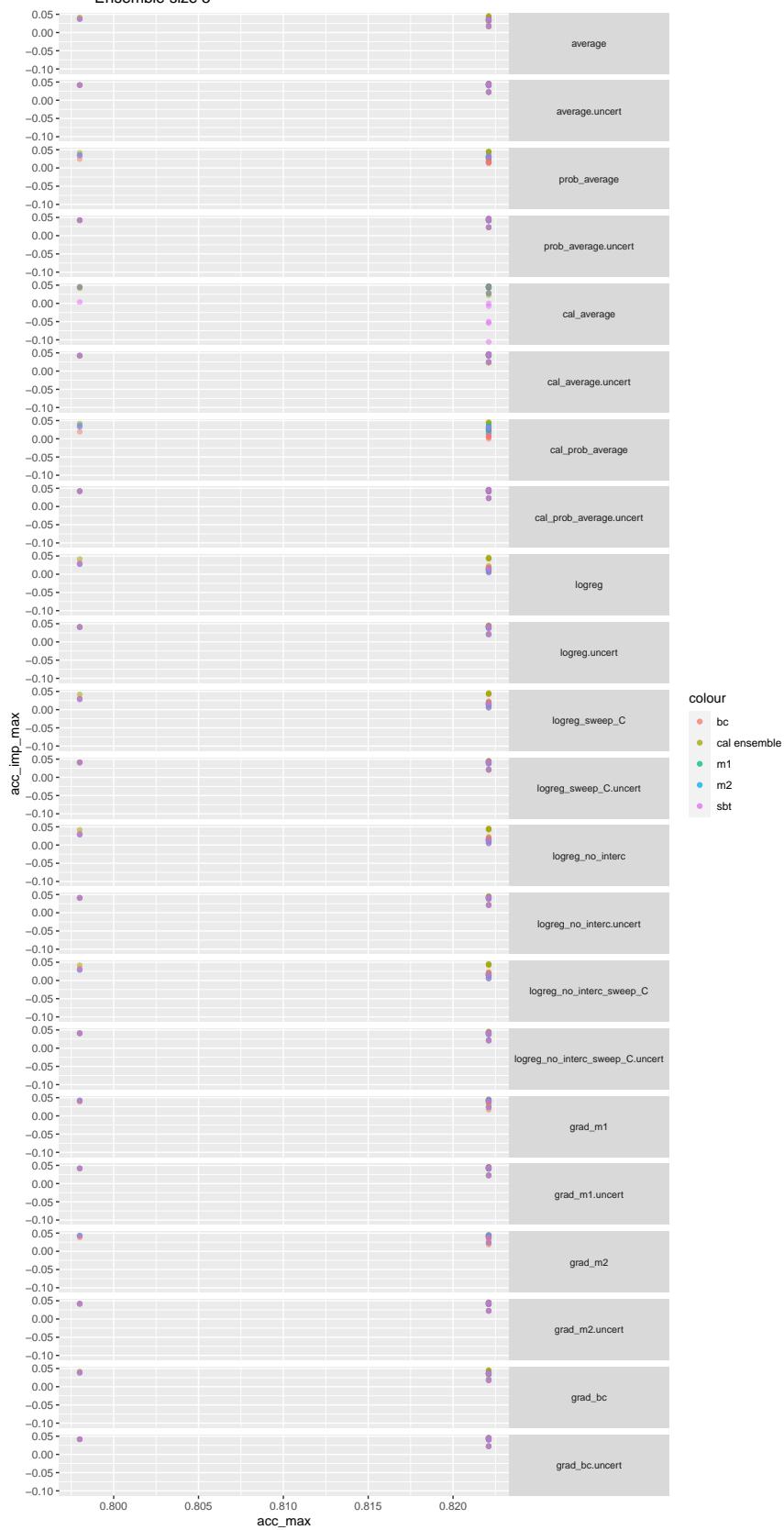
Accuracy improvement of ensemble over
the average of networks vs acc_var.
Ensemble size 5





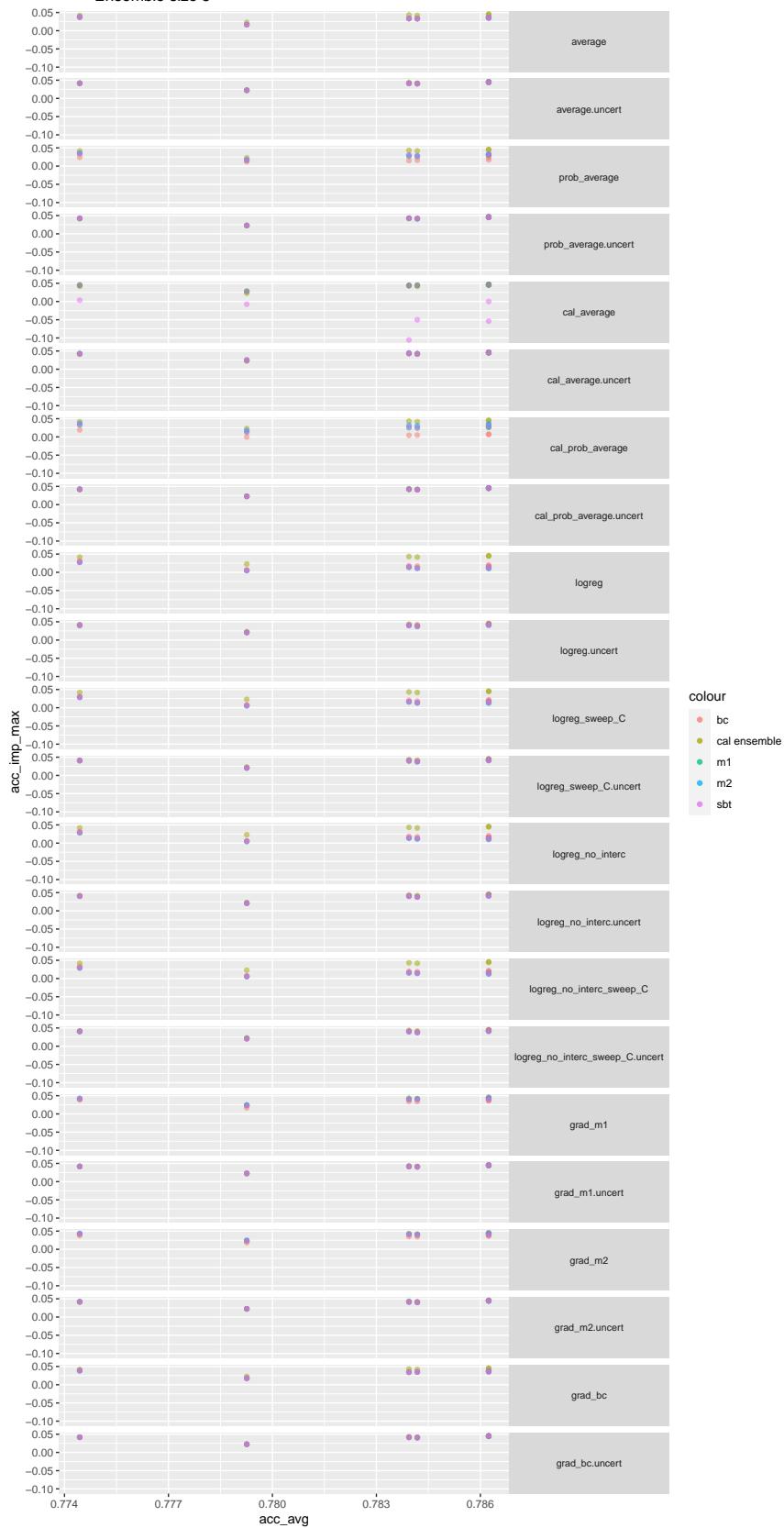


Accuracy improvement of ensemble over
the best of networks vs acc_max.
Ensemble size 5



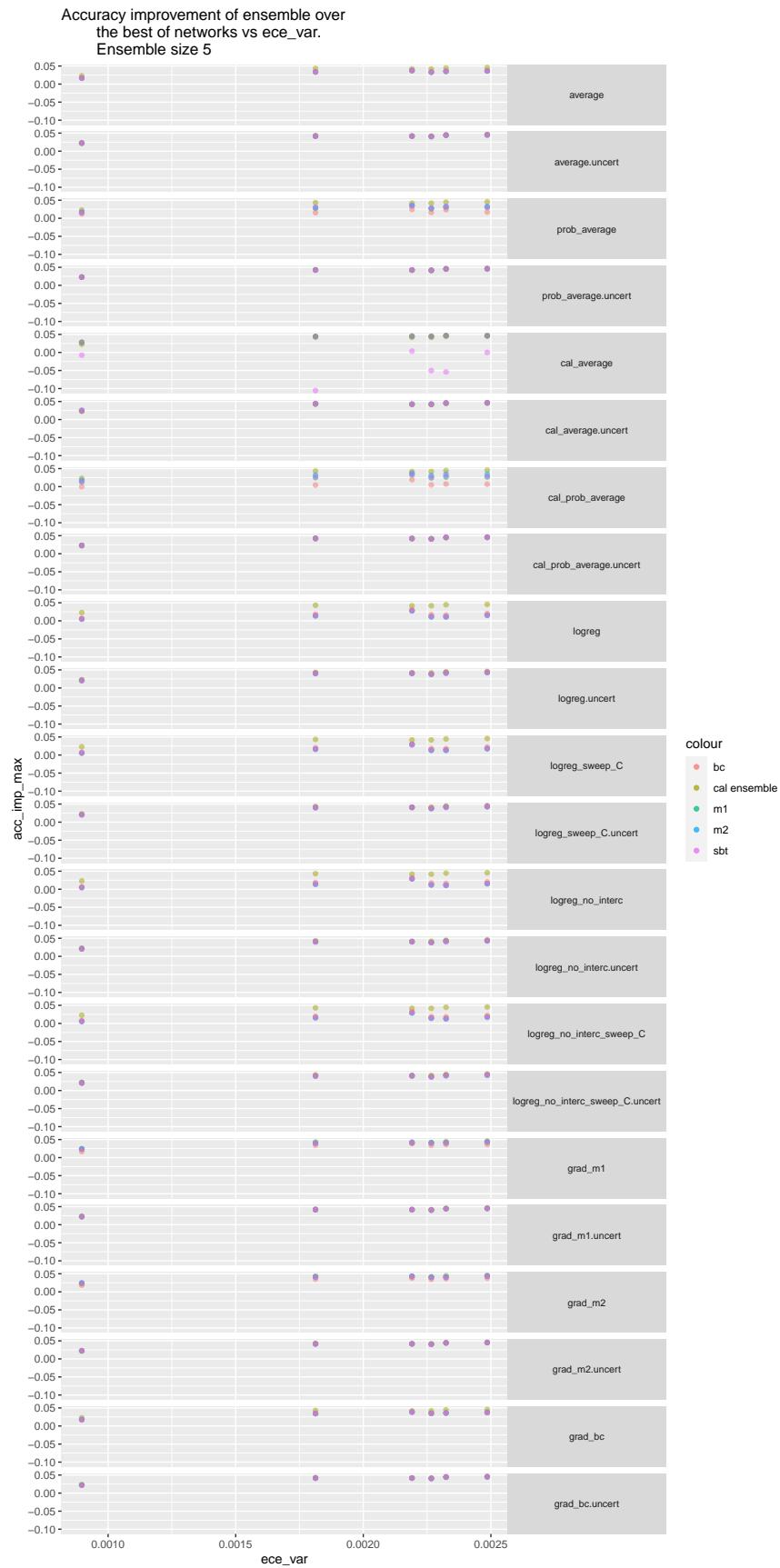


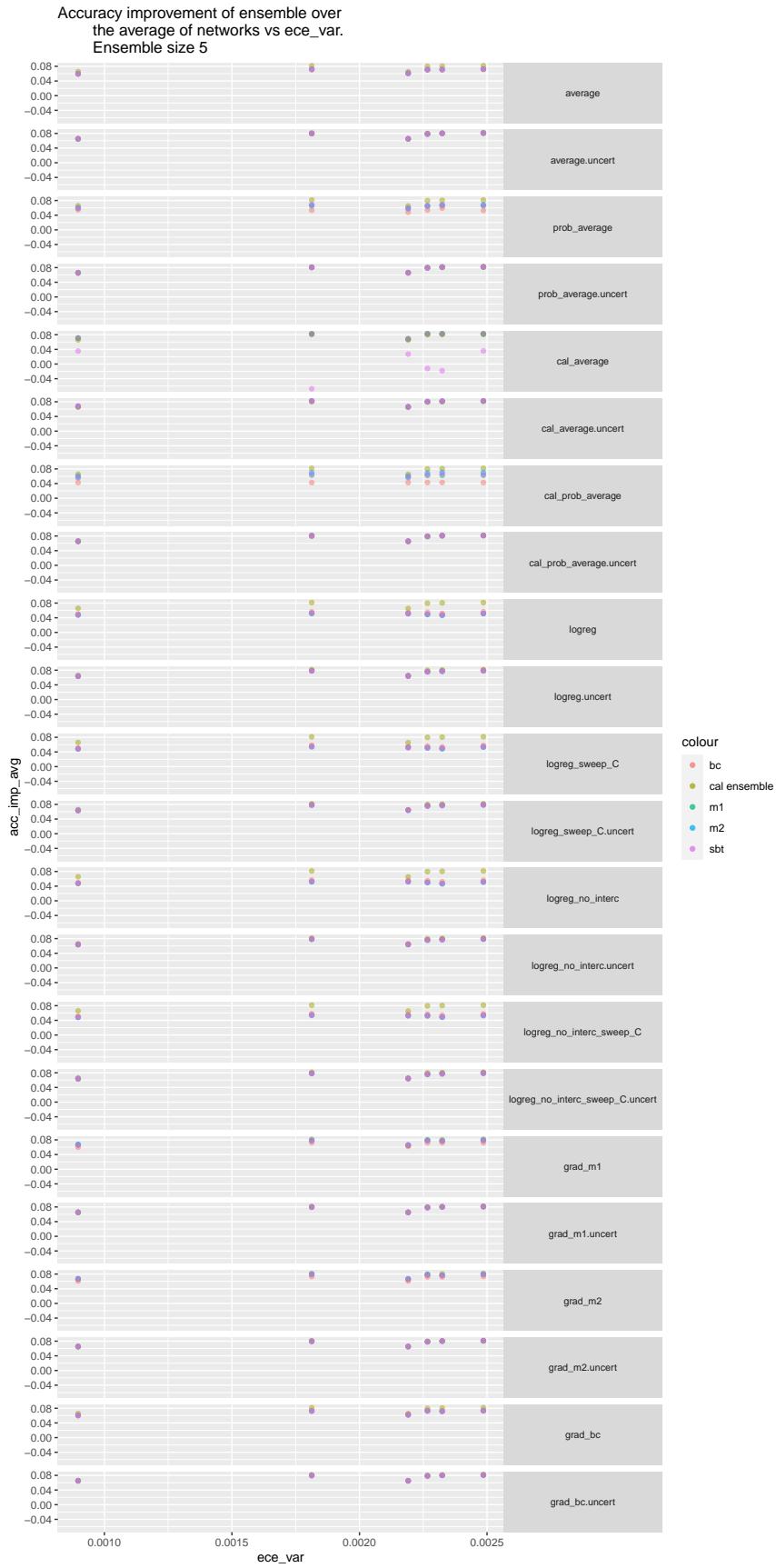
Accuracy improvement of ensemble over
the best of networks vs acc_avg.
Ensemble size 5



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_min.
Ensemble size 5



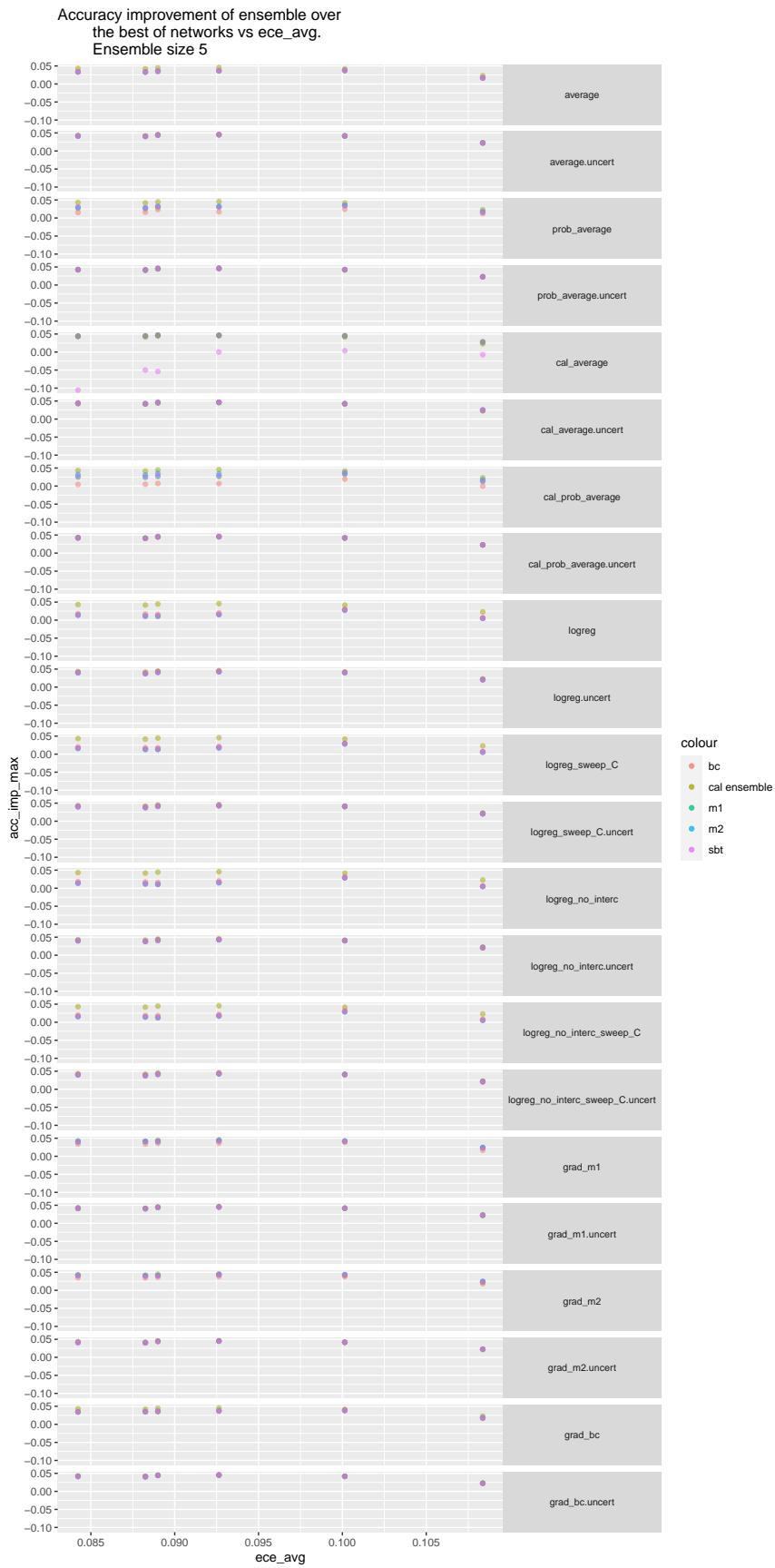
Accuracy improvement of ensemble over
the average of networks vs ece_min.
Ensemble size 5

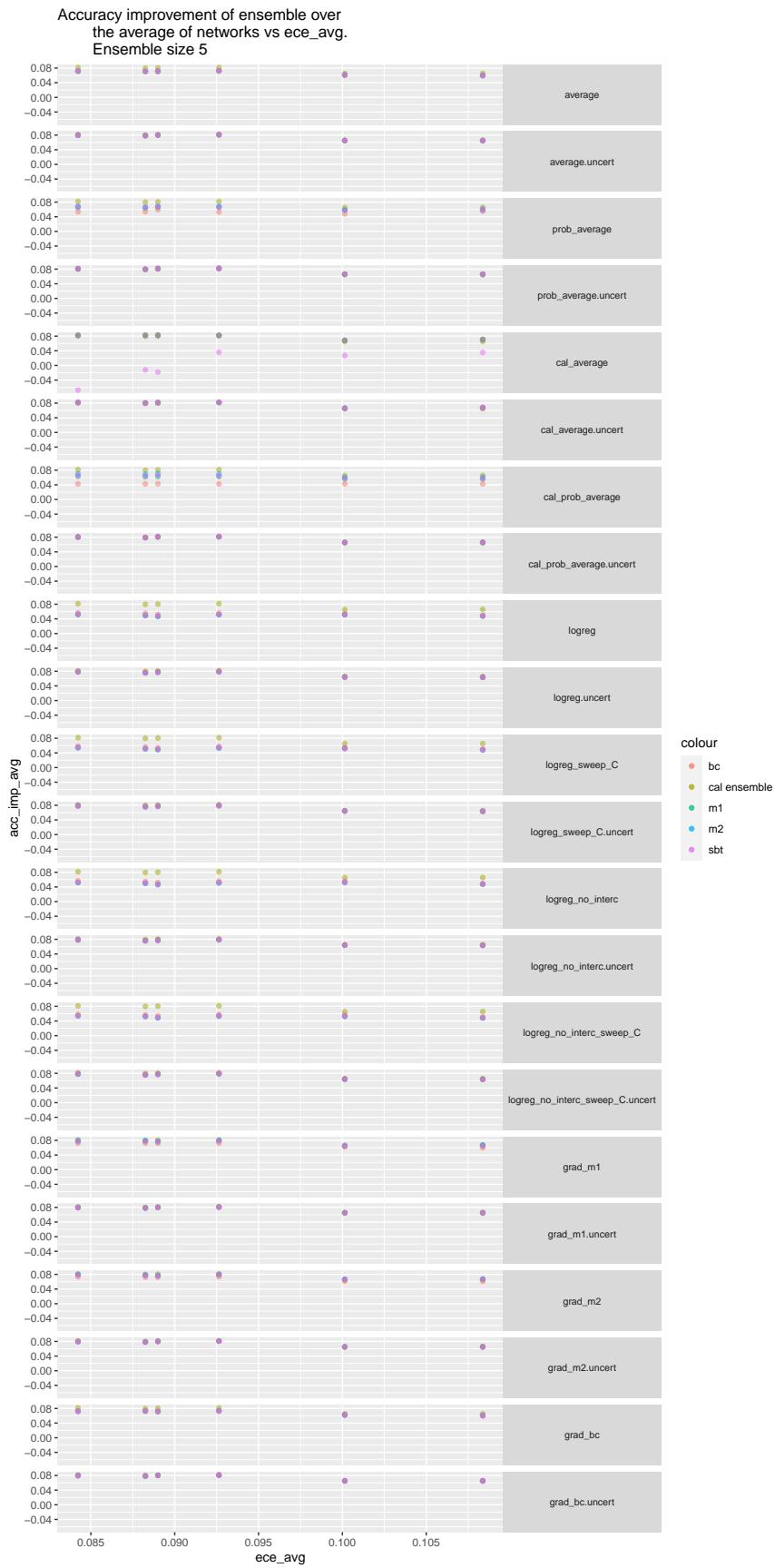




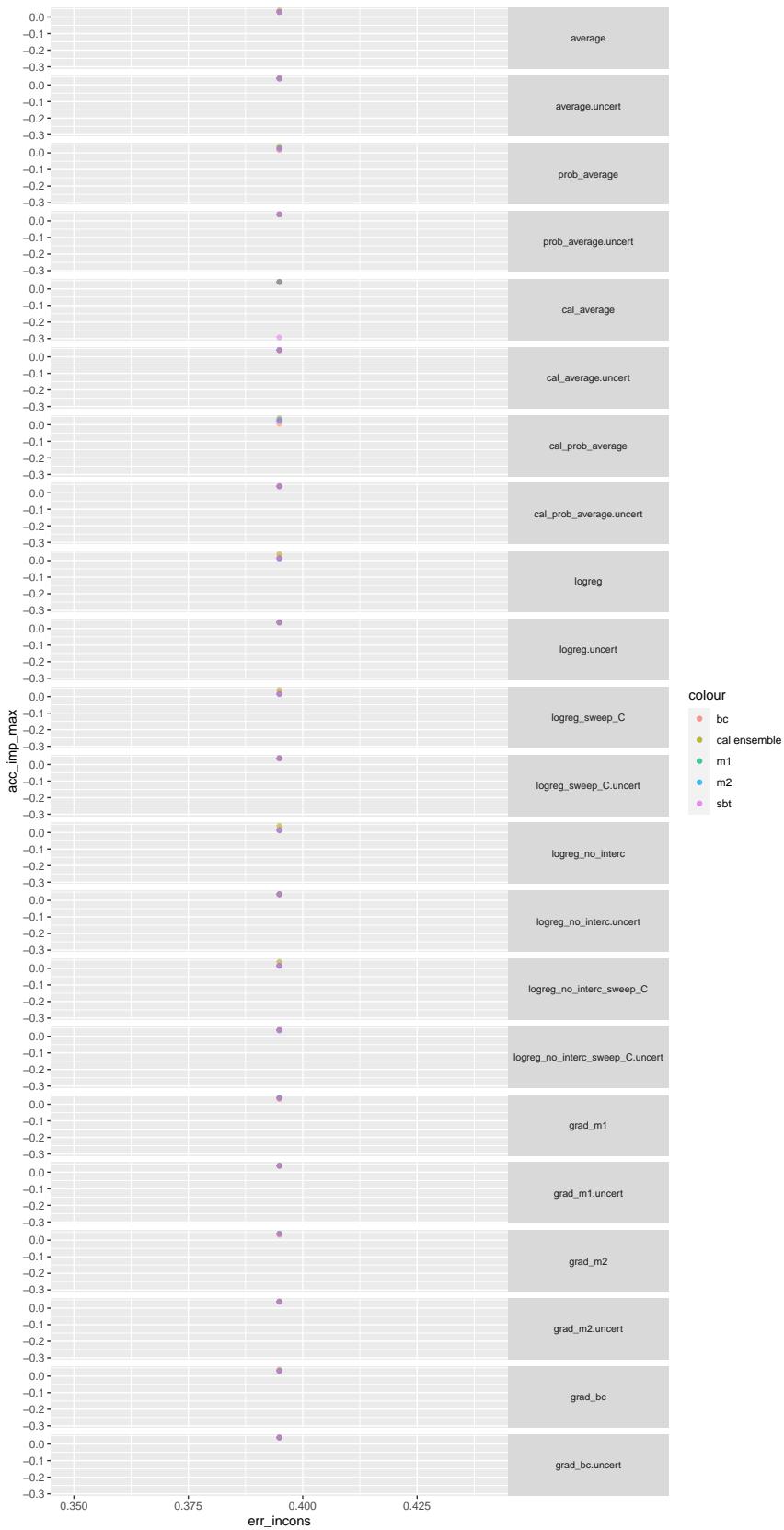
Accuracy improvement of ensemble over
the average of networks vs ece_max.
Ensemble size 5



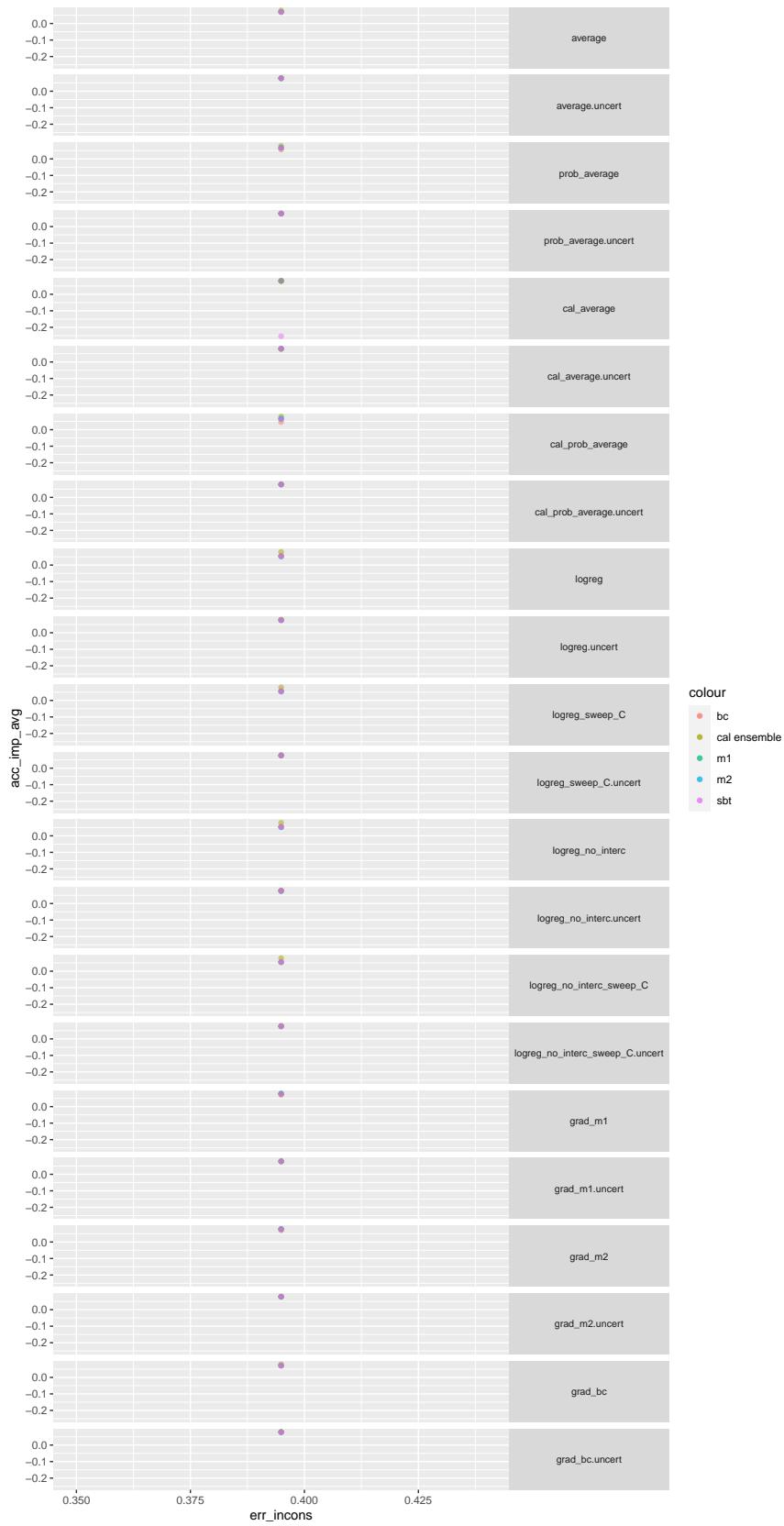




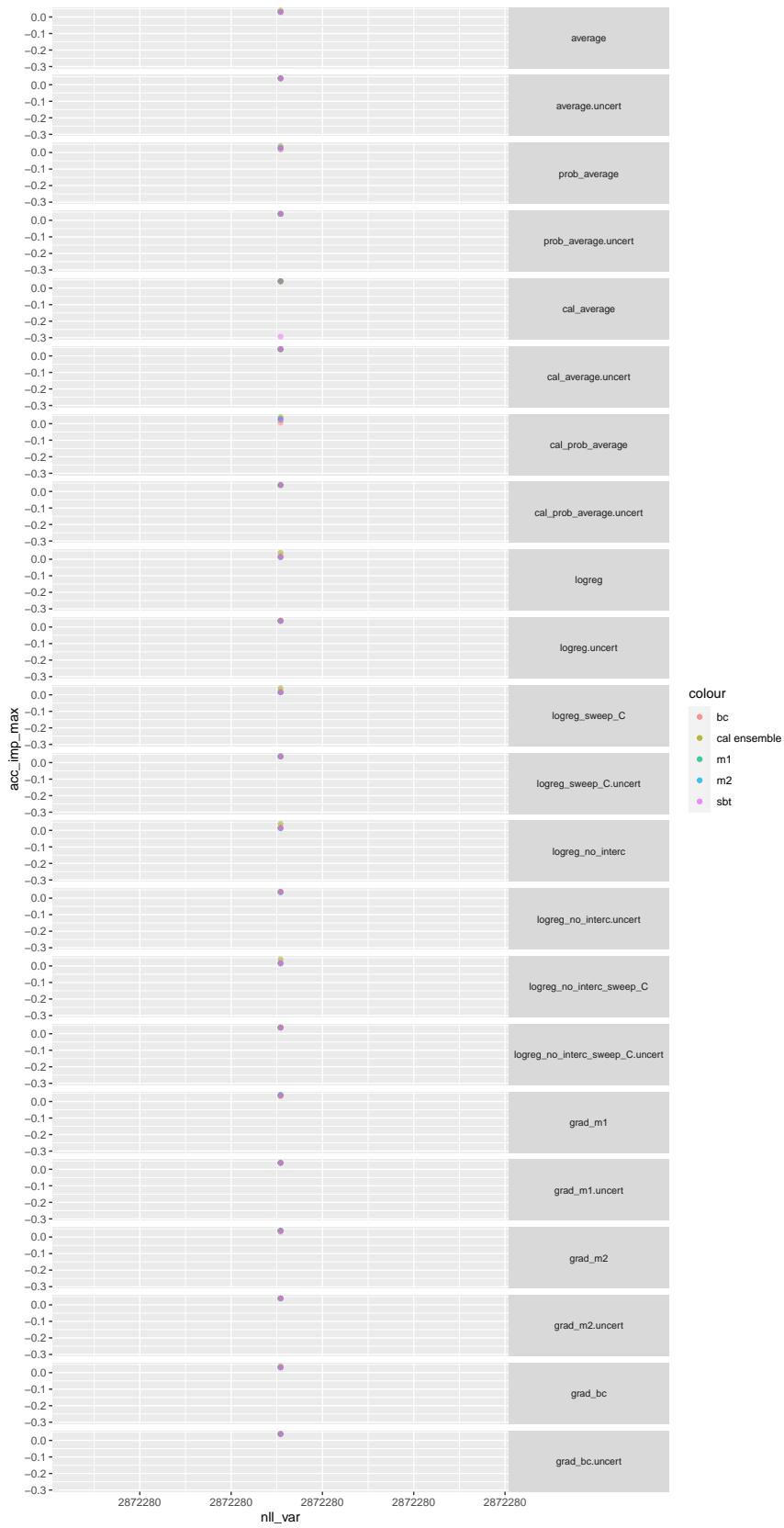
Accuracy improvement of ensemble over
the best of networks vs err_incons.
Ensemble size 6



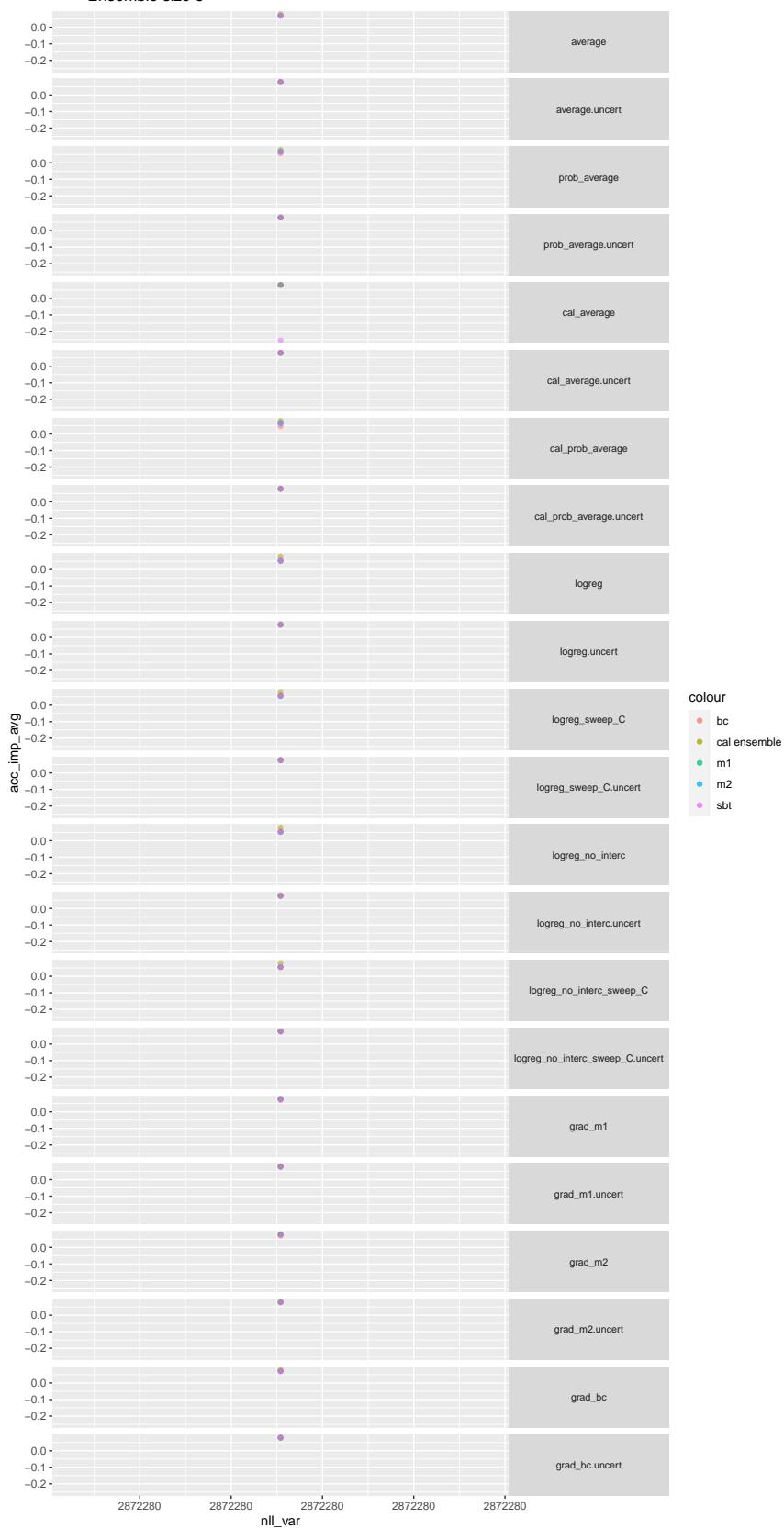
Accuracy improvement of ensemble over
the average of networks vs err_incons.
Ensemble size 6



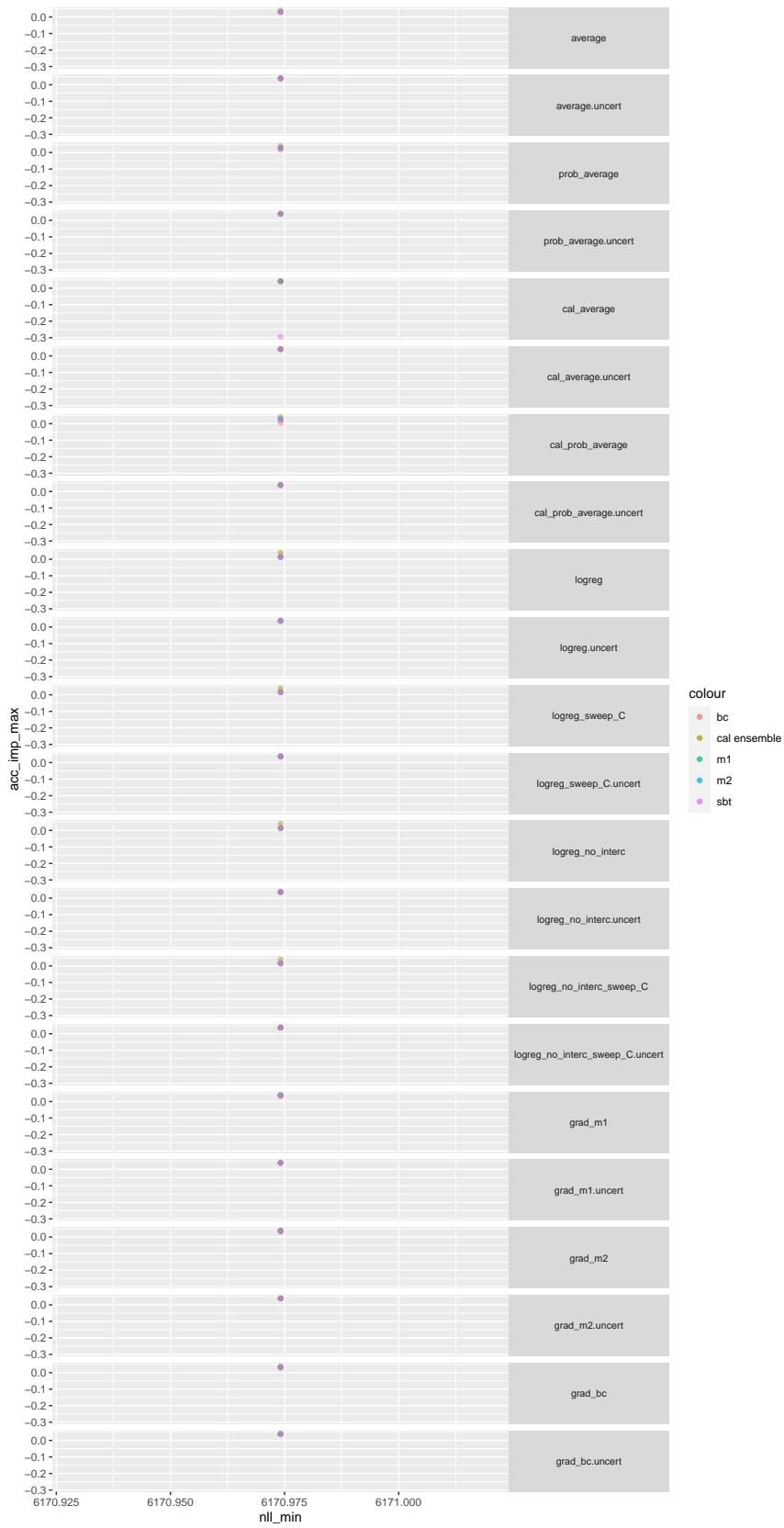
Accuracy improvement of ensemble over
the best of networks vs nll_var.
Ensemble size 6



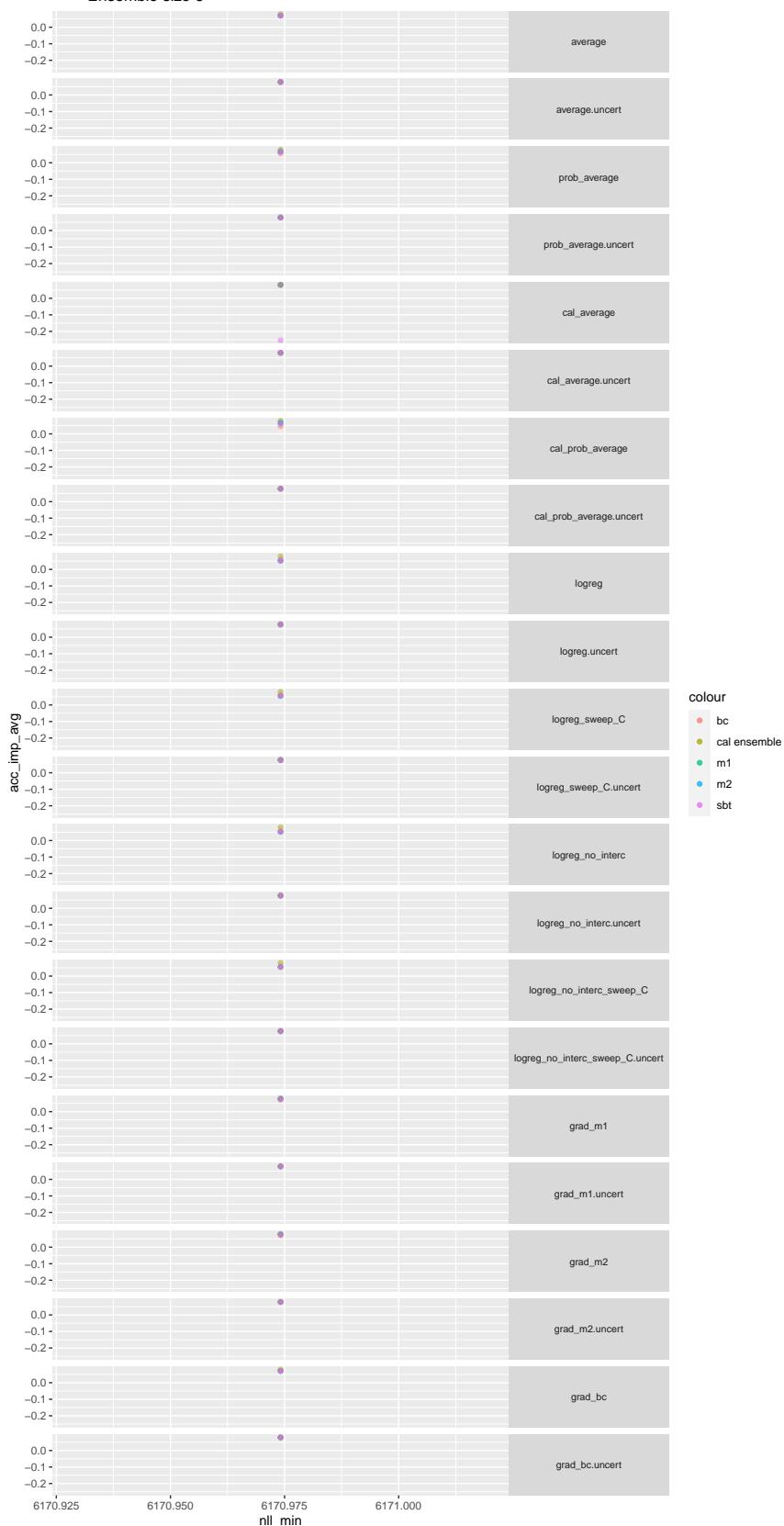
Accuracy improvement of ensemble over
the average of networks vs nll_var.
Ensemble size 6



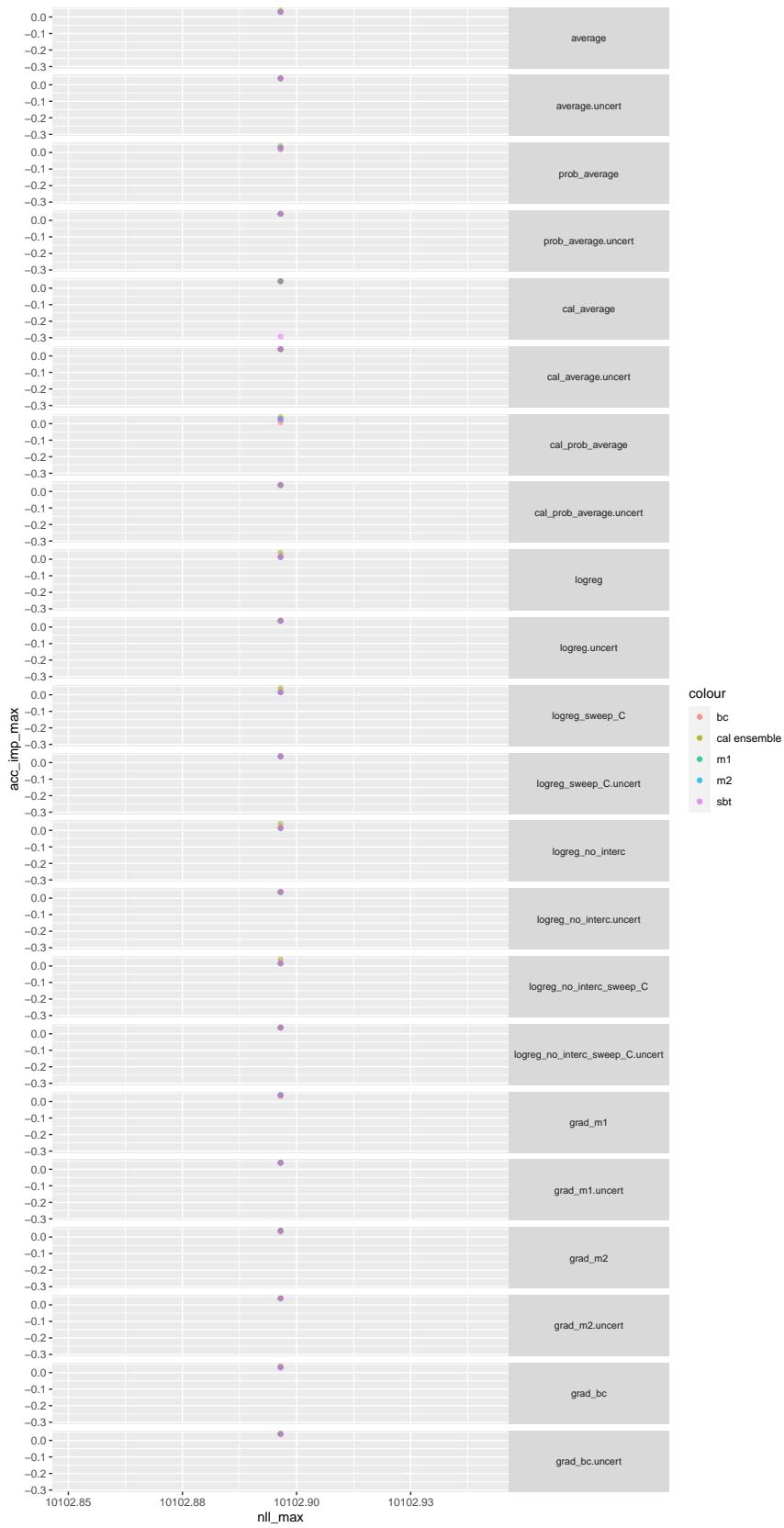
Accuracy improvement of ensemble over
the best of networks vs nll_min.
Ensemble size 6



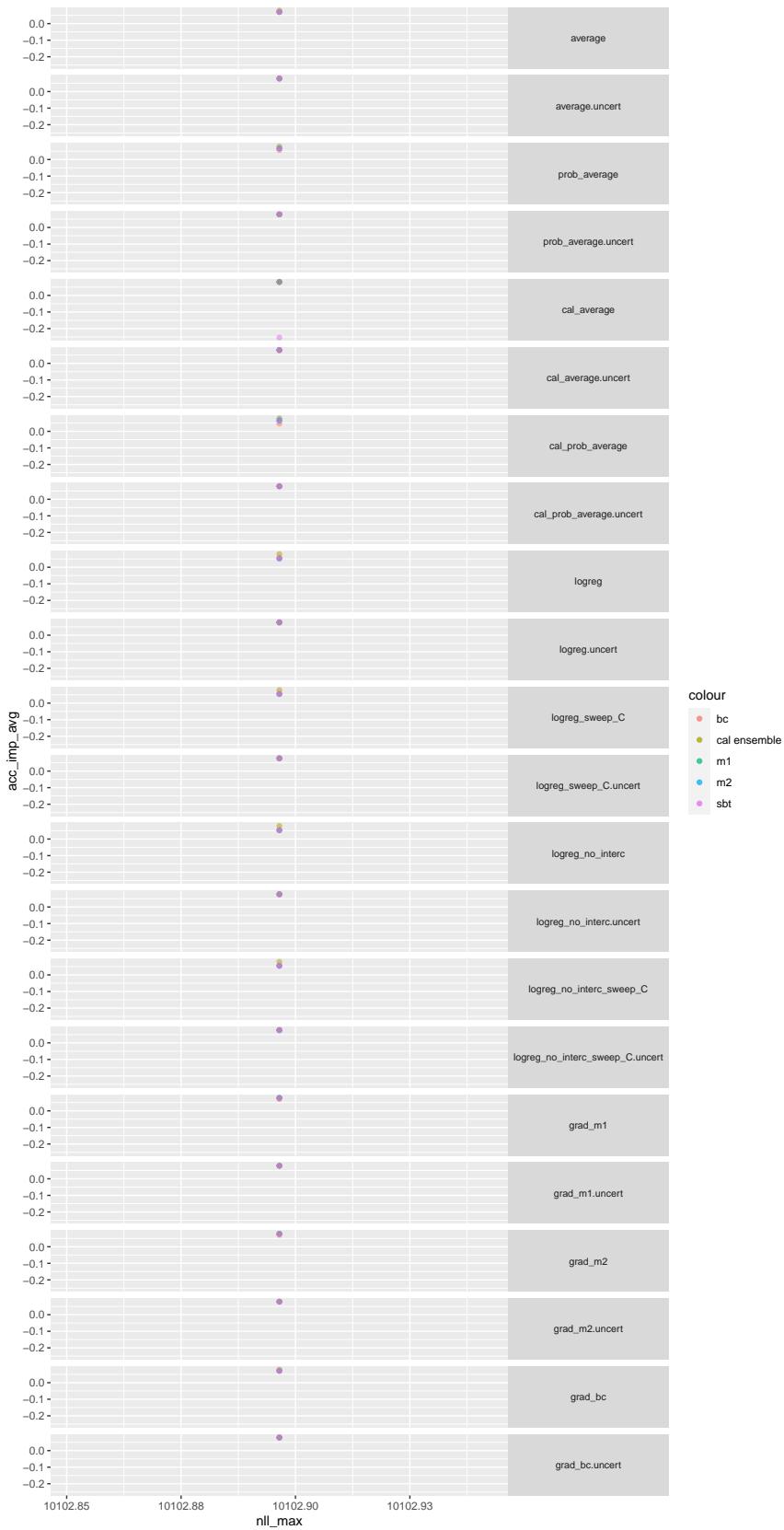
Accuracy improvement of ensemble over
the average of networks vs nll_min.
Ensemble size 6



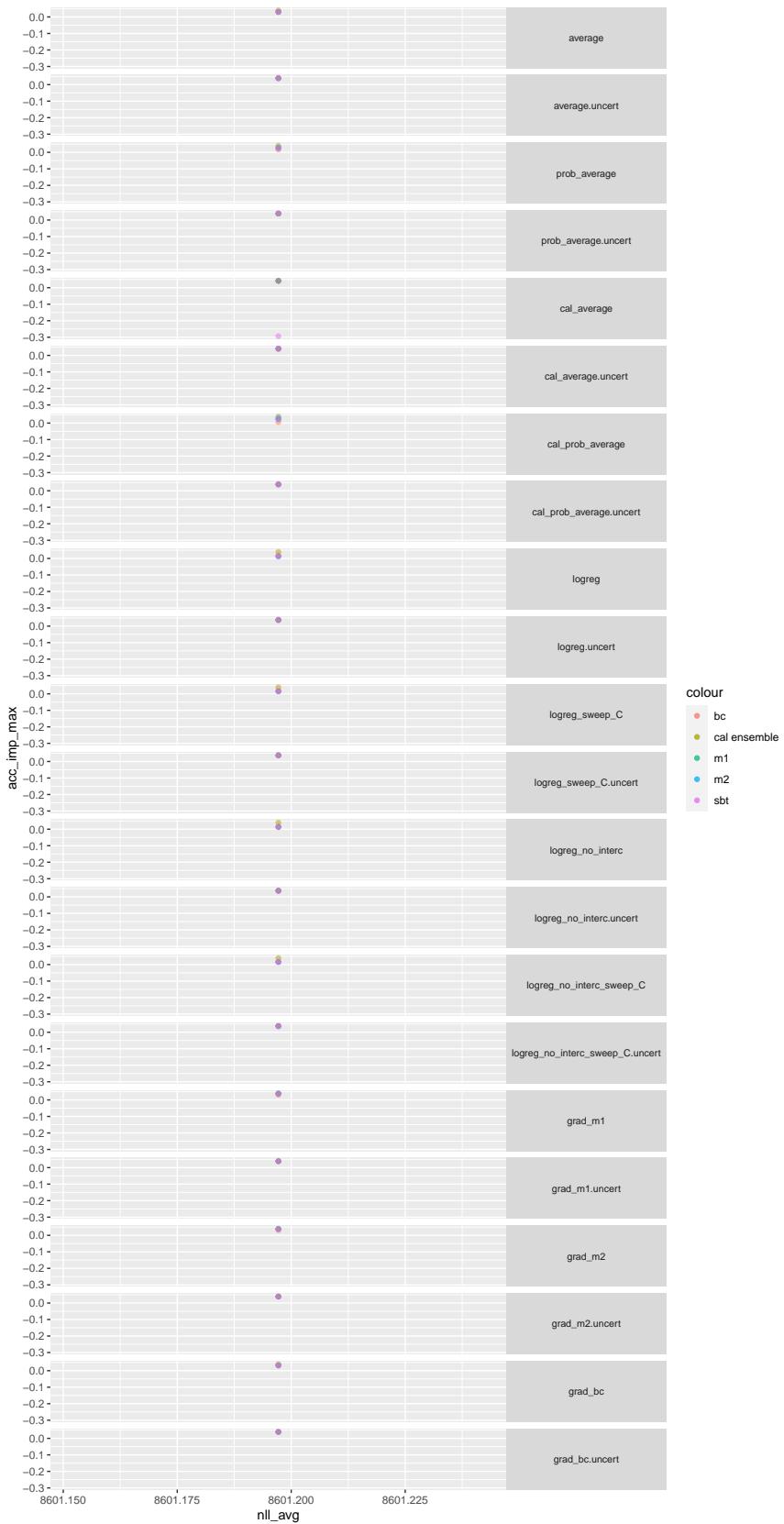
Accuracy improvement of ensemble over
the best of networks vs nll_max.
Ensemble size 6



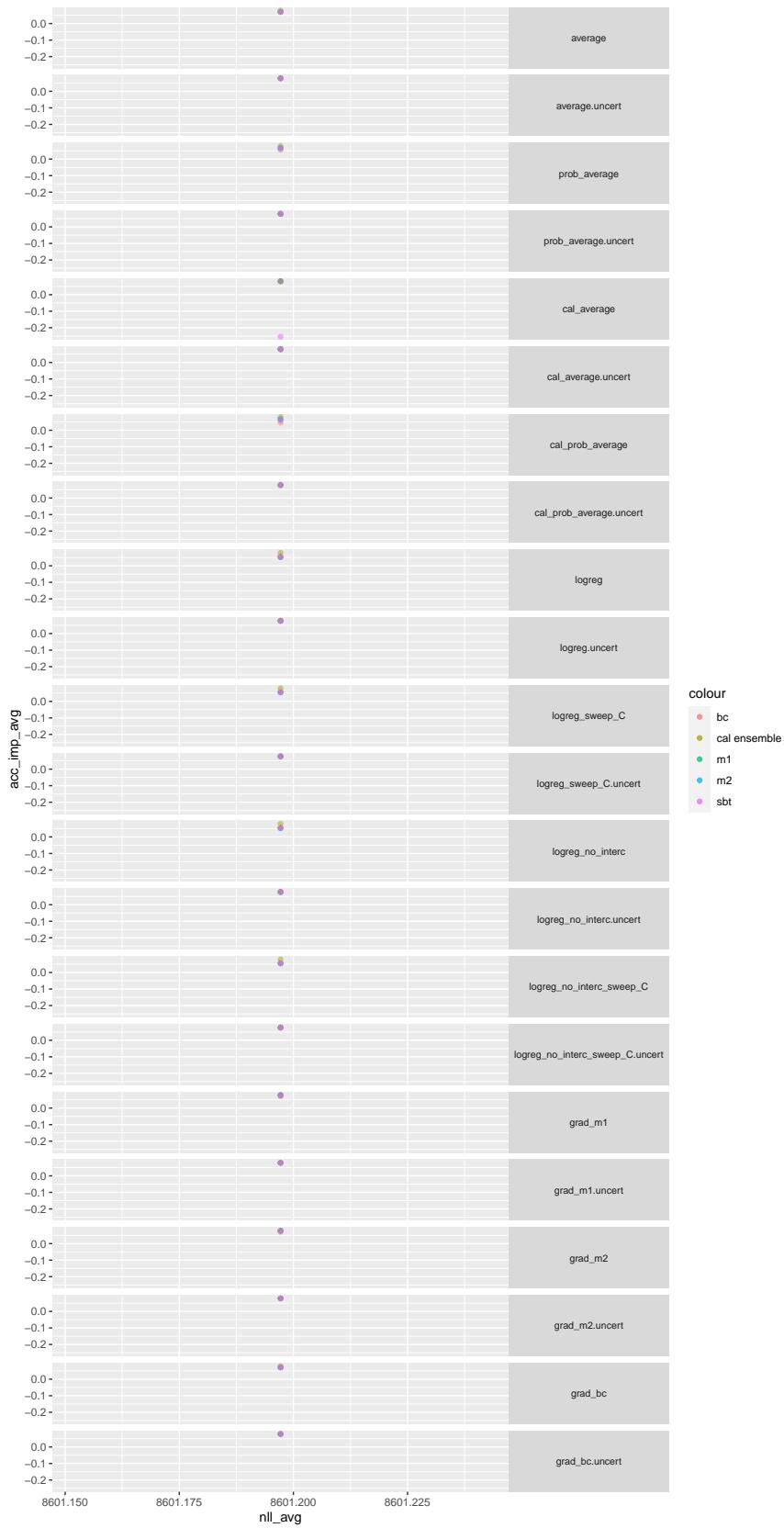
Accuracy improvement of ensemble over
the average of networks vs nll_max.
Ensemble size 6



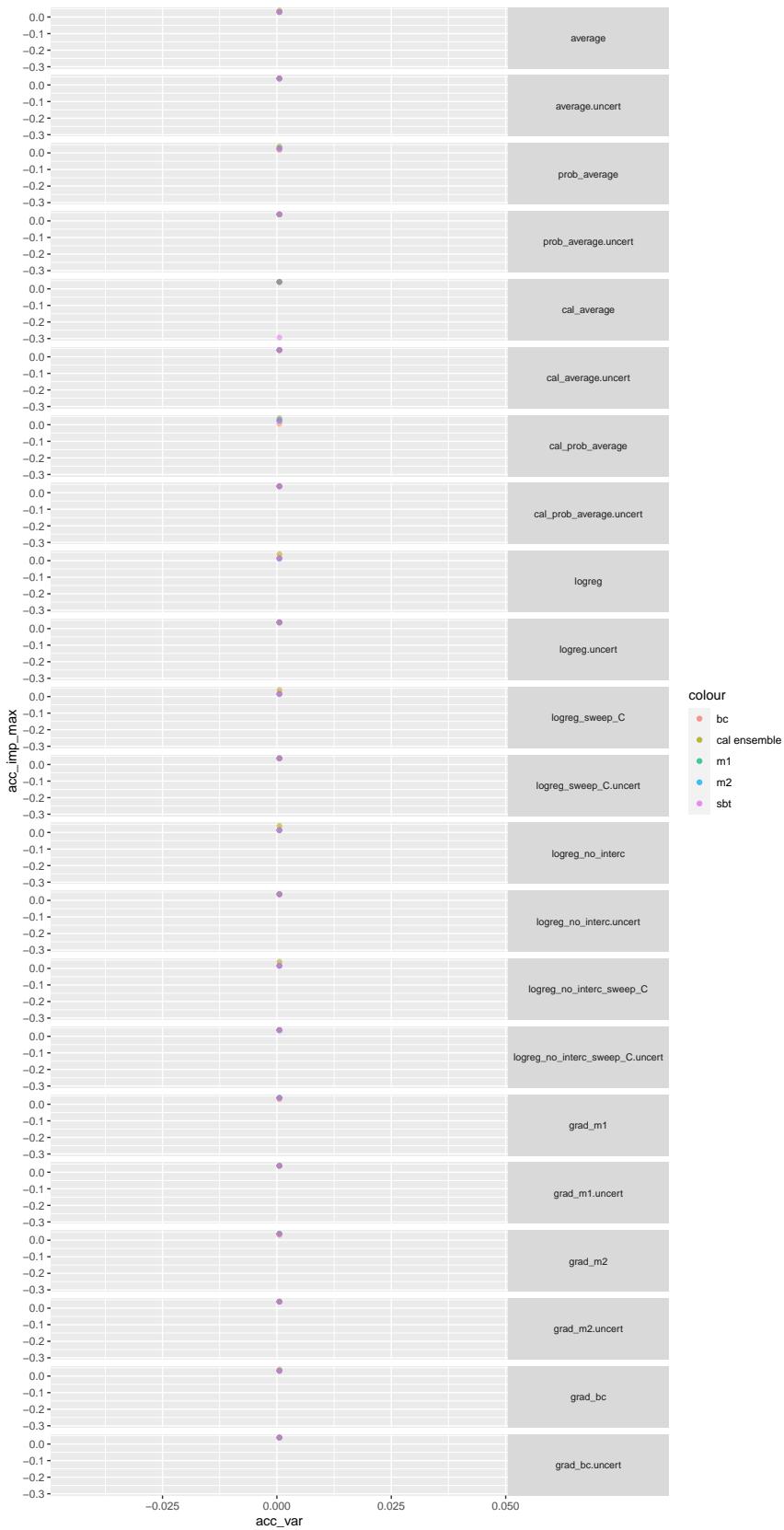
Accuracy improvement of ensemble over
the best of networks vs nll_avg.
Ensemble size 6



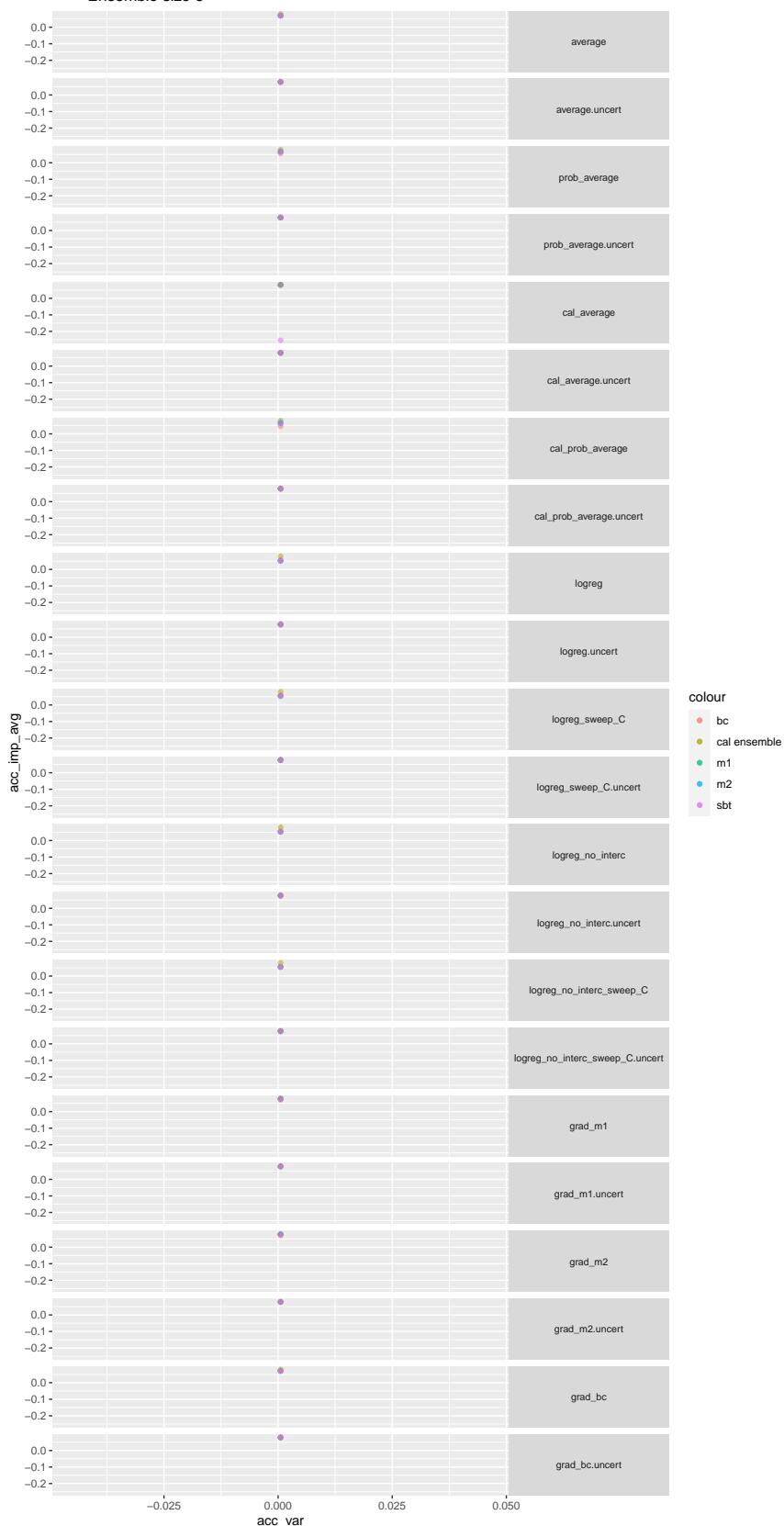
Accuracy improvement of ensemble over
the average of networks vs nll_avg.
Ensemble size 6



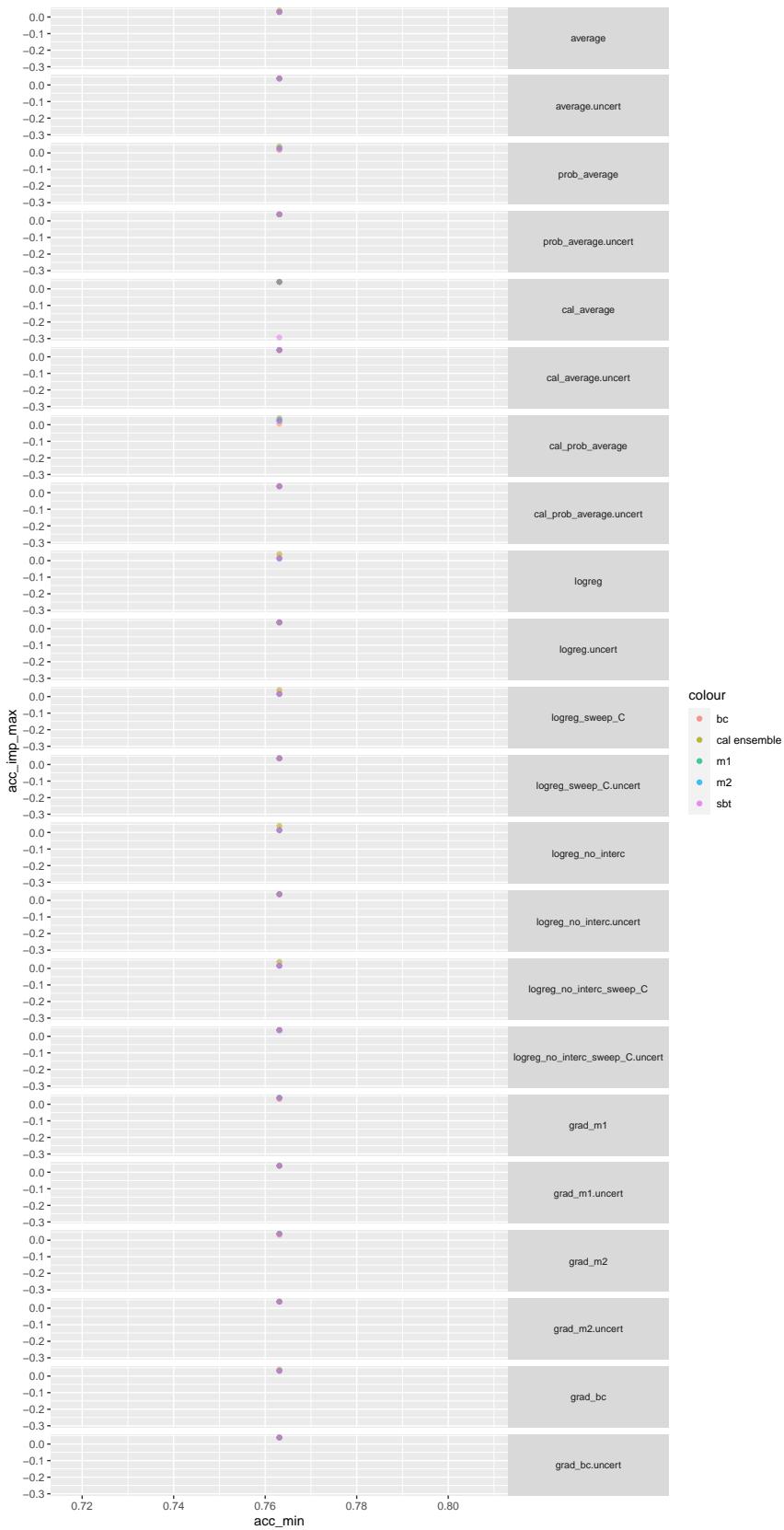
Accuracy improvement of ensemble over
the best of networks vs acc_var.
Ensemble size 6



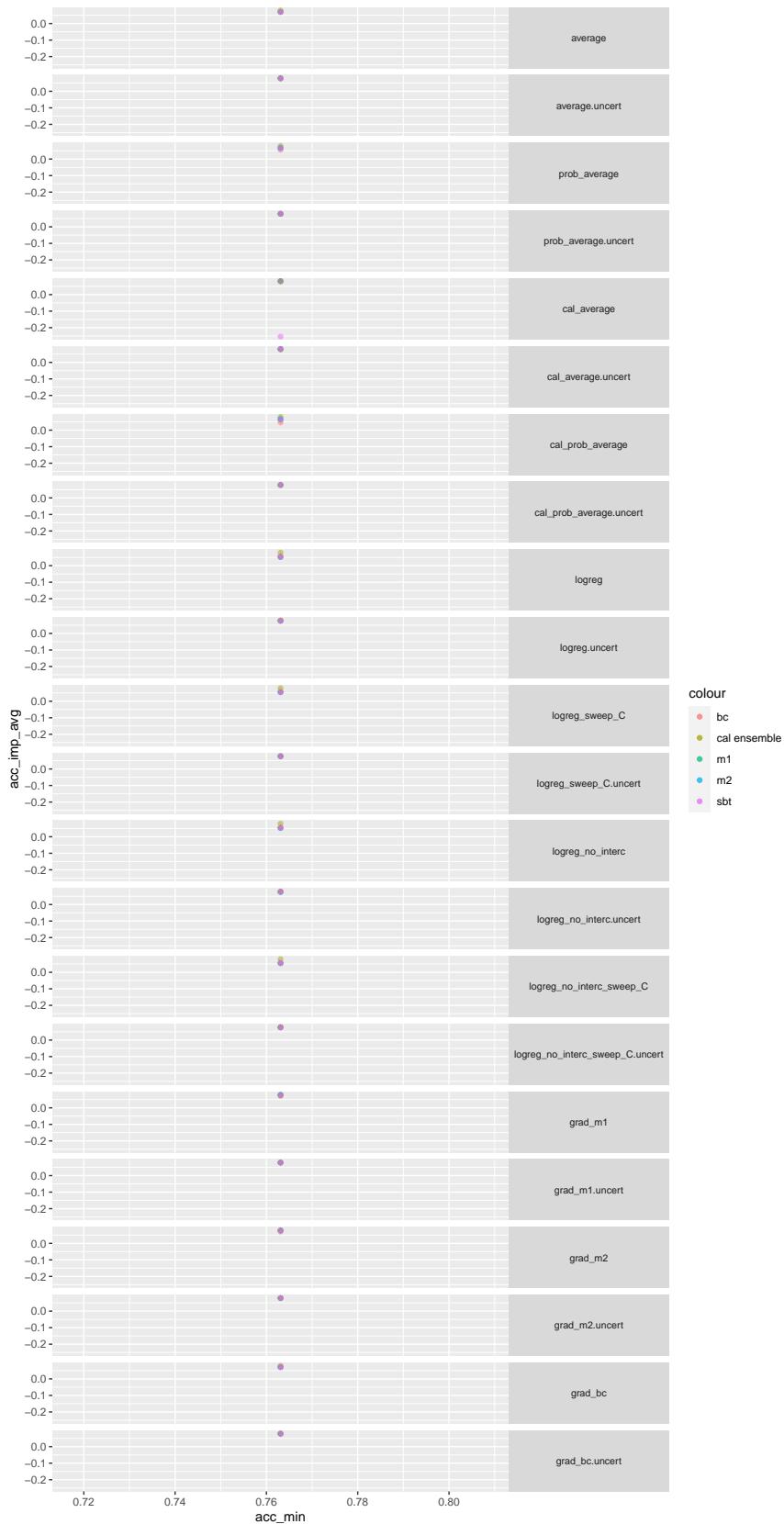
Accuracy improvement of ensemble over
the average of networks vs acc_var.
Ensemble size 6



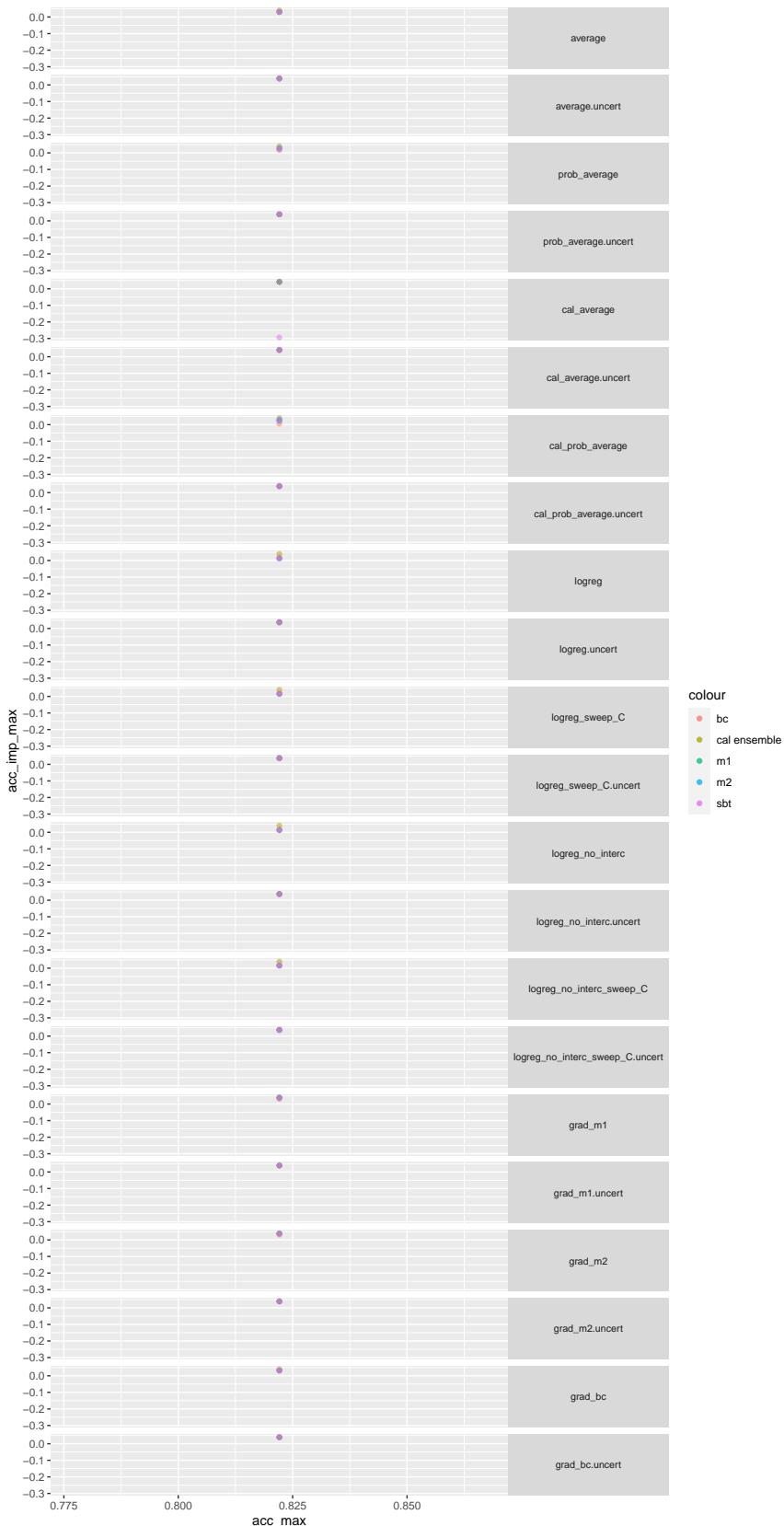
Accuracy improvement of ensemble over
the best of networks vs acc_min.
Ensemble size 6



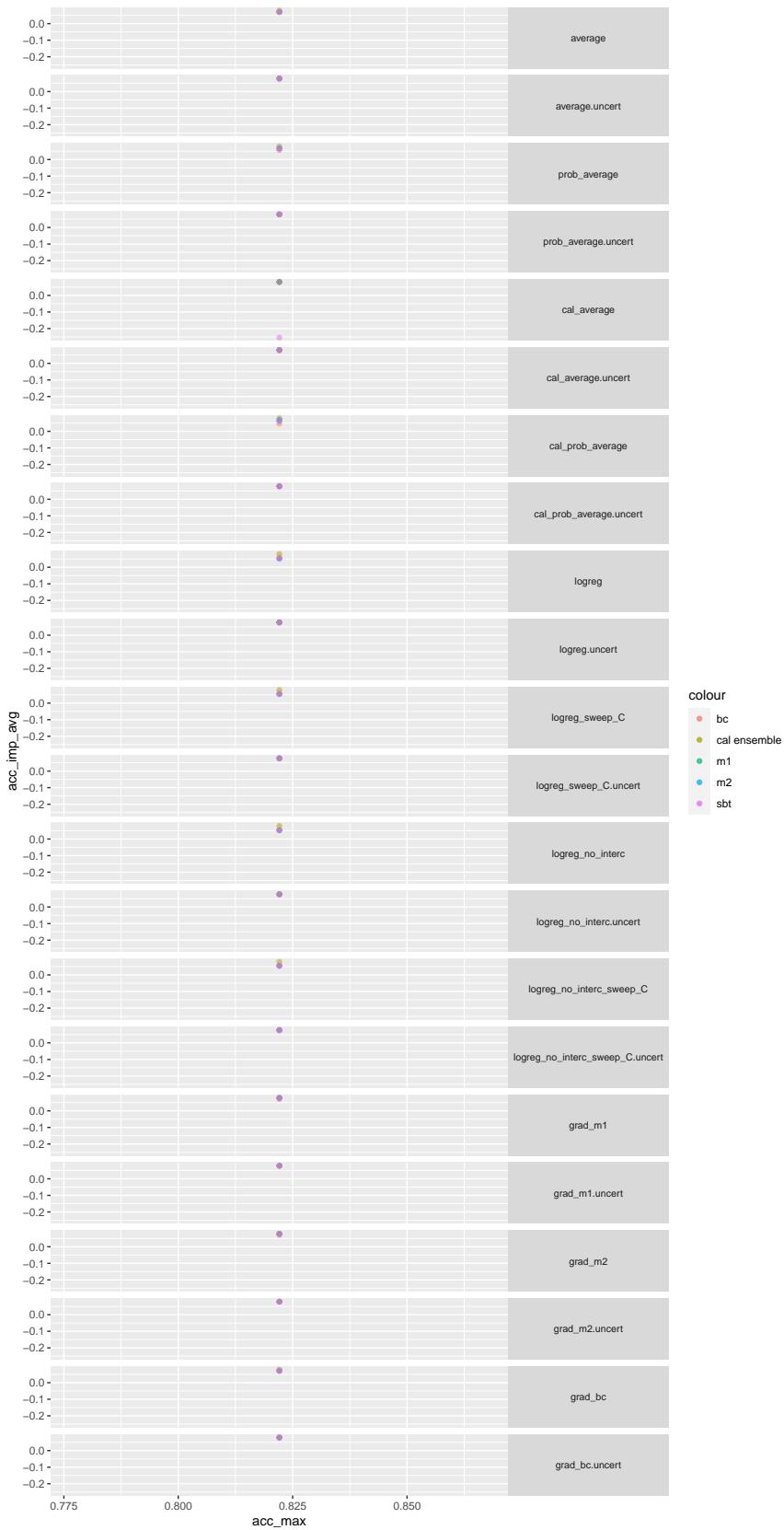
Accuracy improvement of ensemble over
the average of networks vs acc_min.
Ensemble size 6



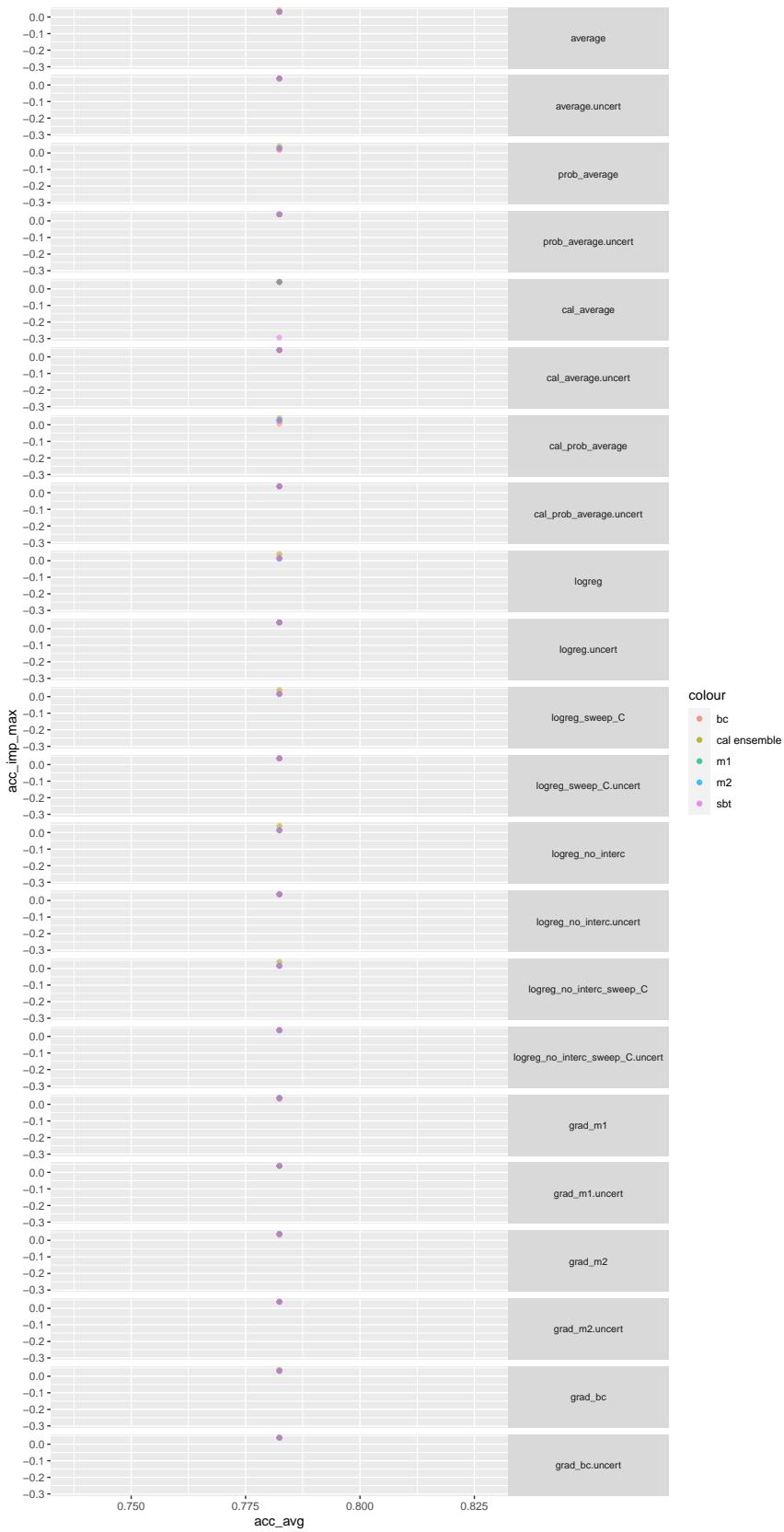
Accuracy improvement of ensemble over
the best of networks vs acc_max.
Ensemble size 6



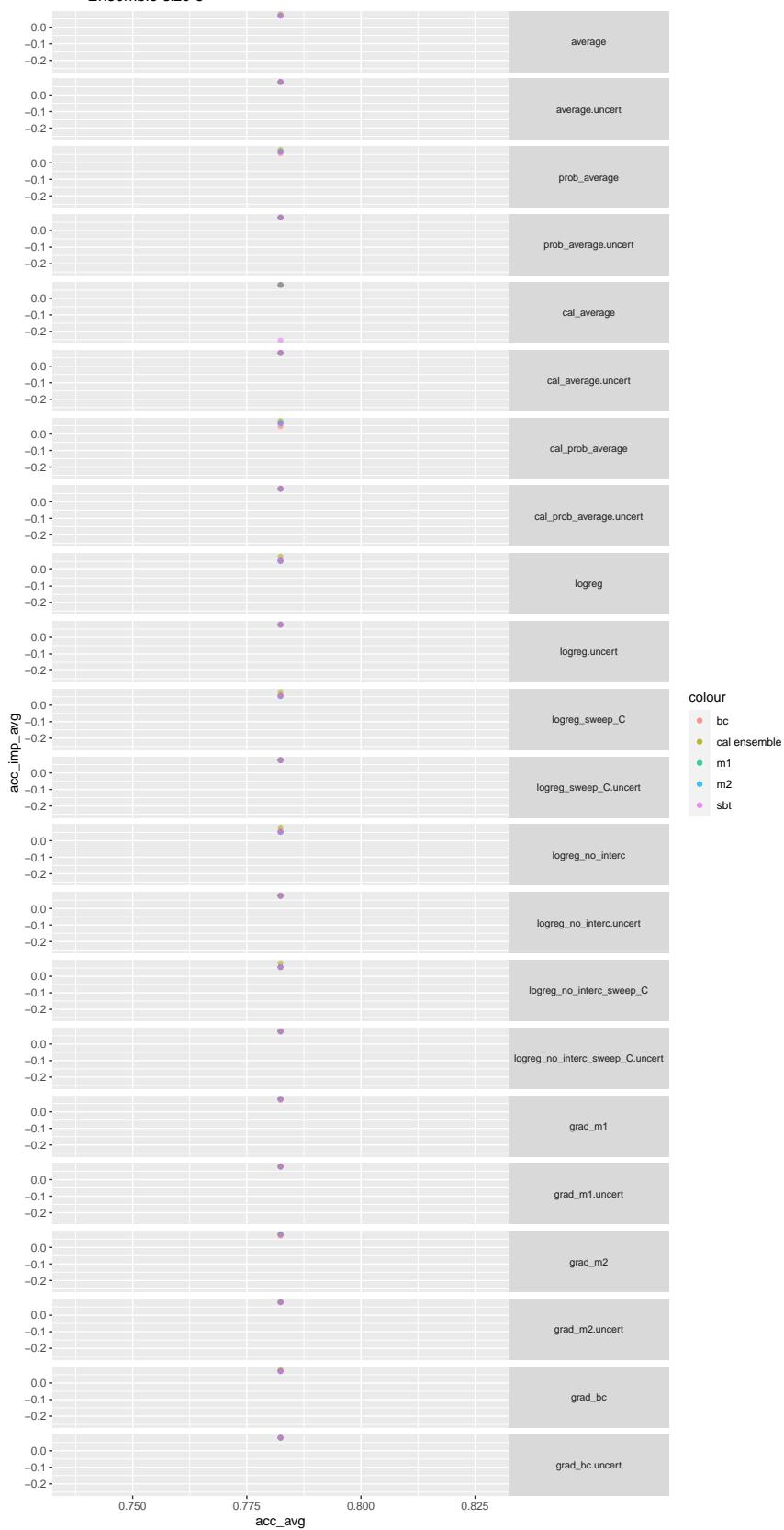
Accuracy improvement of ensemble over
the average of networks vs acc_max.
Ensemble size 6



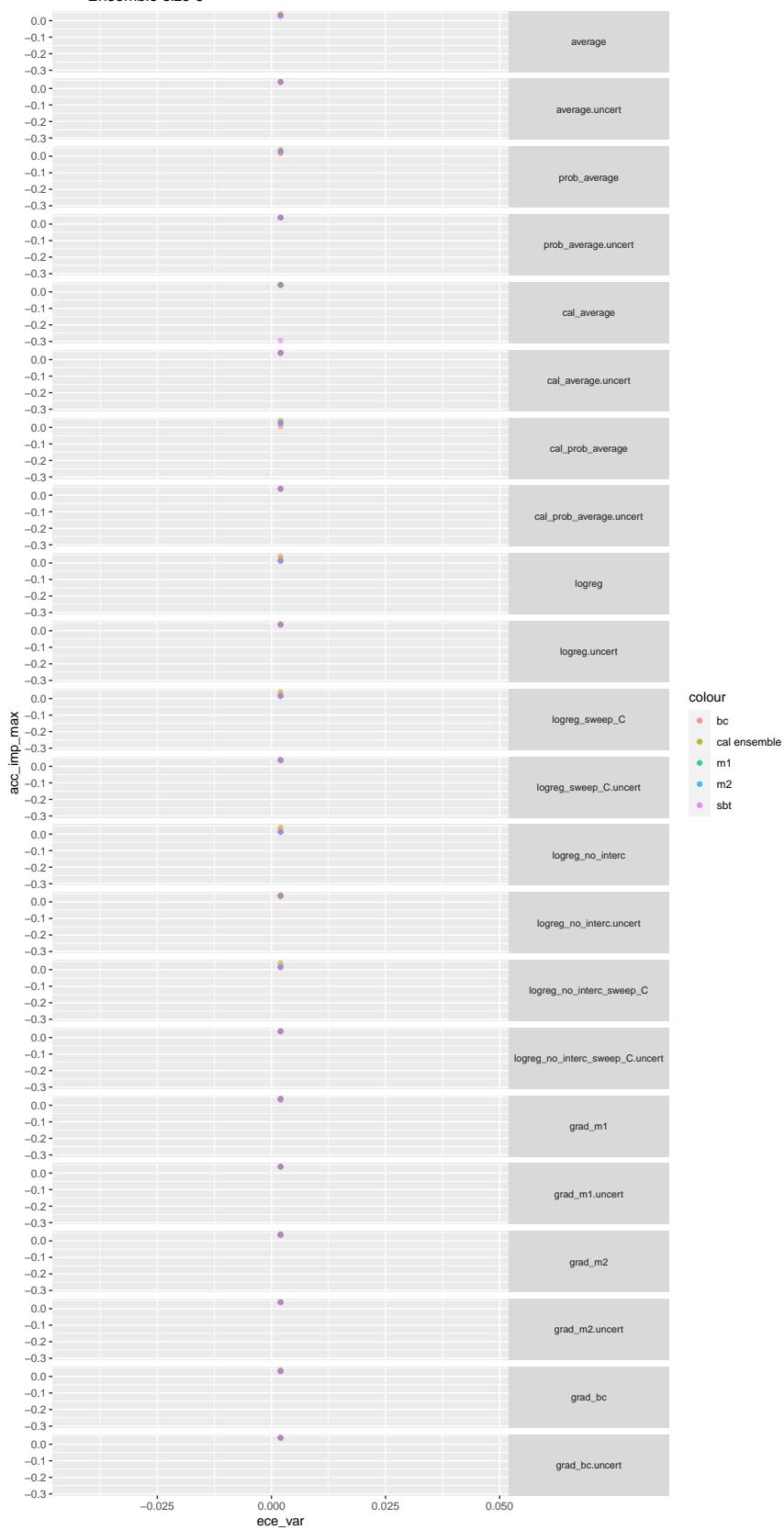
Accuracy improvement of ensemble over
the best of networks vs acc_avg.
Ensemble size 6



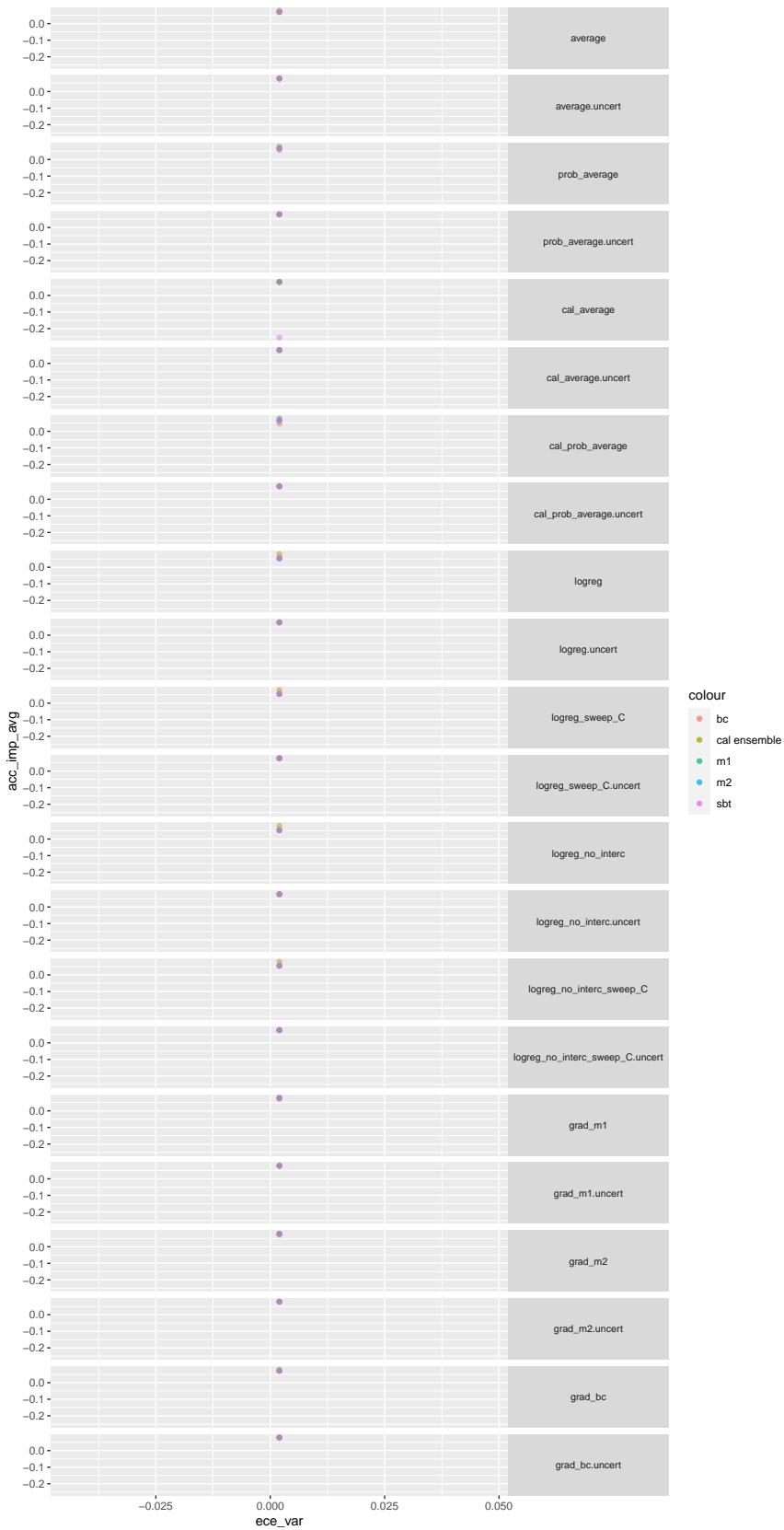
Accuracy improvement of ensemble over
the average of networks vs acc_avg.
Ensemble size 6



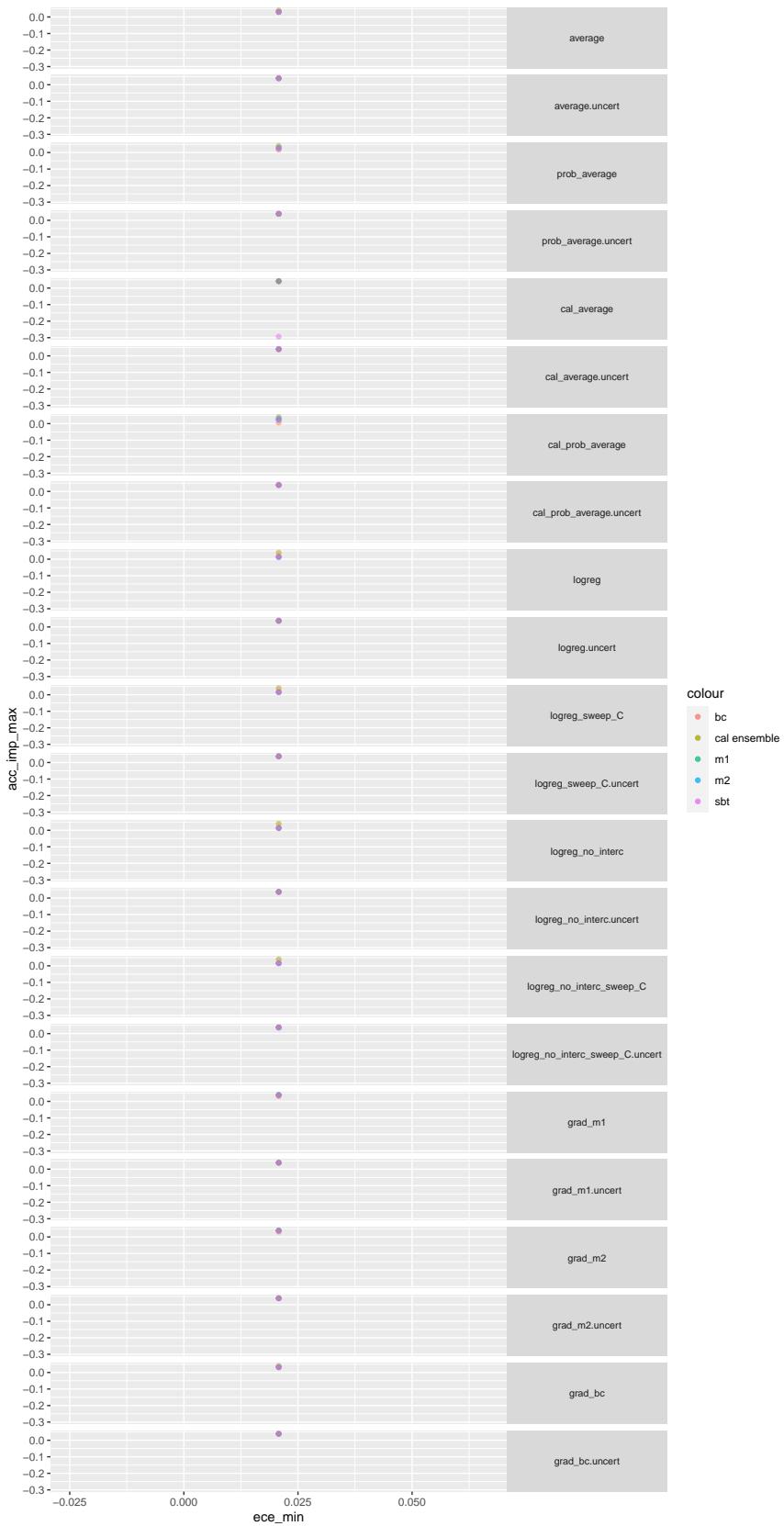
Accuracy improvement of ensemble over
the best of networks vs ece_var.
Ensemble size 6



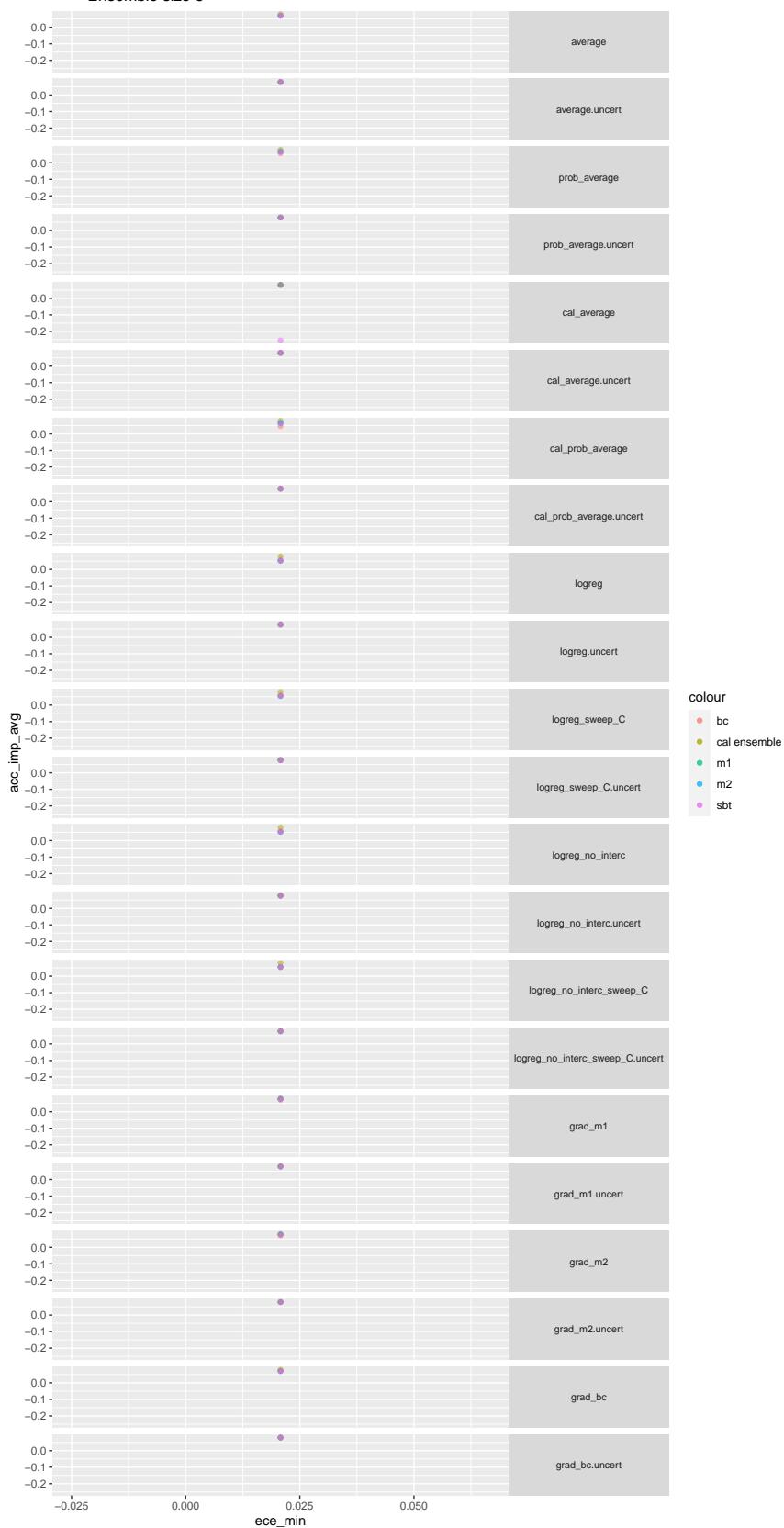
Accuracy improvement of ensemble over
the average of networks vs ece_var.
Ensemble size 6



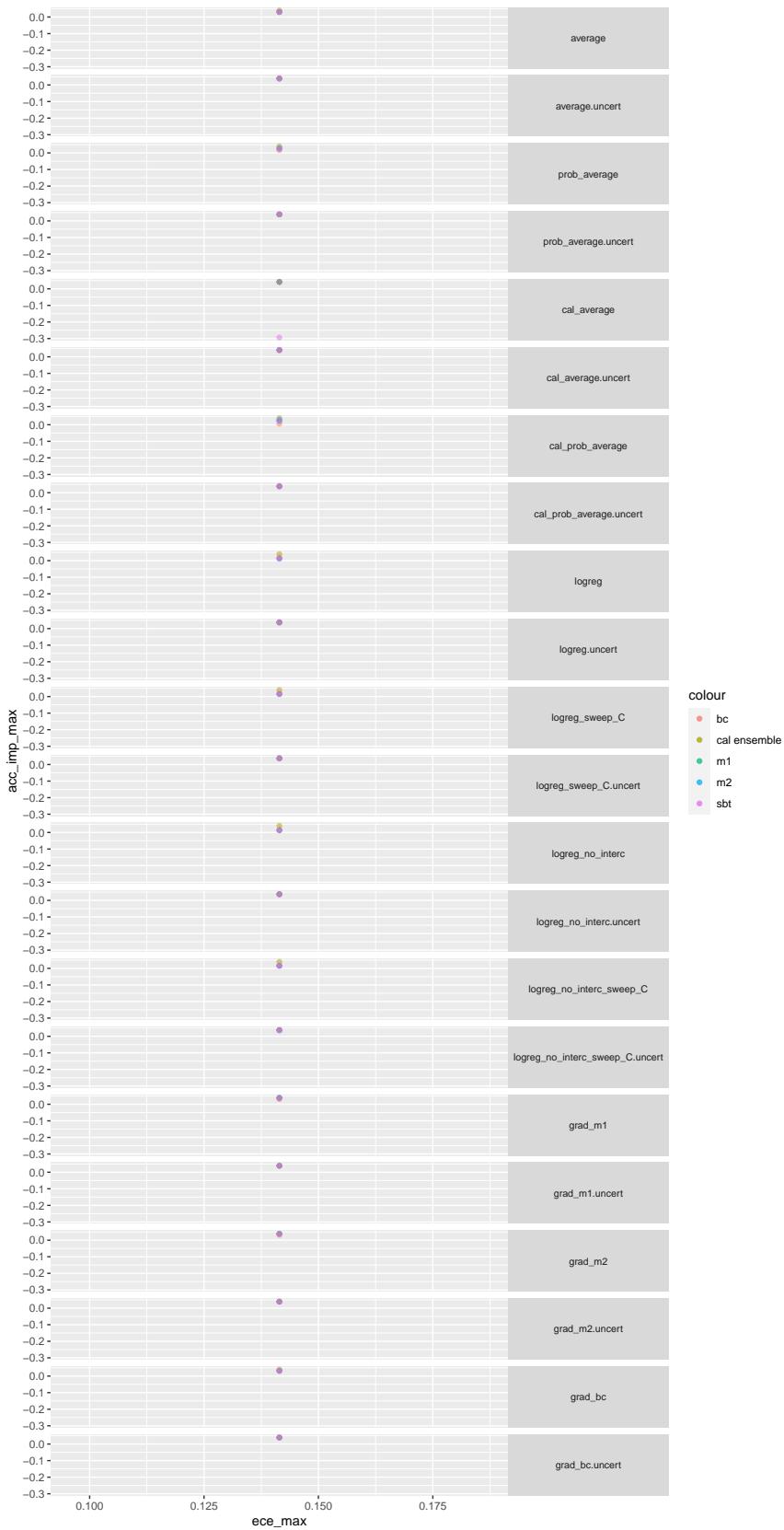
Accuracy improvement of ensemble over
the best of networks vs ece_min.
Ensemble size 6



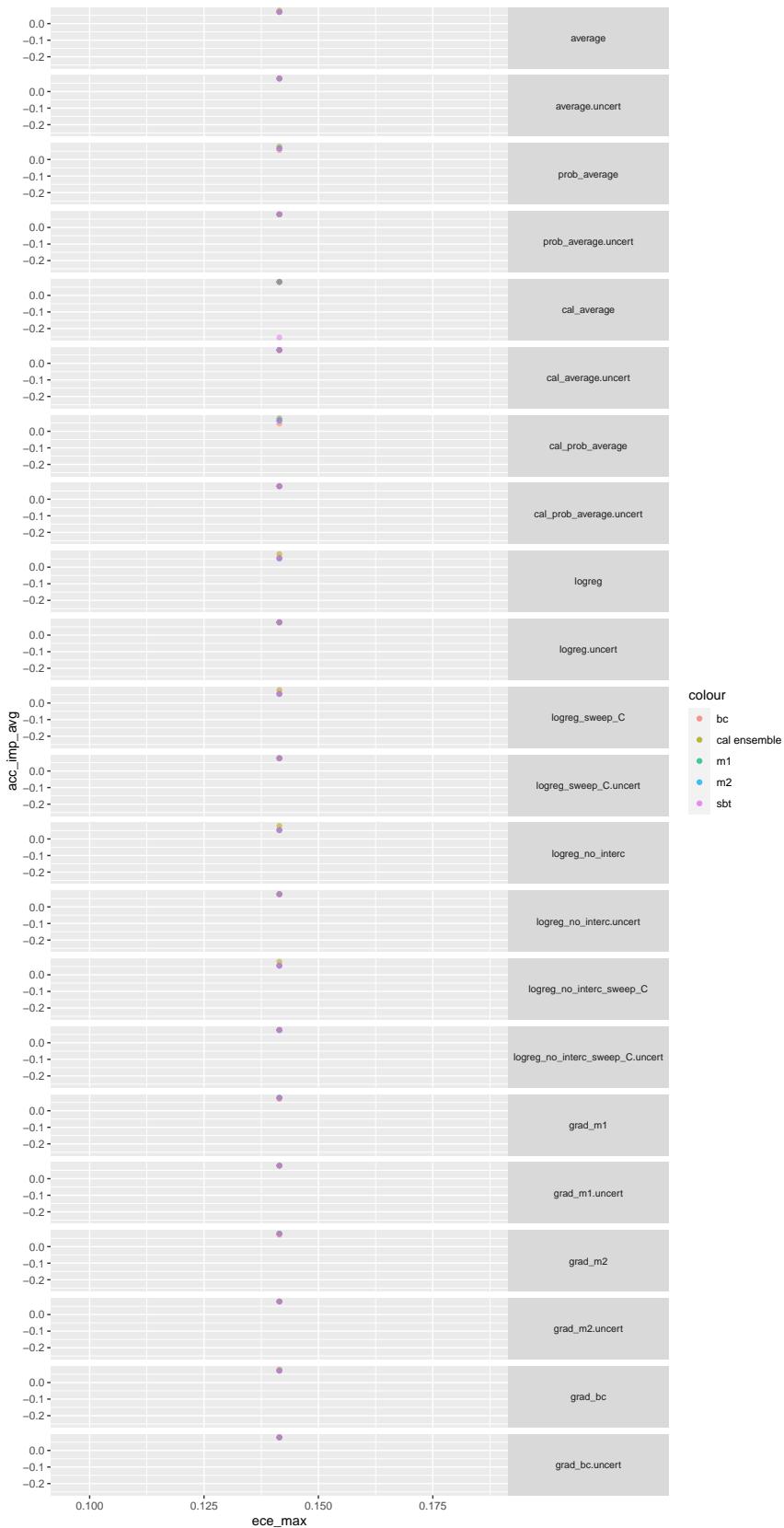
Accuracy improvement of ensemble over
the average of networks vs ece_min.
Ensemble size 6



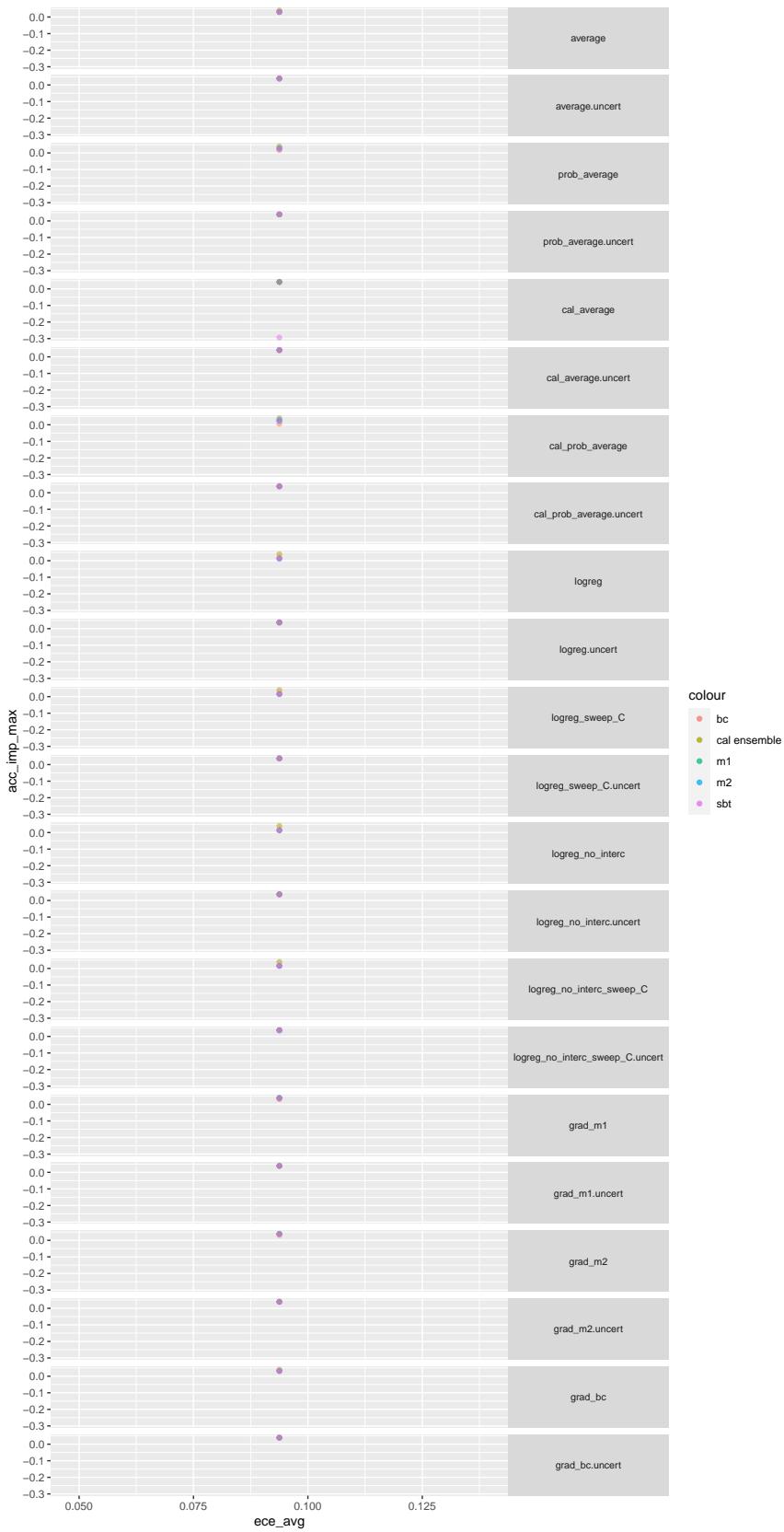
Accuracy improvement of ensemble over
the best of networks vs ece_max.
Ensemble size 6



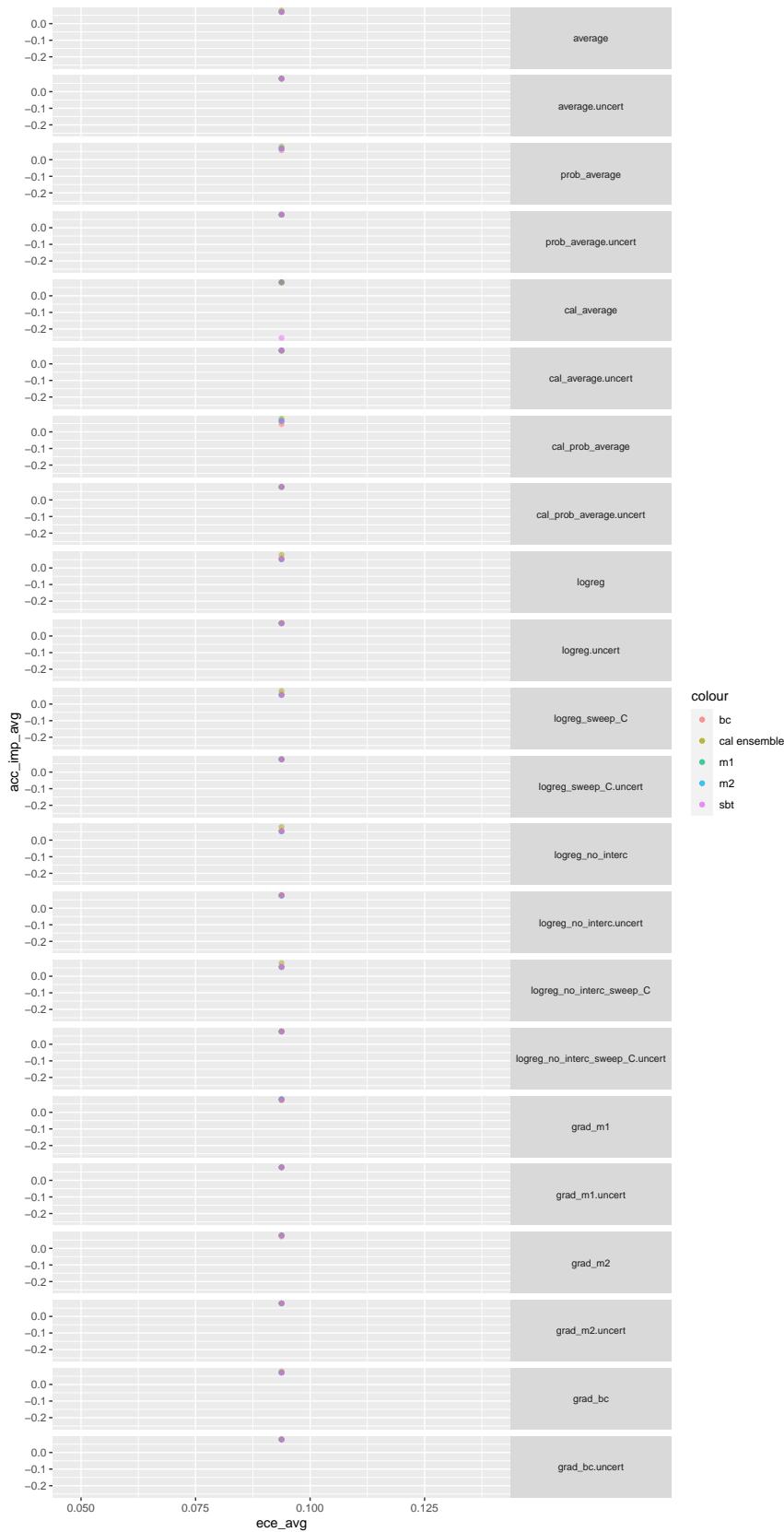
Accuracy improvement of ensemble over
the average of networks vs ece_max.
Ensemble size 6



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



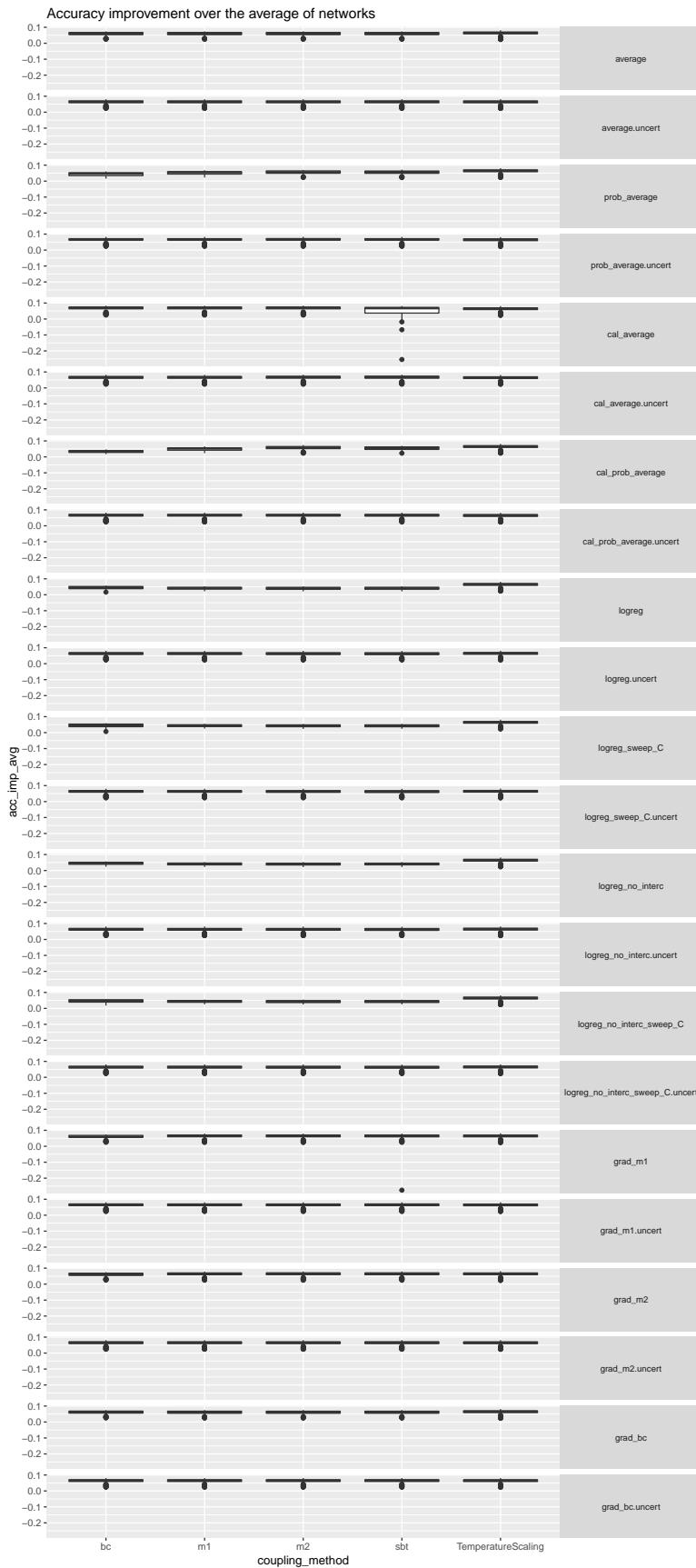
```

avg_imp_table <- rbind(
  ens_pwc_plt_df %>%
    mutate(method = paste(combining_method, coupling_method, sep=" ")) %>%
    group_by(method) %>%
    summarise(imp_o_avg = mean(acc_imp_avg), imp_o_max = mean(acc_imp_max)),
  ens_cal_plt_df %>%
    mutate(method = paste0("average of ", calibrating_method)) %>%
    group_by(method) %>%
    summarise(imp_o_avg = mean(acc_imp_avg), imp_o_max = mean(acc_imp_max))
)

imp_avg_plot <- ggplot() +
  geom_boxplot(data = ens_pwc_plt_df, mapping = aes(x = coupling_method, y = acc_imp_avg)) +
  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") +
  theme(strip.text.y = element_text(size = 8, angle = 0))

print(imp_avg_plot)

```



```
print(xtable(avg_imp_table %>% arrange(desc(imp_o_avg)), digits=c(0, 0, 4, 4)), tabular.environment="longtable")
```

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Tue Jan 11 16:07:29 2022

method	imp_o_avg	imp_o_max
1 cal_average m1	0.0646	0.0407
2 cal_average m2	0.0646	0.0407
3 cal_average bc	0.0646	0.0407
4 cal_average.uncert sbt	0.0629	0.0390
5 cal_average.uncert m2	0.0626	0.0387
6 prob_average.uncert m1	0.0622	0.0383
7 prob_average.uncert bc	0.0622	0.0383
8 prob_average.uncert m2	0.0622	0.0383
9 cal_average.uncert m1	0.0622	0.0383
10 prob_average.uncert sbt	0.0621	0.0382
11 cal_average.uncert bc	0.0620	0.0381
12 cal_prob_average.uncert m2	0.0620	0.0380
13 cal_prob_average.uncert bc	0.0619	0.0380
14 cal_prob_average.uncert m1	0.0619	0.0379
15 cal_prob_average.uncert sbt	0.0618	0.0379
16 grad_m1.uncert bc	0.0613	0.0373
17 grad_bc.uncert bc	0.0613	0.0373
18 average.uncert bc	0.0613	0.0373
19 grad_m2.uncert bc	0.0613	0.0373
20 grad_m2.uncert m1	0.0613	0.0373
21 grad_m1.uncert sbt	0.0612	0.0373
22 grad_m2.uncert sbt	0.0612	0.0373
23 grad_m1.uncert m1	0.0612	0.0373
24 grad_bc.uncert sbt	0.0612	0.0373
25 grad_bc.uncert m1	0.0612	0.0373
26 average.uncert sbt	0.0612	0.0373
27 average.uncert m1	0.0612	0.0373
28 grad_m2.uncert m2	0.0612	0.0373
29 average.uncert m2	0.0612	0.0373
30 grad_m1.uncert m2	0.0612	0.0373
31 grad_bc.uncert m2	0.0612	0.0372
32 grad_m2 m2	0.0612	0.0372
33 average of TemperatureScaling	0.0611	0.0372
34 grad_m2 sbt	0.0610	0.0371
35 grad_m2 m1	0.0608	0.0369
36 grad_m1 m2	0.0608	0.0369
37 grad_m1 m1	0.0608	0.0368
38 logreg.uncert bc	0.0603	0.0364
39 logreg_sweep_C.uncert bc	0.0603	0.0364
40 logreg_no_interc_sweep_C.uncert bc	0.0603	0.0364
41 logreg_no_interc.uncert bc	0.0603	0.0364
42 logreg_no_interc_sweep_C.uncert m1	0.0602	0.0362
43 logreg_no_interc.uncert m1	0.0601	0.0362
44 logreg.uncert m1	0.0601	0.0362
45 logreg_sweep_C.uncert m1	0.0601	0.0362
46 logreg_no_interc_sweep_C.uncert sbt	0.0597	0.0358
47 logreg_no_interc_sweep_C.uncert m2	0.0597	0.0358
48 logreg_sweep_C.uncert m2	0.0597	0.0357

49	logreg.uncert m2	0.0596	0.0356
50	logreg_no_interc.uncert m2	0.0596	0.0356
51	logreg_sweep_C.uncert sbt	0.0596	0.0356
52	logreg_no_interc.uncert sbt	0.0593	0.0353
53	logreg.uncert sbt	0.0592	0.0353
54	grad_bc bc	0.0576	0.0337
55	grad_m2 bc	0.0574	0.0335
56	grad_bc m2	0.0573	0.0333
57	grad_bc m1	0.0573	0.0333
58	grad_m1 bc	0.0572	0.0333
59	grad_bc sbt	0.0572	0.0333
60	average m1	0.0567	0.0327
61	average m2	0.0567	0.0327
62	average sbt	0.0567	0.0327
63	average bc	0.0567	0.0327
64	grad_m1 sbt	0.0551	0.0312
65	cal_prob_average m2	0.0551	0.0311
66	prob_average sbt	0.0532	0.0292
67	prob_average m2	0.0530	0.0291
68	cal_prob_average sbt	0.0517	0.0277
69	prob_average m1	0.0510	0.0271
70	cal_average sbt	0.0498	0.0258
71	cal_prob_average m1	0.0490	0.0250
72	logreg_no_interc_sweep_C bc	0.0452	0.0213
73	logreg_sweep_C bc	0.0441	0.0201
74	logreg_no_interc bc	0.0440	0.0201
75	logreg bc	0.0432	0.0193
76	prob_average bc	0.0431	0.0191
77	logreg_no_interc_sweep_C m1	0.0426	0.0187
78	logreg_sweep_C m1	0.0425	0.0185
79	logreg_no_interc_sweep_C sbt	0.0421	0.0182
80	logreg_no_interc_sweep_C m2	0.0418	0.0179
81	logreg_sweep_C sbt	0.0417	0.0178
82	logreg_sweep_C m2	0.0416	0.0177
83	logreg m1	0.0402	0.0163
84	logreg_no_interc m1	0.0402	0.0162
85	logreg sbt	0.0398	0.0159
86	logreg_no_interc sbt	0.0398	0.0158
87	logreg m2	0.0396	0.0156
88	logreg_no_interc m2	0.0395	0.0156
89	cal_prob_average bc	0.0336	0.0096

```

imp_max_plot <- ggplot() +
  geom_boxplot(data = ens_pwc_plt_df, mapping = aes(x = coupling_method, y = acc_imp_max)) +
  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") +
  theme(strip.text.y = element_text(size = 8, angle = 0))

print(imp_max_plot)

```



```
print(xtable(avg_imp_table %>% arrange(desc(imp_o_max)), digits=c(0, 0, 4, 4)), tabular.environment="longtable")
```

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Tue Jan 11 16:07:31 2022

method	imp_o_avg	imp_o_max
1 cal_average m1	0.0646	0.0407
2 cal_average m2	0.0646	0.0407
3 cal_average bc	0.0646	0.0407
4 cal_average.uncert sbt	0.0629	0.0390
5 cal_average.uncert m2	0.0626	0.0387
6 prob_average.uncert m1	0.0622	0.0383
7 prob_average.uncert bc	0.0622	0.0383
8 prob_average.uncert m2	0.0622	0.0383
9 cal_average.uncert m1	0.0622	0.0383
10 prob_average.uncert sbt	0.0621	0.0382
11 cal_average.uncert bc	0.0620	0.0381
12 cal_prob_average.uncert m2	0.0620	0.0380
13 cal_prob_average.uncert bc	0.0619	0.0380
14 cal_prob_average.uncert m1	0.0619	0.0379
15 cal_prob_average.uncert sbt	0.0618	0.0379
16 grad_m1.uncert bc	0.0613	0.0373
17 grad_bc.uncert bc	0.0613	0.0373
18 average.uncert bc	0.0613	0.0373
19 grad_m2.uncert bc	0.0613	0.0373
20 grad_m2.uncert m1	0.0613	0.0373
21 grad_m1.uncert sbt	0.0612	0.0373
22 grad_m2.uncert sbt	0.0612	0.0373
23 grad_m1.uncert m1	0.0612	0.0373
24 grad_bc.uncert sbt	0.0612	0.0373
25 grad_bc.uncert m1	0.0612	0.0373
26 average.uncert sbt	0.0612	0.0373
27 average.uncert m1	0.0612	0.0373
28 grad_m2.uncert m2	0.0612	0.0373
29 average.uncert m2	0.0612	0.0373
30 grad_m1.uncert m2	0.0612	0.0373
31 grad_bc.uncert m2	0.0612	0.0372
32 grad_m2 m2	0.0612	0.0372
33 average of TemperatureScaling	0.0611	0.0372
34 grad_m2 sbt	0.0610	0.0371
35 grad_m2 m1	0.0608	0.0369
36 grad_m1 m2	0.0608	0.0369
37 grad_m1 m1	0.0608	0.0368
38 logreg.uncert bc	0.0603	0.0364
39 logreg_sweep_C.uncert bc	0.0603	0.0364
40 logreg_no_interc_sweep_C.uncert bc	0.0603	0.0364
41 logreg_no_interc.uncert bc	0.0603	0.0364
42 logreg_no_interc_sweep_C.uncert m1	0.0602	0.0362
43 logreg_no_interc.uncert m1	0.0601	0.0362
44 logreg.uncert m1	0.0601	0.0362
45 logreg_sweep_C.uncert m1	0.0601	0.0362
46 logreg_no_interc_sweep_C.uncert sbt	0.0597	0.0358
47 logreg_no_interc_sweep_C.uncert m2	0.0597	0.0358
48 logreg_sweep_C.uncert m2	0.0597	0.0357

49	logreg.uncert m2	0.0596	0.0356
50	logreg_no_interc.uncert m2	0.0596	0.0356
51	logreg_sweep_C.uncert sbt	0.0596	0.0356
52	logreg_no_interc.uncert sbt	0.0593	0.0353
53	logreg.uncert sbt	0.0592	0.0353
54	grad_bc bc	0.0576	0.0337
55	grad_m2 bc	0.0574	0.0335
56	grad_bc m2	0.0573	0.0333
57	grad_bc m1	0.0573	0.0333
58	grad_m1 bc	0.0572	0.0333
59	grad_bc sbt	0.0572	0.0333
60	average m1	0.0567	0.0327
61	average m2	0.0567	0.0327
62	average sbt	0.0567	0.0327
63	average bc	0.0567	0.0327
64	grad_m1 sbt	0.0551	0.0312
65	cal_prob_average m2	0.0551	0.0311
66	prob_average sbt	0.0532	0.0292
67	prob_average m2	0.0530	0.0291
68	cal_prob_average sbt	0.0517	0.0277
69	prob_average m1	0.0510	0.0271
70	cal_average sbt	0.0498	0.0258
71	cal_prob_average m1	0.0490	0.0250
72	logreg_no_interc_sweep_C bc	0.0452	0.0213
73	logreg_sweep_C bc	0.0441	0.0201
74	logreg_no_interc bc	0.0440	0.0201
75	logreg bc	0.0432	0.0193
76	prob_average bc	0.0431	0.0191
77	logreg_no_interc_sweep_C m1	0.0426	0.0187
78	logreg_sweep_C m1	0.0425	0.0185
79	logreg_no_interc_sweep_C sbt	0.0421	0.0182
80	logreg_no_interc_sweep_C m2	0.0418	0.0179
81	logreg_sweep_C sbt	0.0417	0.0178
82	logreg_sweep_C m2	0.0416	0.0177
83	logreg m1	0.0402	0.0163
84	logreg_no_interc m1	0.0402	0.0162
85	logreg sbt	0.0398	0.0159
86	logreg_no_interc sbt	0.0398	0.0158
87	logreg m2	0.0396	0.0156
88	logreg_no_interc m2	0.0395	0.0156
89	cal_prob_average bc	0.0336	0.0096

```

avg_imp_table_cs <- rbind(
  ens_pwc_plt_df %>%
    mutate(method = paste(combining_method, coupling_method, sep=" ")) %>%
    group_by(method, combination_size) %>%
    summarise(imp_o_avg = mean(acc_imp_avg), imp_o_max = mean(acc_imp_max)),
  ens_cal_plt_df %>%
    mutate(method = paste0("average of ", calibrating_method)) %>%
    group_by(method, combination_size) %>%
    summarise(imp_o_avg = mean(acc_imp_avg), imp_o_max = mean(acc_imp_max))
)

```

```

## `summarise()` has grouped output by 'method'. You can override using the '.groups' argument.
## `summarise()` has grouped output by 'method'. You can override using the '.groups' argument.

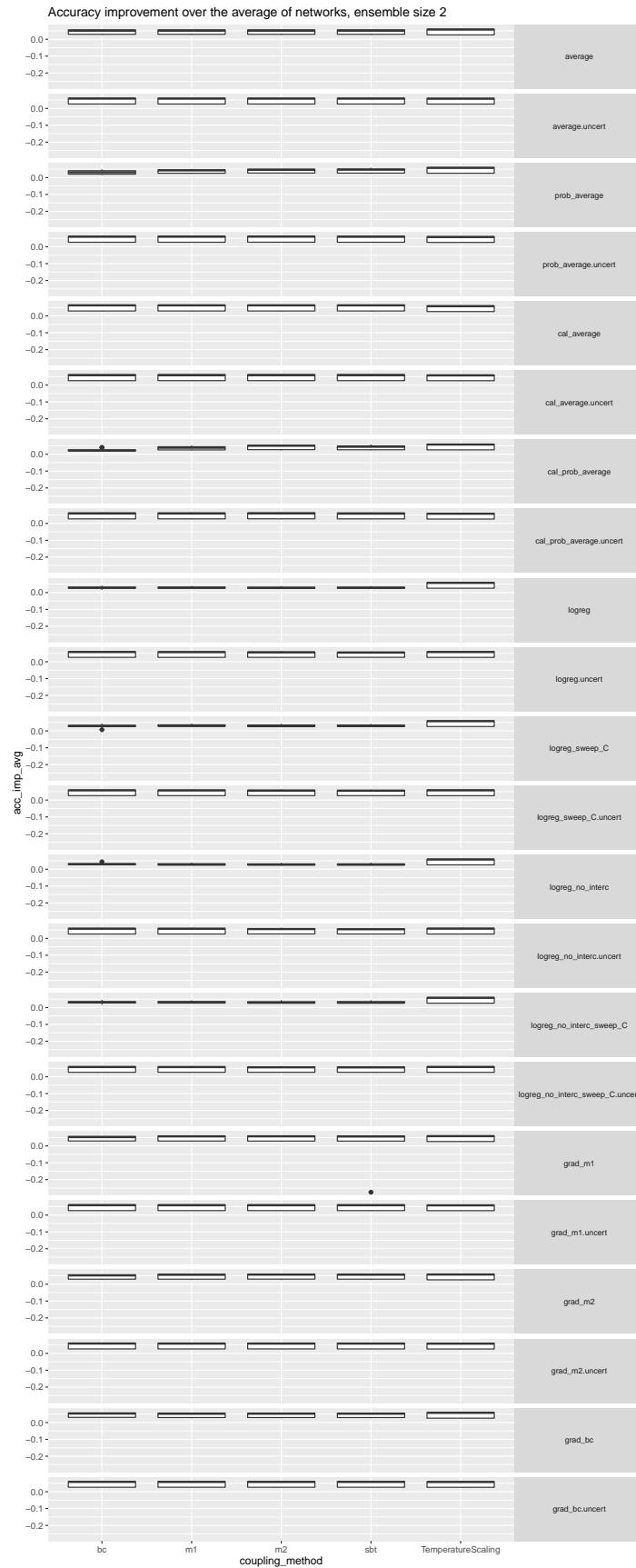
for (sss in unique(ens_cal_plt_df$combination_size))
{
  imp_avg_plot <- ggplot() +
    geom_boxplot(data = ens_pwc_plt_df %>% filter(combination_size == sss), mapping = aes(x = coupling_n,
    geom_boxplot(data = ens_cal_plt_df %>% filter(combination_size == sss), mapping = aes(x = calibrati
    facet_grid(rows = vars(combining_method)) +
    ggtitle(sprintf("Accuracy improvement over the average of networks, ensemble size %s", sss)) +
    theme(strip.text.y = element_text(size = 8, angle = 0))

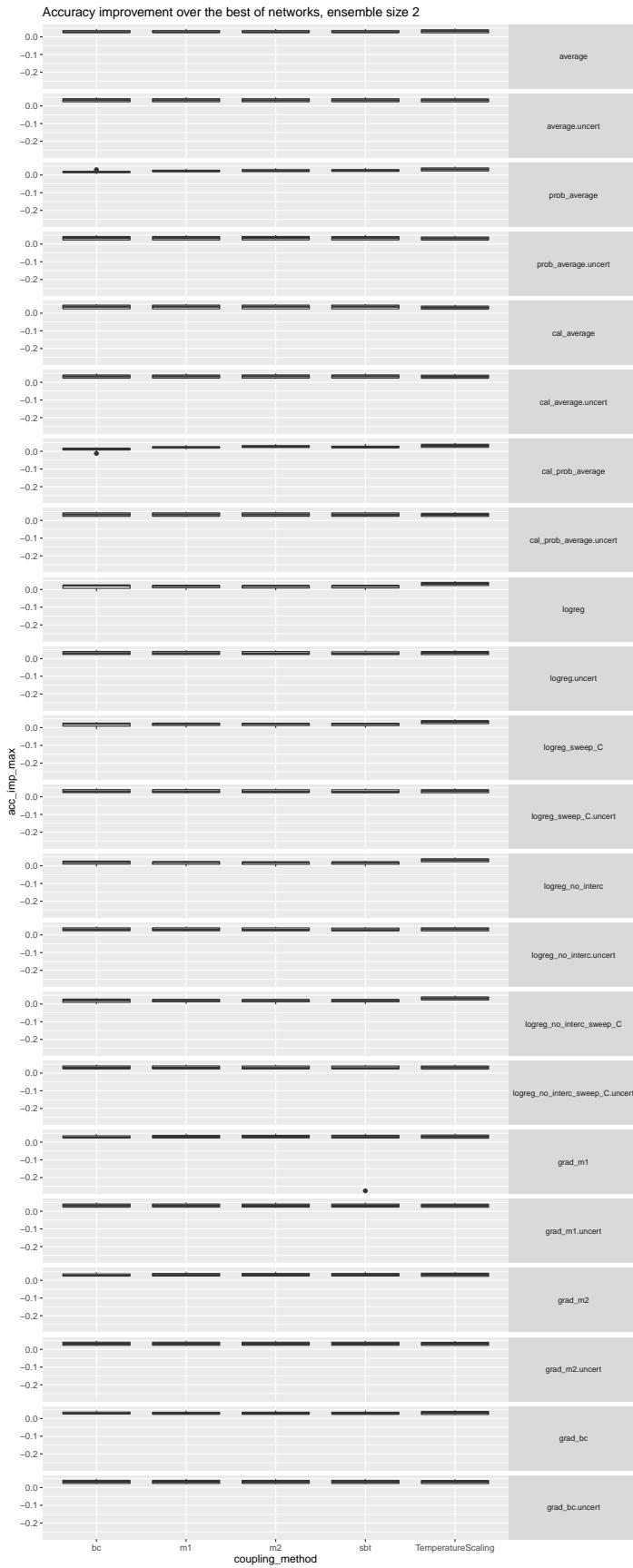
  print(imp_avg_plot)

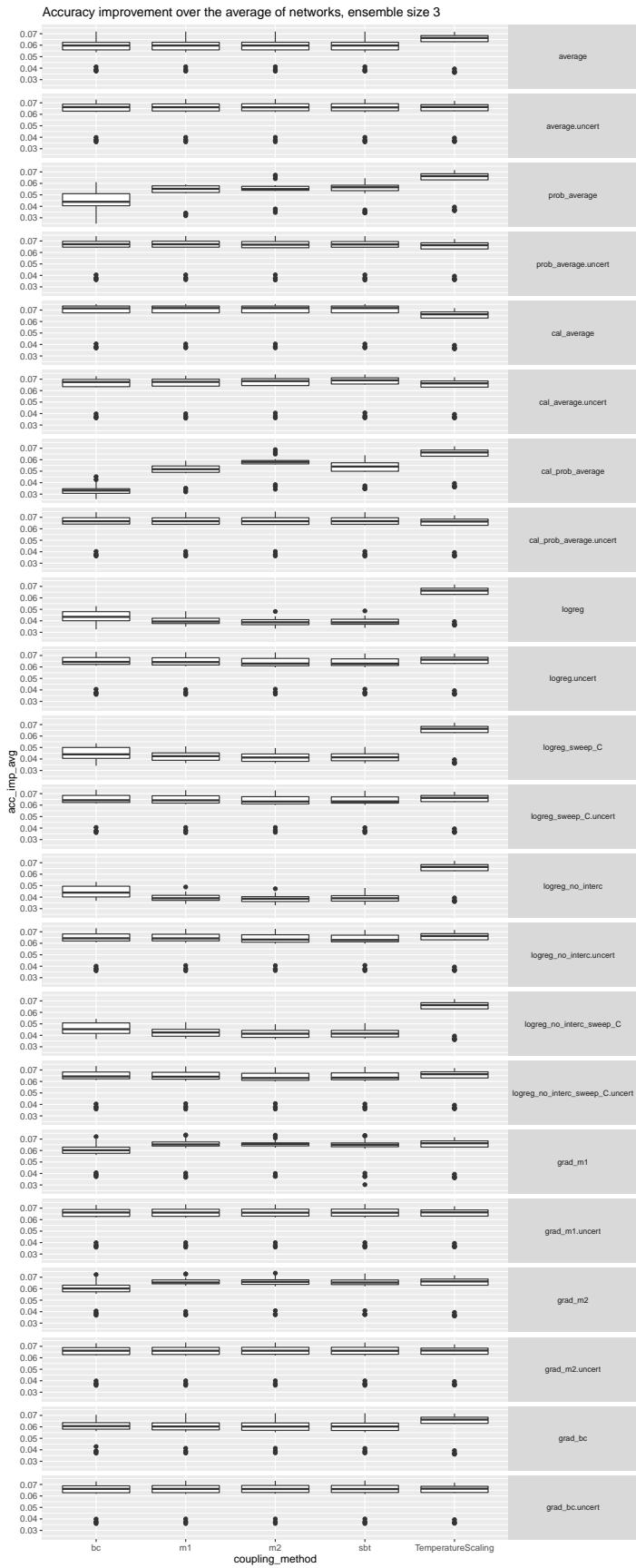
  imp_max_plot <- ggplot() +
    geom_boxplot(data = ens_pwc_plt_df %>% filter(combination_size == sss), mapping = aes(x = coupling_n,
    geom_boxplot(data = ens_cal_plt_df %>% filter(combination_size == sss), mapping = aes(x = calibrati
    facet_grid(rows = vars(combining_method)) +
    ggtitle(sprintf("Accuracy improvement over the best of networks, ensemble size %s", sss)) +
    theme(strip.text.y = element_text(size = 8, angle = 0))

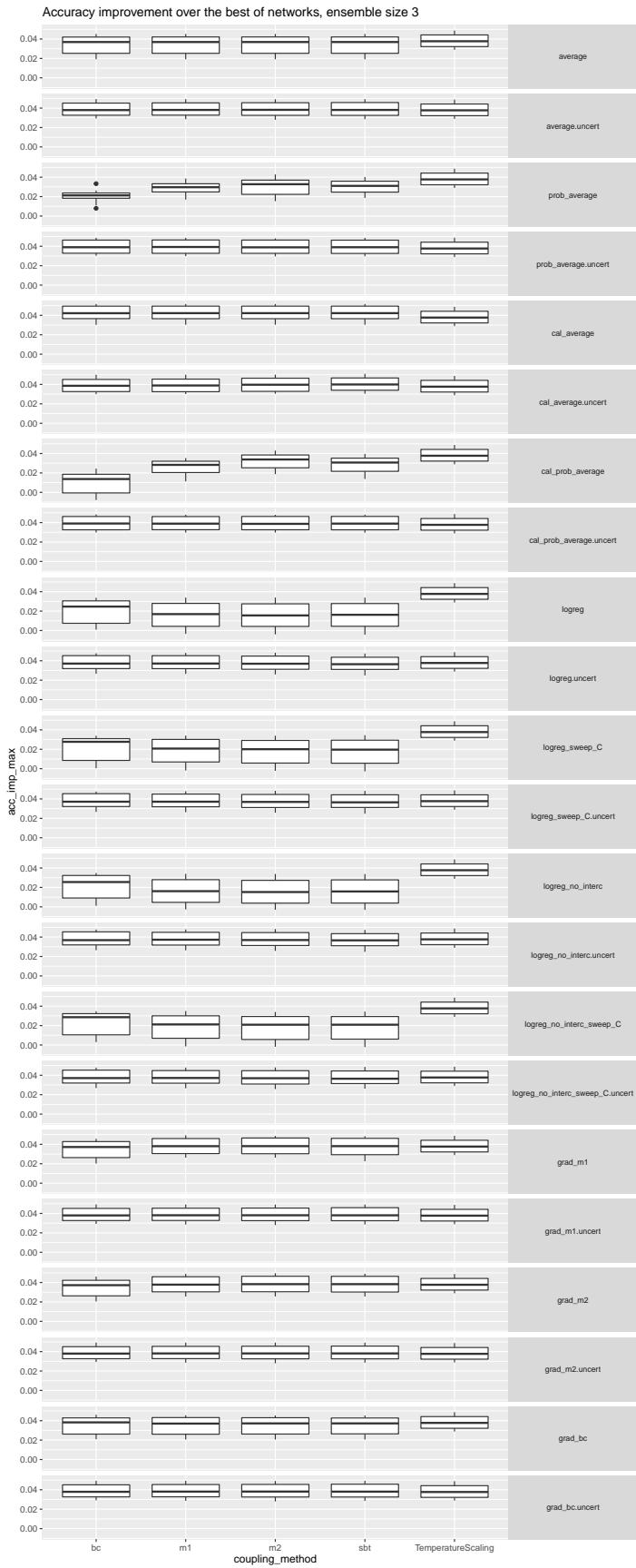
  print(imp_max_plot)
}

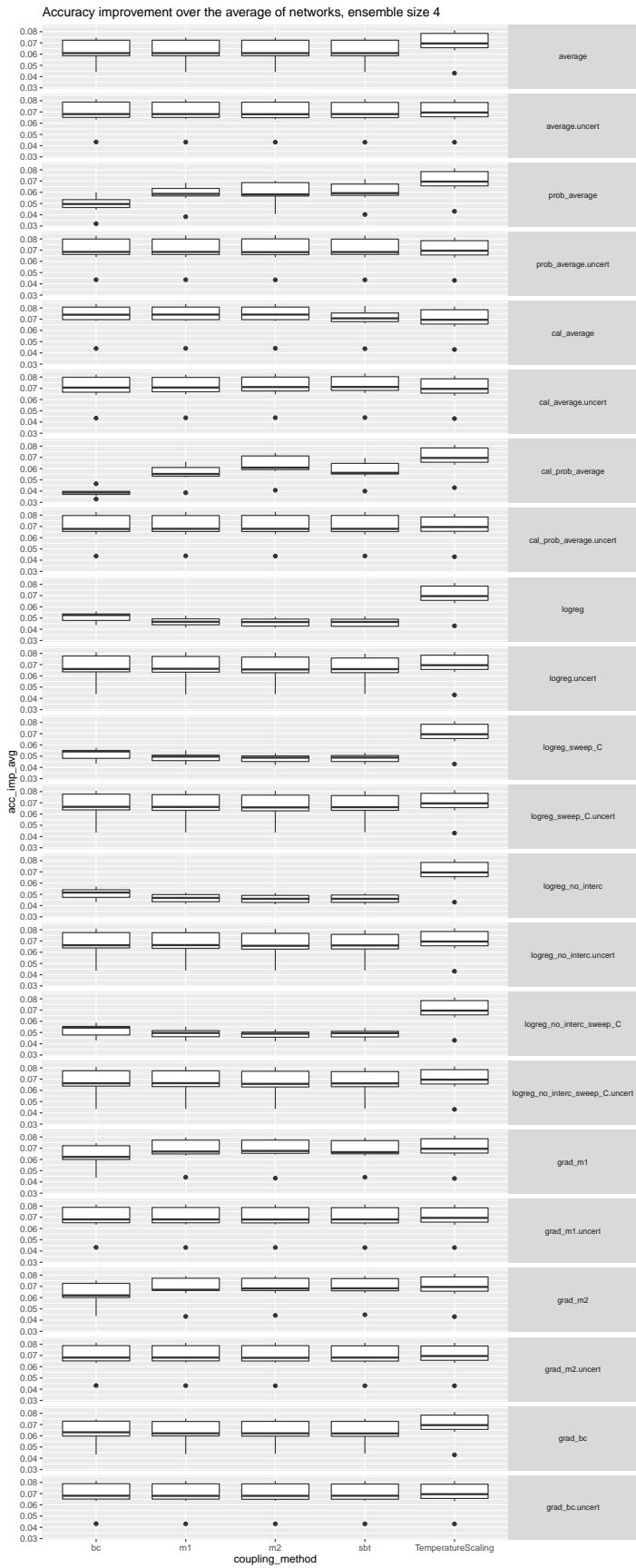
```

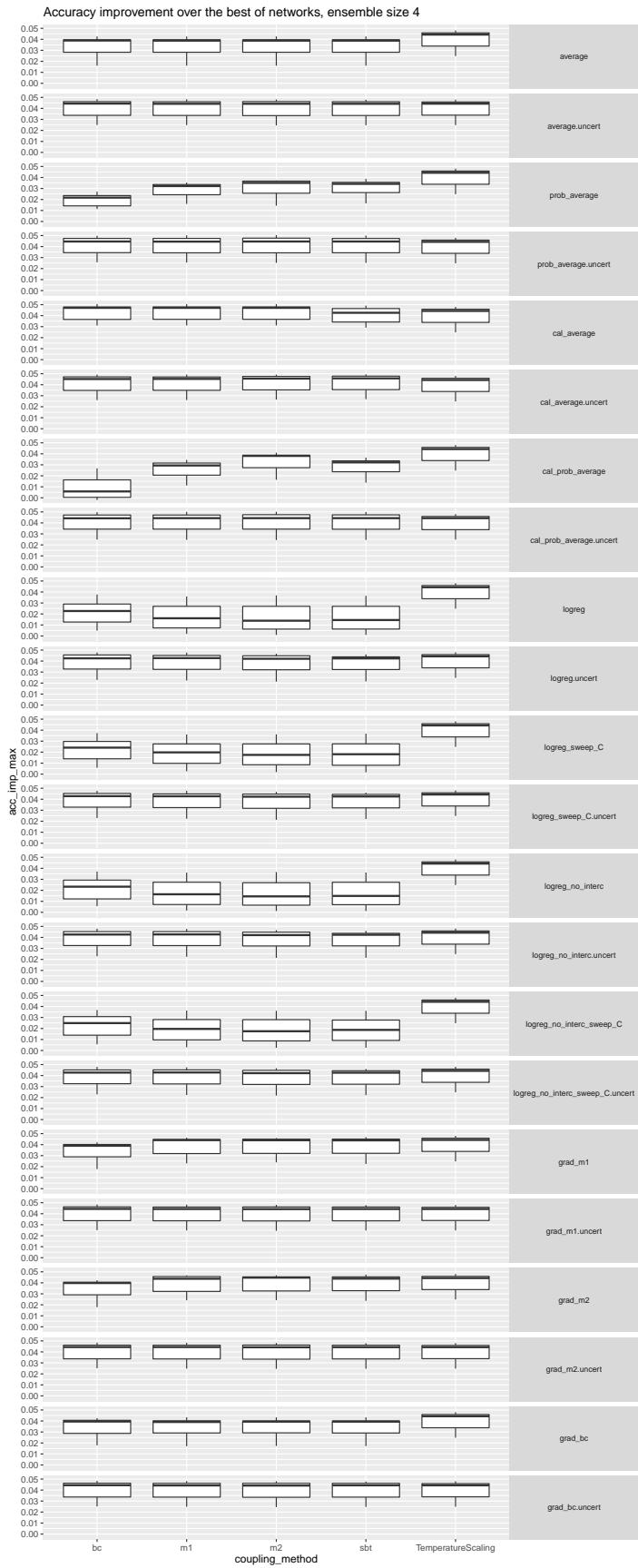


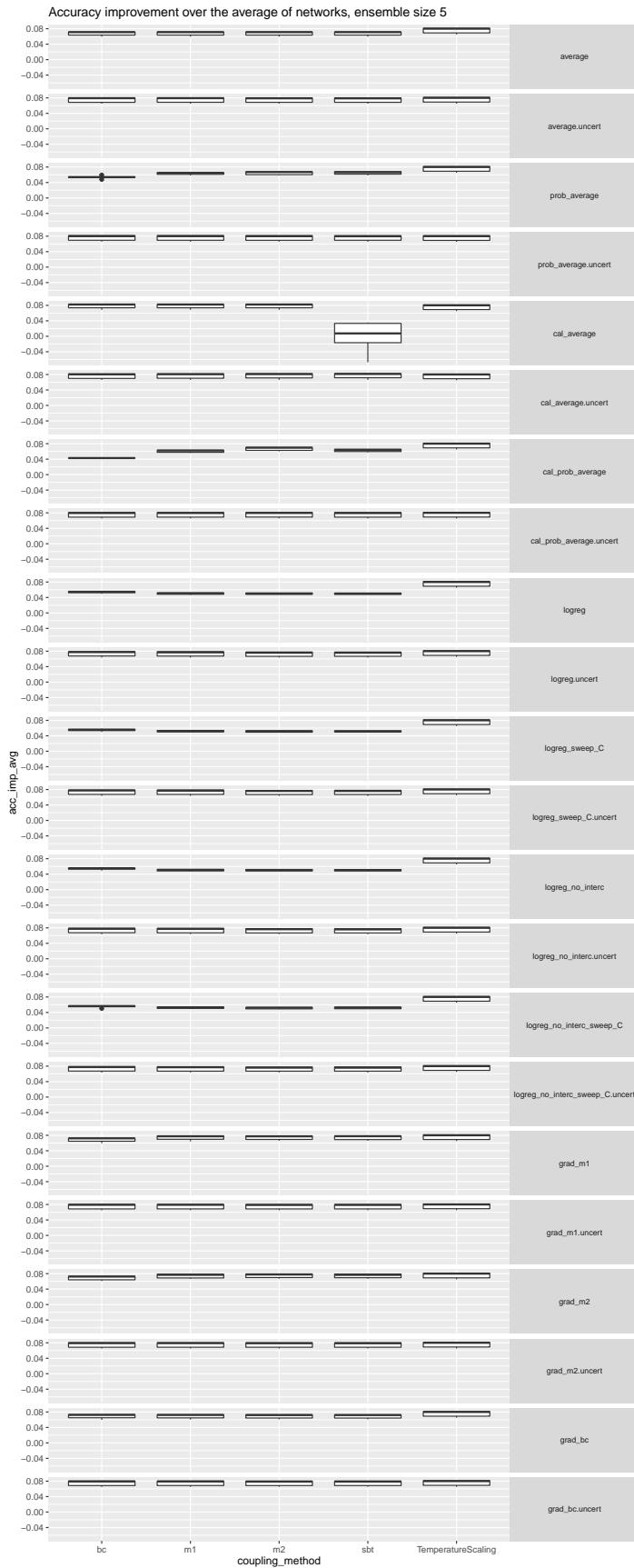


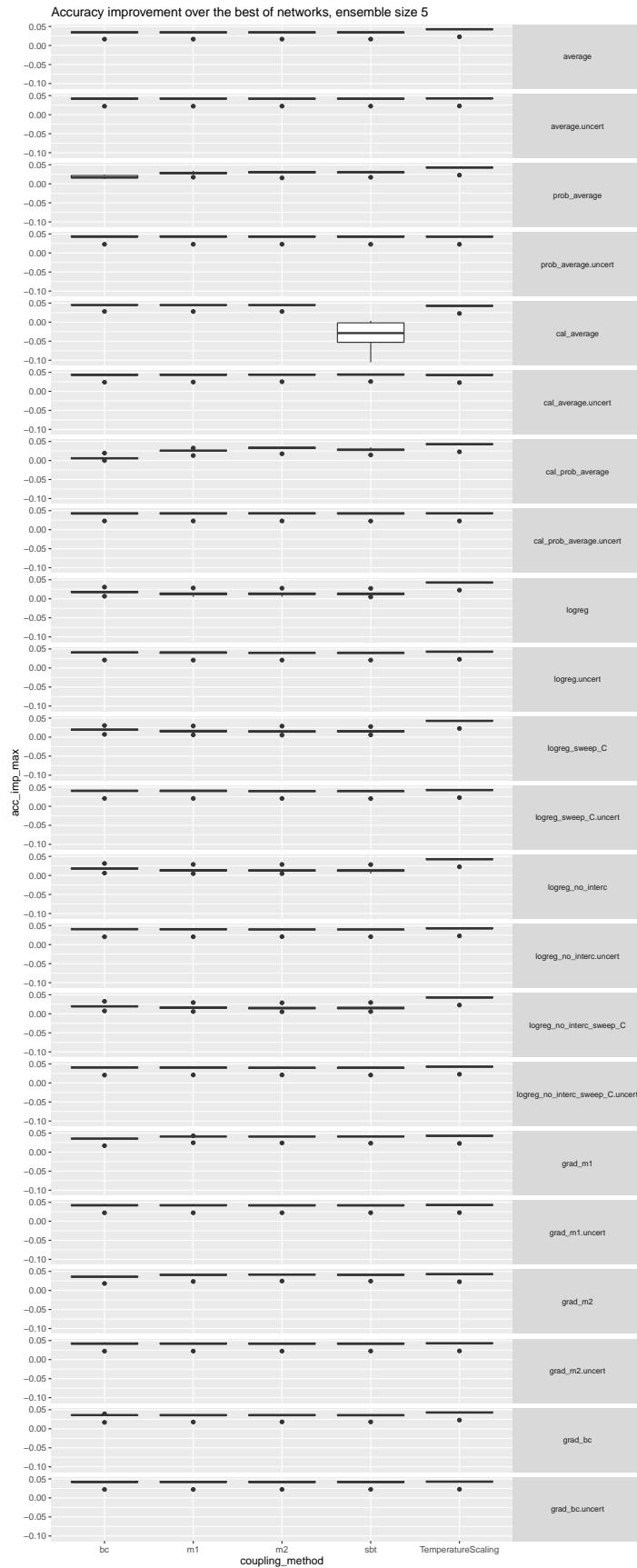


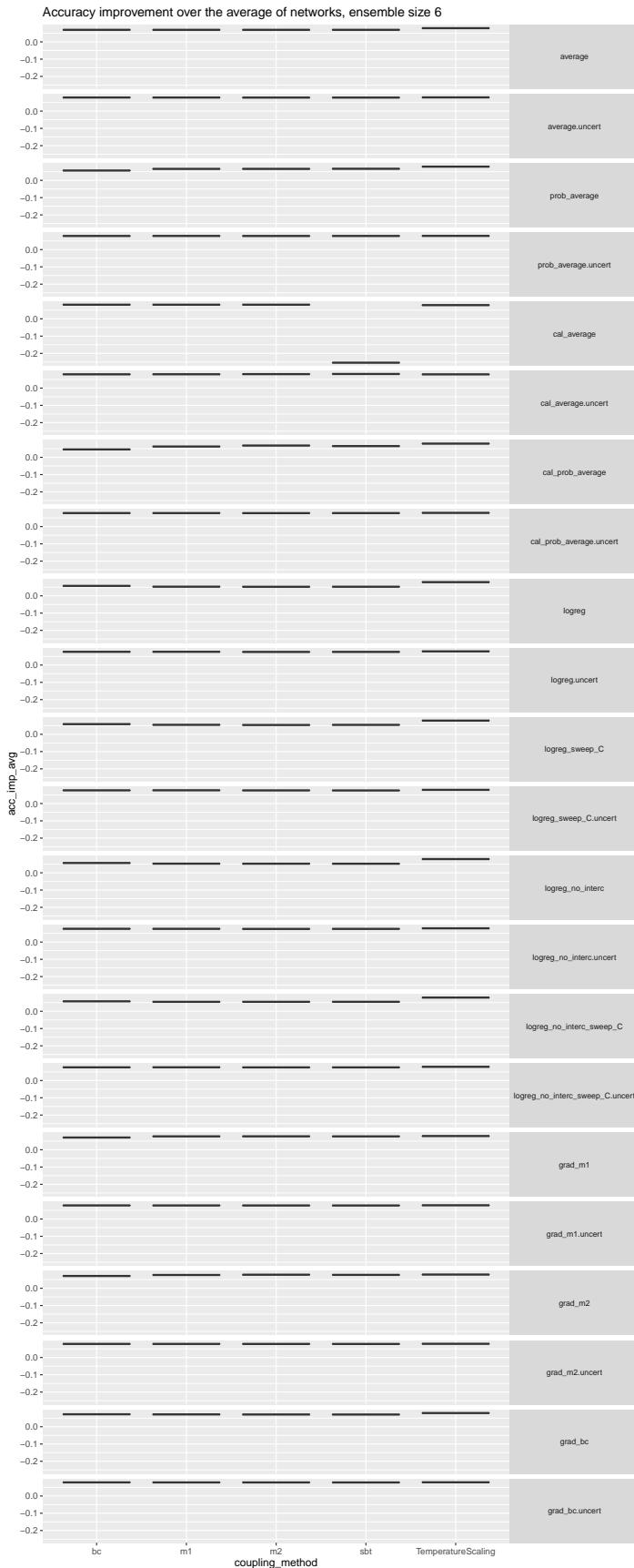














```

for (sss in unique(ens_cal_plt_df$combination_size))
{
  print(xtable(avg_imp_table_cs %>% filter(combination_size == sss) %>% arrange(desc(imp_o_avg)), digits=3))
  print(xtable(avg_imp_table_cs %>% filter(combination_size == sss) %>% arrange(desc(imp_o_max)), digits=3))
}

```

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Tue Jan 11 16:07:49 2022

method	combination_size	imp_o_avg	imp_o_max
1 cal_average m1	2	0.0485	0.0351
2 cal_average m2	2	0.0485	0.0351
3 cal_average sbt	2	0.0485	0.0351
4 cal_average bc	2	0.0484	0.0351
5 prob_average.uncert m2	2	0.0469	0.0335
6 prob_average.uncert m1	2	0.0468	0.0334
7 cal_prob_average.uncert m2	2	0.0467	0.0333
8 prob_average.uncert bc	2	0.0467	0.0333
9 prob_average.uncert sbt	2	0.0466	0.0333
10 cal_prob_average.uncert m1	2	0.0465	0.0331
11 cal_prob_average.uncert bc	2	0.0464	0.0331
12 cal_prob_average.uncert sbt	2	0.0464	0.0330
13 cal_average.uncert sbt	2	0.0461	0.0327
14 cal_average.uncert m2	2	0.0461	0.0327
15 cal_average.uncert m1	2	0.0459	0.0325
16 cal_average.uncert bc	2	0.0458	0.0324
17 grad_m2 m2	2	0.0455	0.0322
18 grad_bc.uncert sbt	2	0.0455	0.0321
19 grad_m1.uncert sbt	2	0.0455	0.0321
20 average.uncert sbt	2	0.0455	0.0321
21 grad_m2.uncert sbt	2	0.0455	0.0321
22 grad_m2.uncert m1	2	0.0455	0.0321
23 average.uncert bc	2	0.0455	0.0321
24 grad_bc.uncert bc	2	0.0455	0.0321
25 grad_m1.uncert bc	2	0.0455	0.0321
26 grad_m2.uncert bc	2	0.0455	0.0321
27 grad_bc.uncert m1	2	0.0454	0.0321
28 grad_m1.uncert m1	2	0.0454	0.0321
29 average.uncert m1	2	0.0454	0.0321
30 grad_m2.uncert m2	2	0.0454	0.0321
31 average.uncert m2	2	0.0454	0.0321
32 grad_bc.uncert m2	2	0.0454	0.0321
33 grad_m1.uncert m2	2	0.0454	0.0320
34 grad_m2 sbt	2	0.0454	0.0320
35 grad_m1 m1	2	0.0453	0.0319
36 grad_m2 m1	2	0.0453	0.0319
37 grad_m1 m2	2	0.0452	0.0318
38 logreg_sweep_C.uncert bc	2	0.0449	0.0316
39 average of TemperatureScaling	2	0.0449	0.0316
40 logreg_no_interc_sweep_C.uncert bc	2	0.0449	0.0315
41 logreg_no_interc_sweep_C.uncert m1	2	0.0448	0.0314
42 logreg.uncert bc	2	0.0448	0.0314
43 logreg_no_interc.uncert bc	2	0.0448	0.0314
44 logreg_sweep_C.uncert m1	2	0.0447	0.0314

45	logreg.uncert m1	2	0.0447	0.0313
46	logreg_no_interc.uncert m1	2	0.0447	0.0313
47	logreg_no_interc_sweep_C.uncert m2	2	0.0444	0.0310
48	grad_bc bc	2	0.0443	0.0309
49	logreg_no_interc_sweep_C.uncert sbt	2	0.0443	0.0309
50	logreg_sweep_C.uncert m2	2	0.0443	0.0309
51	logreg_no_interc.uncert m2	2	0.0441	0.0308
52	logreg.uncert m2	2	0.0441	0.0308
53	logreg_sweep_C.uncert sbt	2	0.0440	0.0306
54	grad_m2 bc	2	0.0438	0.0305
55	grad_m1 bc	2	0.0436	0.0302
56	grad_bc m2	2	0.0435	0.0302
57	logreg_no_interc.uncert sbt	2	0.0435	0.0302
58	grad_bc sbt	2	0.0435	0.0302
59	grad_bc m1	2	0.0435	0.0301
60	logreg.uncert sbt	2	0.0434	0.0300
61	average m1	2	0.0433	0.0300
62	average m2	2	0.0433	0.0300
63	average sbt	2	0.0433	0.0300
64	average bc	2	0.0433	0.0299
65	cal_prob_average m2	2	0.0409	0.0275
66	prob_average sbt	2	0.0397	0.0263
67	cal_prob_average sbt	2	0.0392	0.0259
68	prob_average m2	2	0.0389	0.0256
69	prob_average m1	2	0.0367	0.0233
70	cal_prob_average m1	2	0.0359	0.0226
71	logreg_no_interc_sweep_C bc	2	0.0322	0.0188
72	logreg_sweep_C m1	2	0.0318	0.0184
73	logreg_no_interc_sweep_C m1	2	0.0317	0.0184
74	logreg_no_interc bc	2	0.0314	0.0181
75	logreg_no_interc_sweep_C sbt	2	0.0313	0.0179
76	logreg_no_interc_sweep_C m2	2	0.0310	0.0177
77	prob_average bc	2	0.0310	0.0177
78	logreg_sweep_C m2	2	0.0308	0.0174
79	logreg_sweep_C sbt	2	0.0308	0.0174
80	logreg_sweep_C bc	2	0.0298	0.0165
81	logreg bc	2	0.0296	0.0163
82	logreg m1	2	0.0296	0.0162
83	logreg_no_interc m1	2	0.0293	0.0160
84	logreg_no_interc sbt	2	0.0290	0.0156
85	logreg sbt	2	0.0290	0.0156
86	logreg_no_interc m2	2	0.0287	0.0153
87	logreg m2	2	0.0287	0.0153
88	grad_m1 sbt	2	0.0250	0.0116
89	cal_prob_average bc	2	0.0244	0.0110

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Tue Jan 11 16:07:49 2022

	method	combination_size	imp_o_avg	imp_o_max
1	cal_average m1	2	0.0485	0.0351
2	cal_average m2	2	0.0485	0.0351
3	cal_average sbt	2	0.0485	0.0351
4	cal_average bc	2	0.0484	0.0351

5	prob_average.uncert m2	2	0.0469	0.0335
6	prob_average.uncert m1	2	0.0468	0.0334
7	cal_prob_average.uncert m2	2	0.0467	0.0333
8	prob_average.uncert bc	2	0.0467	0.0333
9	prob_average.uncert sbt	2	0.0466	0.0333
10	cal_prob_average.uncert m1	2	0.0465	0.0331
11	cal_prob_average.uncert bc	2	0.0464	0.0331
12	cal_prob_average.uncert sbt	2	0.0464	0.0330
13	cal_average.uncert sbt	2	0.0461	0.0327
14	cal_average.uncert m2	2	0.0461	0.0327
15	cal_average.uncert m1	2	0.0459	0.0325
16	cal_average.uncert bc	2	0.0458	0.0324
17	grad_m2 m2	2	0.0455	0.0322
18	grad_bc.uncert sbt	2	0.0455	0.0321
19	grad_m1.uncert sbt	2	0.0455	0.0321
20	average.uncert sbt	2	0.0455	0.0321
21	grad_m2.uncert sbt	2	0.0455	0.0321
22	grad_m2.uncert m1	2	0.0455	0.0321
23	average.uncert bc	2	0.0455	0.0321
24	grad_bc.uncert bc	2	0.0455	0.0321
25	grad_m1.uncert bc	2	0.0455	0.0321
26	grad_m2.uncert bc	2	0.0455	0.0321
27	grad_bc.uncert m1	2	0.0454	0.0321
28	grad_m1.uncert m1	2	0.0454	0.0321
29	average.uncert m1	2	0.0454	0.0321
30	grad_m2.uncert m2	2	0.0454	0.0321
31	average.uncert m2	2	0.0454	0.0321
32	grad_bc.uncert m2	2	0.0454	0.0321
33	grad_m1.uncert m2	2	0.0454	0.0320
34	grad_m2 sbt	2	0.0454	0.0320
35	grad_m1 m1	2	0.0453	0.0319
36	grad_m2 m1	2	0.0453	0.0319
37	grad_m1 m2	2	0.0452	0.0318
38	logreg_sweep_C.uncert bc	2	0.0449	0.0316
39	average of TemperatureScaling	2	0.0449	0.0316
40	logreg_no_interc_sweep_C.uncert bc	2	0.0449	0.0315
41	logreg_no_interc_sweep_C.uncert m1	2	0.0448	0.0314
42	logreg.uncert bc	2	0.0448	0.0314
43	logreg_no_interc.uncert bc	2	0.0448	0.0314
44	logreg_sweep_C.uncert m1	2	0.0447	0.0314
45	logreg.uncert m1	2	0.0447	0.0313
46	logreg_no_interc.uncert m1	2	0.0447	0.0313
47	logreg_no_interc_sweep_C.uncert m2	2	0.0444	0.0310
48	grad_bc bc	2	0.0443	0.0309
49	logreg_no_interc_sweep_C.uncert sbt	2	0.0443	0.0309
50	logreg_sweep_C.uncert m2	2	0.0443	0.0309
51	logreg_no_interc.uncert m2	2	0.0441	0.0308
52	logreg.uncert m2	2	0.0441	0.0308
53	logreg_sweep_C.uncert sbt	2	0.0440	0.0306
54	grad_m2 bc	2	0.0438	0.0305
55	grad_m1 bc	2	0.0436	0.0302
56	grad_bc m2	2	0.0435	0.0302
57	logreg_no_interc.uncert sbt	2	0.0435	0.0302
58	grad_bc sbt	2	0.0435	0.0302

59	grad_bc m1	2	0.0435	0.0301
60	logreg.uncert sbt	2	0.0434	0.0300
61	average m1	2	0.0433	0.0300
62	average m2	2	0.0433	0.0300
63	average sbt	2	0.0433	0.0300
64	average bc	2	0.0433	0.0299
65	cal_prob_average m2	2	0.0409	0.0275
66	prob_average sbt	2	0.0397	0.0263
67	cal_prob_average sbt	2	0.0392	0.0259
68	prob_average m2	2	0.0389	0.0256
69	prob_average m1	2	0.0367	0.0233
70	cal_prob_average m1	2	0.0359	0.0226
71	logreg_no_interc_sweep_C bc	2	0.0322	0.0188
72	logreg_sweep_C m1	2	0.0318	0.0184
73	logreg_no_interc_sweep_C m1	2	0.0317	0.0184
74	logreg_no_interc bc	2	0.0314	0.0181
75	logreg_no_interc_sweep_C sbt	2	0.0313	0.0179
76	logreg_no_interc_sweep_C m2	2	0.0310	0.0177
77	prob_average bc	2	0.0310	0.0177
78	logreg_sweep_C m2	2	0.0308	0.0174
79	logreg_sweep_C sbt	2	0.0308	0.0174
80	logreg_sweep_C bc	2	0.0298	0.0165
81	logreg bc	2	0.0296	0.0163
82	logreg m1	2	0.0296	0.0162
83	logreg_no_interc m1	2	0.0293	0.0160
84	logreg_no_interc sbt	2	0.0290	0.0156
85	logreg sbt	2	0.0290	0.0156
86	logreg_no_interc m2	2	0.0287	0.0153
87	logreg m2	2	0.0287	0.0153
88	grad_m1 sbt	2	0.0250	0.0116
89	cal_prob_average bc	2	0.0244	0.0110

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Tue Jan 11 16:07:49 2022

	method	combination_size	imp_o_avg	imp_o_max
1	cal_average m1	3	0.0653	0.0424
2	cal_average m2	3	0.0653	0.0424
3	cal_average sbt	3	0.0653	0.0424
4	cal_average bc	3	0.0653	0.0424
5	cal_average.uncert sbt	3	0.0635	0.0405
6	cal_average.uncert m2	3	0.0629	0.0400
7	cal_average.uncert m1	3	0.0624	0.0395
8	prob_average.uncert bc	3	0.0623	0.0394
9	prob_average.uncert m1	3	0.0623	0.0394
10	prob_average.uncert sbt	3	0.0622	0.0393
11	cal_average.uncert bc	3	0.0621	0.0392
12	prob_average.uncert m2	3	0.0621	0.0392
13	cal_prob_average.uncert bc	3	0.0620	0.0391
14	cal_prob_average.uncert m2	3	0.0620	0.0390
15	cal_prob_average.uncert m1	3	0.0619	0.0390
16	cal_prob_average.uncert sbt	3	0.0619	0.0390
17	grad_m2 m2	3	0.0615	0.0386
18	grad_m1.uncert sbt	3	0.0615	0.0386

19	grad_m2.uncert sbt	3	0.0615	0.0385
20	average.uncert sbt	3	0.0615	0.0385
21	grad_bc.uncert sbt	3	0.0615	0.0385
22	average.uncert bc	3	0.0614	0.0385
23	grad_m1.uncert bc	3	0.0614	0.0385
24	grad_bc.uncert bc	3	0.0614	0.0385
25	grad_m2.uncert bc	3	0.0614	0.0385
26	grad_m2.uncert m2	3	0.0614	0.0385
27	grad_m1.uncert m1	3	0.0614	0.0385
28	grad_bc.uncert m1	3	0.0614	0.0385
29	grad_m2.uncert m1	3	0.0614	0.0385
30	average.uncert m2	3	0.0614	0.0385
31	average.uncert m1	3	0.0614	0.0385
32	grad_bc.uncert m2	3	0.0614	0.0385
33	grad_m1.uncert m2	3	0.0614	0.0385
34	grad_m2 sbt	3	0.0614	0.0385
35	grad_m1 m1	3	0.0613	0.0384
36	grad_m1 m2	3	0.0613	0.0384
37	grad_m2 m1	3	0.0613	0.0383
38	average of TemperatureScaling	3	0.0612	0.0383
39	logreg.uncert bc	3	0.0607	0.0378
40	logreg_sweep_C.uncert bc	3	0.0607	0.0377
41	logreg_no_interc_sweep_C.uncert bc	3	0.0607	0.0377
42	grad_m1 sbt	3	0.0606	0.0377
43	logreg_no_interc.uncert bc	3	0.0606	0.0377
44	logreg.uncert m1	3	0.0605	0.0375
45	logreg_no_interc.uncert m1	3	0.0605	0.0375
46	logreg_no_interc_sweep_C.uncert m1	3	0.0605	0.0375
47	logreg_sweep_C.uncert m1	3	0.0604	0.0375
48	logreg_no_interc_sweep_C.uncert sbt	3	0.0601	0.0372
49	logreg_sweep_C.uncert m2	3	0.0601	0.0371
50	logreg.uncert m2	3	0.0600	0.0371
51	logreg_no_interc_sweep_C.uncert m2	3	0.0600	0.0371
52	logreg_sweep_C.uncert sbt	3	0.0599	0.0370
53	logreg_no_interc.uncert m2	3	0.0599	0.0370
54	logreg_no_interc.uncert sbt	3	0.0597	0.0367
55	logreg.uncert sbt	3	0.0596	0.0367
56	grad_bc bc	3	0.0583	0.0354
57	grad_m1 bc	3	0.0580	0.0350
58	grad_m2 bc	3	0.0580	0.0350
59	grad_bc m2	3	0.0580	0.0350
60	grad_bc m1	3	0.0579	0.0350
61	grad_bc sbt	3	0.0579	0.0350
62	average m1	3	0.0574	0.0344
63	average m2	3	0.0574	0.0344
64	average sbt	3	0.0574	0.0344
65	average bc	3	0.0574	0.0344
66	cal_prob_average m2	3	0.0554	0.0325
67	prob_average m2	3	0.0537	0.0307
68	prob_average sbt	3	0.0535	0.0306
69	cal_prob_average sbt	3	0.0520	0.0290
70	prob_average m1	3	0.0515	0.0285
71	cal_prob_average m1	3	0.0493	0.0264
72	logreg_no_interc_sweep_C bc	3	0.0460	0.0231

73	logreg_sweep_C bc	3	0.0448	0.0218
74	logreg_no_interc bc	3	0.0446	0.0216
75	logreg bc	3	0.0437	0.0208
76	prob_average bc	3	0.0433	0.0203
77	logreg_no_interc_sweep_C m1	3	0.0427	0.0198
78	logreg_sweep_C m1	3	0.0425	0.0196
79	logreg_no_interc_sweep_C sbt	3	0.0421	0.0192
80	logreg_no_interc_sweep_C m2	3	0.0418	0.0189
81	logreg_sweep_C sbt	3	0.0418	0.0189
82	logreg_sweep_C m2	3	0.0416	0.0187
83	logreg m1	3	0.0399	0.0169
84	logreg_no_interc m1	3	0.0398	0.0169
85	logreg sbt	3	0.0395	0.0166
86	logreg_no_interc sbt	3	0.0394	0.0165
87	logreg m2	3	0.0392	0.0163
88	logreg_no_interc m2	3	0.0390	0.0160
89	cal_prob_average bc	3	0.0334	0.0104

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Tue Jan 11 16:07:49 2022

method	combination_size	imp_o_avg	imp_o_max
1 cal_average m1	3	0.0653	0.0424
2 cal_average m2	3	0.0653	0.0424
3 cal_average sbt	3	0.0653	0.0424
4 cal_average bc	3	0.0653	0.0424
5 cal_average.uncert sbt	3	0.0635	0.0405
6 cal_average.uncert m2	3	0.0629	0.0400
7 cal_average.uncert m1	3	0.0624	0.0395
8 prob_average.uncert bc	3	0.0623	0.0394
9 prob_average.uncert m1	3	0.0623	0.0394
10 prob_average.uncert sbt	3	0.0622	0.0393
11 cal_average.uncert bc	3	0.0621	0.0392
12 prob_average.uncert m2	3	0.0621	0.0392
13 cal_prob_average.uncert bc	3	0.0620	0.0391
14 cal_prob_average.uncert m2	3	0.0620	0.0390
15 cal_prob_average.uncert m1	3	0.0619	0.0390
16 cal_prob_average.uncert sbt	3	0.0619	0.0390
17 grad_m2 m2	3	0.0615	0.0386
18 grad_m1.uncert sbt	3	0.0615	0.0386
19 grad_m2.uncert sbt	3	0.0615	0.0385
20 average.uncert sbt	3	0.0615	0.0385
21 grad_bc.uncert sbt	3	0.0615	0.0385
22 average.uncert bc	3	0.0614	0.0385
23 grad_m1.uncert bc	3	0.0614	0.0385
24 grad_bc.uncert bc	3	0.0614	0.0385
25 grad_m2.uncert bc	3	0.0614	0.0385
26 grad_m2.uncert m2	3	0.0614	0.0385
27 grad_m1.uncert m1	3	0.0614	0.0385
28 grad_bc.uncert m1	3	0.0614	0.0385
29 grad_m2.uncert m1	3	0.0614	0.0385
30 average.uncert m2	3	0.0614	0.0385
31 average.uncert m1	3	0.0614	0.0385
32 grad_bc.uncert m2	3	0.0614	0.0385

33	grad_m1.uncert m2	3	0.0614	0.0385
34	grad_m2 sbt	3	0.0614	0.0385
35	grad_m1 m1	3	0.0613	0.0384
36	grad_m1 m2	3	0.0613	0.0384
37	grad_m2 m1	3	0.0613	0.0383
38	average of TemperatureScaling	3	0.0612	0.0383
39	logreg.uncert bc	3	0.0607	0.0378
40	logreg_sweep_C.uncert bc	3	0.0607	0.0377
41	logreg_no_interc_sweep_C.uncert bc	3	0.0607	0.0377
42	grad_m1 sbt	3	0.0606	0.0377
43	logreg_no_interc.uncert bc	3	0.0606	0.0377
44	logreg.uncert m1	3	0.0605	0.0375
45	logreg_no_interc.uncert m1	3	0.0605	0.0375
46	logreg_no_interc_sweep_C.uncert m1	3	0.0605	0.0375
47	logreg_sweep_C.uncert m1	3	0.0604	0.0375
48	logreg_no_interc_sweep_C.uncert sbt	3	0.0601	0.0372
49	logreg_sweep_C.uncert m2	3	0.0601	0.0371
50	logreg.uncert m2	3	0.0600	0.0371
51	logreg_no_interc_sweep_C.uncert m2	3	0.0600	0.0371
52	logreg_sweep_C.uncert sbt	3	0.0599	0.0370
53	logreg_no_interc.uncert m2	3	0.0599	0.0370
54	logreg_no_interc.uncert sbt	3	0.0597	0.0367
55	logreg.uncert sbt	3	0.0596	0.0367
56	grad_bc bc	3	0.0583	0.0354
57	grad_m2 bc	3	0.0580	0.0350
58	grad_m1 bc	3	0.0580	0.0350
59	grad_bc m2	3	0.0580	0.0350
60	grad_bc m1	3	0.0579	0.0350
61	grad_bc sbt	3	0.0579	0.0350
62	average m1	3	0.0574	0.0344
63	average m2	3	0.0574	0.0344
64	average sbt	3	0.0574	0.0344
65	average bc	3	0.0574	0.0344
66	cal_prob_average m2	3	0.0554	0.0325
67	prob_average m2	3	0.0537	0.0307
68	prob_average sbt	3	0.0535	0.0306
69	cal_prob_average sbt	3	0.0520	0.0290
70	prob_average m1	3	0.0515	0.0285
71	cal_prob_average m1	3	0.0493	0.0264
72	logreg_no_interc_sweep_C bc	3	0.0460	0.0231
73	logreg_sweep_C bc	3	0.0448	0.0218
74	logreg_no_interc bc	3	0.0446	0.0216
75	logreg bc	3	0.0437	0.0208
76	prob_average bc	3	0.0433	0.0203
77	logreg_no_interc_sweep_C m1	3	0.0427	0.0198
78	logreg_sweep_C m1	3	0.0425	0.0196
79	logreg_no_interc_sweep_C sbt	3	0.0421	0.0192
80	logreg_no_interc_sweep_C m2	3	0.0418	0.0189
81	logreg_sweep_C sbt	3	0.0418	0.0189
82	logreg_sweep_C m2	3	0.0416	0.0187
83	logreg m1	3	0.0399	0.0169
84	logreg_no_interc m1	3	0.0398	0.0169
85	logreg sbt	3	0.0395	0.0166
86	logreg_no_interc sbt	3	0.0394	0.0165

87	logreg m2	3	0.0392	0.0163
88	logreg_no_interc m2	3	0.0390	0.0160
89	cal_prob_average bc	3	0.0334	0.0104

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Tue Jan 11 16:07:49 2022

method	combination_size	imp_o_avg	imp_o_max
1 cal_average m1	4	0.0734	0.0432
2 cal_average m2	4	0.0734	0.0432
3 cal_average bc	4	0.0733	0.0432
4 cal_average.uncert sbt	4	0.0722	0.0420
5 cal_average.uncert m2	4	0.0718	0.0417
6 cal_average.uncert m1	4	0.0714	0.0412
7 cal_average.uncert bc	4	0.0712	0.0411
8 prob_average.uncert m2	4	0.0711	0.0409
9 prob_average.uncert bc	4	0.0710	0.0409
10 prob_average.uncert m1	4	0.0710	0.0409
11 prob_average.uncert sbt	4	0.0709	0.0408
12 cal_prob_average.uncert m2	4	0.0707	0.0406
13 cal_prob_average.uncert m1	4	0.0707	0.0405
14 cal_prob_average.uncert bc	4	0.0707	0.0405
15 cal_prob_average.uncert sbt	4	0.0706	0.0405
16 cal_average sbt	4	0.0704	0.0403
17 average of TemperatureScaling	4	0.0702	0.0401
18 grad_bc.uncert bc	4	0.0702	0.0401
19 grad_m1.uncert bc	4	0.0702	0.0400
20 grad_m2.uncert bc	4	0.0702	0.0400
21 grad_m2.uncert m1	4	0.0702	0.0400
22 average.uncert bc	4	0.0702	0.0400
23 grad_m1.uncert m1	4	0.0701	0.0400
24 average.uncert m1	4	0.0701	0.0400
25 grad_bc.uncert m1	4	0.0701	0.0400
26 grad_m2.uncert sbt	4	0.0701	0.0400
27 grad_m2.uncert m2	4	0.0701	0.0400
28 grad_bc.uncert sbt	4	0.0701	0.0400
29 average.uncert m2	4	0.0701	0.0400
30 grad_m1.uncert m2	4	0.0701	0.0400
31 average.uncert sbt	4	0.0701	0.0400
32 grad_m1.uncert sbt	4	0.0701	0.0400
33 grad_bc.uncert m2	4	0.0701	0.0400
34 grad_m2 m2	4	0.0696	0.0395
35 grad_m2 sbt	4	0.0696	0.0395
36 grad_m2 m1	4	0.0694	0.0393
37 grad_m1 m2	4	0.0693	0.0391
38 grad_m1 m1	4	0.0691	0.0389
39 logreg.uncert bc	4	0.0690	0.0389
40 logreg_no_interc.uncert bc	4	0.0690	0.0389
41 logreg_sweep_C.uncert bc	4	0.0690	0.0389
42 logreg_no_interc_sweep_C.uncert bc	4	0.0690	0.0388
43 logreg_no_interc.uncert m1	4	0.0688	0.0387
44 grad_m1 sbt	4	0.0688	0.0387
45 logreg_no_interc_sweep_C.uncert m1	4	0.0688	0.0387
46 logreg.uncert m1	4	0.0688	0.0386

47	logreg_sweep_C.uncert m1	4	0.0687	0.0386
48	logreg_no_interc_sweep_C.uncert sbt	4	0.0683	0.0382
49	logreg_no_interc_sweep_C.uncert m2	4	0.0683	0.0382
50	logreg_sweep_C.uncert sbt	4	0.0683	0.0382
51	logreg_sweep_C.uncert m2	4	0.0682	0.0381
52	logreg.uncert m2	4	0.0682	0.0380
53	logreg_no_interc.uncert m2	4	0.0682	0.0380
54	logreg.uncert sbt	4	0.0681	0.0380
55	logreg_no_interc.uncert sbt	4	0.0680	0.0379
56	grad_m2 bc	4	0.0647	0.0346
57	grad_bc bc	4	0.0646	0.0344
58	grad_bc m2	4	0.0645	0.0344
59	grad_bc m1	4	0.0645	0.0344
60	grad_bc sbt	4	0.0645	0.0344
61	grad_m1 bc	4	0.0644	0.0343
62	average bc	4	0.0638	0.0336
63	average m1	4	0.0638	0.0336
64	average m2	4	0.0638	0.0336
65	average sbt	4	0.0638	0.0336
66	cal_prob_average m2	4	0.0632	0.0330
67	prob_average m2	4	0.0608	0.0307
68	prob_average sbt	4	0.0606	0.0305
69	prob_average m1	4	0.0588	0.0287
70	cal_prob_average sbt	4	0.0584	0.0283
71	cal_prob_average m1	4	0.0561	0.0260
72	logreg_no_interc_sweep_C bc	4	0.0523	0.0222
73	logreg_sweep_C bc	4	0.0521	0.0219
74	logreg_no_interc bc	4	0.0511	0.0210
75	logreg bc	4	0.0511	0.0209
76	prob_average bc	4	0.0497	0.0195
77	logreg_no_interc_sweep_C m1	4	0.0488	0.0187
78	logreg_sweep_C m1	4	0.0485	0.0184
79	logreg_no_interc_sweep_C sbt	4	0.0483	0.0182
80	logreg_no_interc_sweep_C m2	4	0.0480	0.0179
81	logreg_sweep_C sbt	4	0.0479	0.0178
82	logreg_sweep_C m2	4	0.0477	0.0176
83	logreg_no_interc m1	4	0.0466	0.0165
84	logreg m1	4	0.0466	0.0165
85	logreg_no_interc sbt	4	0.0463	0.0161
86	logreg sbt	4	0.0462	0.0161
87	logreg_no_interc m2	4	0.0461	0.0160
88	logreg m2	4	0.0460	0.0159
89	cal_prob_average bc	4	0.0386	0.0085

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Tue Jan 11 16:07:49 2022

	method	combination_size	imp_o_avg	imp_o_max
1	cal_average m1	4	0.0734	0.0432
2	cal_average m2	4	0.0734	0.0432
3	cal_average bc	4	0.0733	0.0432
4	cal_average.uncert sbt	4	0.0722	0.0420
5	cal_average.uncert m2	4	0.0718	0.0417
6	cal_average.uncert m1	4	0.0714	0.0412

7	cal_average.uncert bc	4	0.0712	0.0411
8	prob_average.uncert m2	4	0.0711	0.0409
9	prob_average.uncert bc	4	0.0710	0.0409
10	prob_average.uncert m1	4	0.0710	0.0409
11	prob_average.uncert sbt	4	0.0709	0.0408
12	cal_prob_average.uncert m2	4	0.0707	0.0406
13	cal_prob_average.uncert m1	4	0.0707	0.0405
14	cal_prob_average.uncert bc	4	0.0707	0.0405
15	cal_prob_average.uncert sbt	4	0.0706	0.0405
16	cal_average sbt	4	0.0704	0.0403
17	average of TemperatureScaling	4	0.0702	0.0401
18	grad_bc.uncert bc	4	0.0702	0.0401
19	grad_m1.uncert bc	4	0.0702	0.0400
20	grad_m2.uncert bc	4	0.0702	0.0400
21	grad_m2.uncert m1	4	0.0702	0.0400
22	average.uncert bc	4	0.0702	0.0400
23	grad_m1.uncert m1	4	0.0701	0.0400
24	average.uncert m1	4	0.0701	0.0400
25	grad_bc.uncert m1	4	0.0701	0.0400
26	grad_m2.uncert sbt	4	0.0701	0.0400
27	grad_m2.uncert m2	4	0.0701	0.0400
28	grad_bc.uncert sbt	4	0.0701	0.0400
29	average.uncert m2	4	0.0701	0.0400
30	grad_m1.uncert m2	4	0.0701	0.0400
31	average.uncert sbt	4	0.0701	0.0400
32	grad_m1.uncert sbt	4	0.0701	0.0400
33	grad_bc.uncert m2	4	0.0701	0.0400
34	grad_m2 m2	4	0.0696	0.0395
35	grad_m2 sbt	4	0.0696	0.0395
36	grad_m2 m1	4	0.0694	0.0393
37	grad_m1 m2	4	0.0693	0.0391
38	grad_m1 m1	4	0.0691	0.0389
39	logreg.uncert bc	4	0.0690	0.0389
40	logreg_no_interc.uncert bc	4	0.0690	0.0389
41	logreg_sweep_C.uncert bc	4	0.0690	0.0389
42	logreg_no_interc_sweep_C.uncert bc	4	0.0690	0.0388
43	logreg_no_interc.uncert m1	4	0.0688	0.0387
44	grad_m1 sbt	4	0.0688	0.0387
45	logreg_no_interc_sweep_C.uncert m1	4	0.0688	0.0387
46	logreg.uncert m1	4	0.0688	0.0386
47	logreg_sweep_C.uncert m1	4	0.0687	0.0386
48	logreg_no_interc_sweep_C.uncert sbt	4	0.0683	0.0382
49	logreg_no_interc_sweep_C.uncert m2	4	0.0683	0.0382
50	logreg_sweep_C.uncert sbt	4	0.0683	0.0382
51	logreg_sweep_C.uncert m2	4	0.0682	0.0381
52	logreg.uncert m2	4	0.0682	0.0380
53	logreg_no_interc.uncert m2	4	0.0682	0.0380
54	logreg.uncert sbt	4	0.0681	0.0380
55	logreg_no_interc.uncert sbt	4	0.0680	0.0379
56	grad_m2 bc	4	0.0647	0.0346
57	grad_bc bc	4	0.0646	0.0344
58	grad_bc m2	4	0.0645	0.0344
59	grad_bc m1	4	0.0645	0.0344
60	grad_bc sbt	4	0.0645	0.0344

61	grad_m1 bc	4	0.0644	0.0343
62	average bc	4	0.0638	0.0336
63	average m1	4	0.0638	0.0336
64	average m2	4	0.0638	0.0336
65	average sbt	4	0.0638	0.0336
66	cal_prob_average m2	4	0.0632	0.0330
67	prob_average m2	4	0.0608	0.0307
68	prob_average sbt	4	0.0606	0.0305
69	prob_average m1	4	0.0588	0.0287
70	cal_prob_average sbt	4	0.0584	0.0283
71	cal_prob_average m1	4	0.0561	0.0260
72	logreg_no_interc_sweep_C bc	4	0.0523	0.0222
73	logreg_sweep_C bc	4	0.0521	0.0219
74	logreg_no_interc bc	4	0.0511	0.0210
75	logreg bc	4	0.0511	0.0209
76	prob_average bc	4	0.0497	0.0195
77	logreg_no_interc_sweep_C m1	4	0.0488	0.0187
78	logreg_sweep_C m1	4	0.0485	0.0184
79	logreg_no_interc_sweep_C sbt	4	0.0483	0.0182
80	logreg_no_interc_sweep_C m2	4	0.0480	0.0179
81	logreg_sweep_C sbt	4	0.0479	0.0178
82	logreg_sweep_C m2	4	0.0477	0.0176
83	logreg_no_interc m1	4	0.0466	0.0165
84	logreg m1	4	0.0466	0.0165
85	logreg_no_interc sbt	4	0.0463	0.0161
86	logreg sbt	4	0.0462	0.0161
87	logreg_no_interc m2	4	0.0461	0.0160
88	logreg m2	4	0.0460	0.0159
89	cal_prob_average bc	4	0.0386	0.0085

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Tue Jan 11 16:07:49 2022

	method	combination	size	imp_o_avg	imp_o_max
1	cal_average bc		5	0.0782	0.0425
2	cal_average m1		5	0.0781	0.0424
3	cal_average m2		5	0.0781	0.0424
4	cal_average.uncert sbt		5	0.0771	0.0414
5	cal_average.uncert m2		5	0.0768	0.0411
6	cal_average.uncert m1		5	0.0765	0.0407
7	cal_average.uncert bc		5	0.0763	0.0406
8	prob_average.uncert bc		5	0.0761	0.0403
9	prob_average.uncert m1		5	0.0760	0.0403
10	prob_average.uncert m2		5	0.0759	0.0402
11	prob_average.uncert sbt		5	0.0759	0.0402
12	average of TemperatureScaling		5	0.0756	0.0399
13	cal_prob_average.uncert m2		5	0.0755	0.0398
14	cal_prob_average.uncert bc		5	0.0755	0.0398
15	cal_prob_average.uncert m1		5	0.0755	0.0398
16	cal_prob_average.uncert sbt		5	0.0754	0.0397
17	grad_m1.uncert bc		5	0.0751	0.0394
18	grad_bc.uncert bc		5	0.0751	0.0393
19	average.uncert bc		5	0.0750	0.0393
20	grad_m2.uncert bc		5	0.0750	0.0393

21	average.uncert m1	5	0.0750	0.0393
22	grad_bc.uncert m1	5	0.0750	0.0393
23	grad_m1.uncert m1	5	0.0750	0.0393
24	grad_m2.uncert m1	5	0.0750	0.0393
25	grad_m1.uncert sbt	5	0.0749	0.0392
26	grad_m2.uncert sbt	5	0.0749	0.0392
27	grad_bc.uncert sbt	5	0.0749	0.0392
28	average.uncert m2	5	0.0749	0.0392
29	grad_m2.uncert m2	5	0.0749	0.0392
30	average.uncert sbt	5	0.0749	0.0392
31	grad_bc.uncert m2	5	0.0749	0.0392
32	grad_m1.uncert m2	5	0.0749	0.0392
33	grad_m2 m2	5	0.0749	0.0391
34	grad_m2 sbt	5	0.0745	0.0388
35	grad_m1 m2	5	0.0743	0.0386
36	grad_m1 sbt	5	0.0743	0.0385
37	grad_m1 m1	5	0.0741	0.0384
38	grad_m2 m1	5	0.0741	0.0384
39	logreg.uncert bc	5	0.0737	0.0380
40	logreg_no_interc_sweep_C.uncert bc	5	0.0736	0.0379
41	logreg_no_interc.uncert bc	5	0.0736	0.0379
42	logreg_sweep_C.uncert bc	5	0.0736	0.0379
43	logreg_no_interc_sweep_C.uncert m1	5	0.0734	0.0377
44	logreg_sweep_C.uncert m1	5	0.0733	0.0376
45	logreg_no_interc.uncert m1	5	0.0733	0.0376
46	logreg.uncert m1	5	0.0733	0.0376
47	logreg_no_interc_sweep_C.uncert sbt	5	0.0729	0.0372
48	logreg_no_interc.uncert m2	5	0.0728	0.0371
49	logreg_no_interc_sweep_C.uncert m2	5	0.0728	0.0371
50	logreg_sweep_C.uncert sbt	5	0.0728	0.0371
51	logreg_sweep_C.uncert m2	5	0.0728	0.0371
52	logreg_no_interc.uncert sbt	5	0.0727	0.0370
53	logreg.uncert sbt	5	0.0727	0.0370
54	logreg.uncert m2	5	0.0726	0.0369
55	grad_m2 bc	5	0.0691	0.0334
56	grad_bc m1	5	0.0690	0.0333
57	grad_bc bc	5	0.0690	0.0333
58	grad_bc m2	5	0.0689	0.0332
59	grad_bc sbt	5	0.0688	0.0331
60	grad_m1 bc	5	0.0685	0.0328
61	average m1	5	0.0678	0.0321
62	average m2	5	0.0678	0.0321
63	average sbt	5	0.0678	0.0321
64	average bc	5	0.0678	0.0320
65	cal_prob_average m2	5	0.0669	0.0312
66	prob_average sbt	5	0.0648	0.0291
67	prob_average m2	5	0.0645	0.0288
68	prob_average m1	5	0.0633	0.0276
69	cal_prob_average sbt	5	0.0626	0.0269
70	cal_prob_average m1	5	0.0605	0.0248
71	logreg_no_interc_sweep_C bc	5	0.0555	0.0198
72	logreg_sweep_C bc	5	0.0550	0.0193
73	logreg_no_interc bc	5	0.0541	0.0183
74	prob_average bc	5	0.0538	0.0181

75	logreg bc	5	0.0538	0.0181
76	logreg_no_interc_sweep_C m1	5	0.0522	0.0165
77	logreg_sweep_C m1	5	0.0520	0.0163
78	logreg_no_interc_sweep_C sbt	5	0.0517	0.0160
79	logreg_sweep_C sbt	5	0.0515	0.0158
80	logreg_no_interc_sweep_C m2	5	0.0514	0.0157
81	logreg_sweep_C m2	5	0.0513	0.0156
82	logreg_no_interc m1	5	0.0502	0.0145
83	logreg_no_interc m2	5	0.0501	0.0144
84	logreg m2	5	0.0500	0.0143
85	logreg_no_interc sbt	5	0.0500	0.0143
86	logreg m1	5	0.0499	0.0142
87	logreg sbt	5	0.0498	0.0141
88	cal_prob_average bc	5	0.0428	0.0071
89	cal_average sbt	5	0.0002	-0.0356

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Tue Jan 11 16:07:49 2022

	method	combination_size	imp_o_avg	imp_o_max
1	cal_average bc	5	0.0782	0.0425
2	cal_average m1	5	0.0781	0.0424
3	cal_average m2	5	0.0781	0.0424
4	cal_average.uncert sbt	5	0.0771	0.0414
5	cal_average.uncert m2	5	0.0768	0.0411
6	cal_average.uncert m1	5	0.0765	0.0407
7	cal_average.uncert bc	5	0.0763	0.0406
8	prob_average.uncert bc	5	0.0761	0.0403
9	prob_average.uncert m1	5	0.0760	0.0403
10	prob_average.uncert m2	5	0.0759	0.0402
11	prob_average.uncert sbt	5	0.0759	0.0402
12	average of TemperatureScaling	5	0.0756	0.0399
13	cal_prob_average.uncert m2	5	0.0755	0.0398
14	cal_prob_average.uncert bc	5	0.0755	0.0398
15	cal_prob_average.uncert m1	5	0.0755	0.0398
16	cal_prob_average.uncert sbt	5	0.0754	0.0397
17	grad_m1.uncert bc	5	0.0751	0.0394
18	grad_bc.uncert bc	5	0.0751	0.0393
19	average.uncert bc	5	0.0750	0.0393
20	grad_m2.uncert bc	5	0.0750	0.0393
21	average.uncert m1	5	0.0750	0.0393
22	grad_bc.uncert m1	5	0.0750	0.0393
23	grad_m1.uncert m1	5	0.0750	0.0393
24	grad_m2.uncert m1	5	0.0750	0.0393
25	grad_m1.uncert sbt	5	0.0749	0.0392
26	grad_m2.uncert sbt	5	0.0749	0.0392
27	grad_bc.uncert sbt	5	0.0749	0.0392
28	average.uncert m2	5	0.0749	0.0392
29	grad_m2.uncert m2	5	0.0749	0.0392
30	average.uncert sbt	5	0.0749	0.0392
31	grad_bc.uncert m2	5	0.0749	0.0392
32	grad_m1.uncert m2	5	0.0749	0.0392
33	grad_m2 m2	5	0.0749	0.0391
34	grad_m2 sbt	5	0.0745	0.0388

35	grad_m1 m2	5	0.0743	0.0386
36	grad_m1 sbt	5	0.0743	0.0385
37	grad_m1 m1	5	0.0741	0.0384
38	grad_m2 m1	5	0.0741	0.0384
39	logreg.uncert bc	5	0.0737	0.0380
40	logreg_no_interc_sweep_C.uncert bc	5	0.0736	0.0379
41	logreg_no_interc.uncert bc	5	0.0736	0.0379
42	logreg_sweep_C.uncert bc	5	0.0736	0.0379
43	logreg_no_interc_sweep_C.uncert m1	5	0.0734	0.0377
44	logreg_sweep_C.uncert m1	5	0.0733	0.0376
45	logreg_no_interc.uncert m1	5	0.0733	0.0376
46	logreg.uncert m1	5	0.0733	0.0376
47	logreg_no_interc_sweep_C.uncert sbt	5	0.0729	0.0372
48	logreg_no_interc.uncert m2	5	0.0728	0.0371
49	logreg_no_interc_sweep_C.uncert m2	5	0.0728	0.0371
50	logreg_sweep_C.uncert sbt	5	0.0728	0.0371
51	logreg_sweep_C.uncert m2	5	0.0728	0.0371
52	logreg_no_interc.uncert sbt	5	0.0727	0.0370
53	logreg.uncert sbt	5	0.0727	0.0370
54	logreg.uncert m2	5	0.0726	0.0369
55	grad_m2 bc	5	0.0691	0.0334
56	grad_bc m1	5	0.0690	0.0333
57	grad_bc bc	5	0.0690	0.0333
58	grad_bc m2	5	0.0689	0.0332
59	grad_bc sbt	5	0.0688	0.0331
60	grad_m1 bc	5	0.0685	0.0328
61	average m1	5	0.0678	0.0321
62	average m2	5	0.0678	0.0321
63	average sbt	5	0.0678	0.0321
64	average bc	5	0.0678	0.0320
65	cal_prob_average m2	5	0.0669	0.0312
66	prob_average sbt	5	0.0648	0.0291
67	prob_average m2	5	0.0645	0.0288
68	prob_average m1	5	0.0633	0.0276
69	cal_prob_average sbt	5	0.0626	0.0269
70	cal_prob_average m1	5	0.0605	0.0248
71	logreg_no_interc_sweep_C bc	5	0.0555	0.0198
72	logreg_sweep_C bc	5	0.0550	0.0193
73	logreg_no_interc bc	5	0.0541	0.0183
74	prob_average bc	5	0.0538	0.0181
75	logreg bc	5	0.0538	0.0181
76	logreg_no_interc_sweep_C m1	5	0.0522	0.0165
77	logreg_sweep_C m1	5	0.0520	0.0163
78	logreg_no_interc_sweep_C sbt	5	0.0517	0.0160
79	logreg_sweep_C sbt	5	0.0515	0.0158
80	logreg_no_interc_sweep_C m2	5	0.0514	0.0157
81	logreg_sweep_C m2	5	0.0513	0.0156
82	logreg_no_interc m1	5	0.0502	0.0145
83	logreg_no_interc m2	5	0.0501	0.0144
84	logreg m2	5	0.0500	0.0143
85	logreg_no_interc sbt	5	0.0500	0.0143
86	logreg m1	5	0.0499	0.0142
87	logreg sbt	5	0.0498	0.0141
88	cal_prob_average bc	5	0.0428	0.0071

89	cal_average sbt	5	0.0002	-0.0356
----	-----------------	---	--------	---------

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Tue Jan 11 16:07:49 2022

method	combination_size	imp_o_avg	imp_o_max
1 cal_average bc	6	0.0819	0.0422
2 cal_average m1	6	0.0819	0.0422
3 cal_average m2	6	0.0819	0.0422
4 cal_average.uncert sbt	6	0.0816	0.0419
5 cal_average.uncert m2	6	0.0807	0.0410
6 cal_average.uncert m1	6	0.0799	0.0402
7 cal_average.uncert bc	6	0.0796	0.0399
8 average of TemperatureScaling	6	0.0793	0.0396
9 prob_average.uncert m1	6	0.0790	0.0393
10 prob_average.uncert bc	6	0.0788	0.0391
11 prob_average.uncert sbt	6	0.0788	0.0391
12 average.uncert bc	6	0.0787	0.0390
13 grad_bc.uncert bc	6	0.0787	0.0390
14 grad_m1.uncert bc	6	0.0787	0.0390
15 prob_average.uncert m2	6	0.0787	0.0390
16 grad_m2.uncert bc	6	0.0786	0.0389
17 average.uncert m1	6	0.0785	0.0388
18 grad_bc.uncert m1	6	0.0785	0.0388
19 grad_m1.uncert m1	6	0.0785	0.0388
20 grad_m2.uncert m1	6	0.0785	0.0388
21 grad_bc.uncert m2	6	0.0784	0.0387
22 grad_m1.uncert m2	6	0.0784	0.0387
23 grad_m2 m2	6	0.0784	0.0387
24 grad_m2.uncert m2	6	0.0784	0.0387
25 average.uncert m2	6	0.0783	0.0386
26 average.uncert sbt	6	0.0783	0.0386
27 grad_bc.uncert sbt	6	0.0783	0.0386
28 grad_m1.uncert sbt	6	0.0783	0.0386
29 grad_m2.uncert sbt	6	0.0783	0.0386
30 cal_prob_average.uncert bc	6	0.0782	0.0385
31 cal_prob_average.uncert m1	6	0.0781	0.0384
32 cal_prob_average.uncert sbt	6	0.0781	0.0384
33 cal_prob_average.uncert m2	6	0.0778	0.0381
34 grad_m1 m2	6	0.0778	0.0381
35 grad_m1 m1	6	0.0776	0.0379
36 grad_m1 sbt	6	0.0775	0.0378
37 grad_m2 sbt	6	0.0774	0.0377
38 grad_m2 m1	6	0.0768	0.0371
39 logreg.uncert m1	6	0.0768	0.0371
40 logreg_no_interc.uncert bc	6	0.0767	0.0370
41 logreg.uncert bc	6	0.0766	0.0369
42 logreg_no_interc_sweep_C.uncert m1	6	0.0766	0.0369
43 logreg_sweep_C.uncert m1	6	0.0766	0.0369
44 logreg_no_interc.uncert m1	6	0.0765	0.0368
45 logreg_sweep_C.uncert bc	6	0.0764	0.0367
46 logreg_no_interc_sweep_C.uncert bc	6	0.0763	0.0366
47 logreg_no_interc.uncert sbt	6	0.0762	0.0365
48 logreg.uncert sbt	6	0.0761	0.0364

49	logreg_sweep_C.uncert m2	6	0.0761	0.0364
50	logreg_no_interc_sweep_C.uncert m2	6	0.0760	0.0363
51	logreg_no_interc_sweep_C.uncert sbt	6	0.0760	0.0363
52	logreg.uncert m2	6	0.0759	0.0362
53	logreg_no_interc.uncert m2	6	0.0759	0.0362
54	logreg_sweep_C.uncert sbt	6	0.0758	0.0361
55	grad_bc bc	6	0.0721	0.0324
56	grad_bc m1	6	0.0713	0.0316
57	grad_m2 bc	6	0.0710	0.0313
58	grad_m1 bc	6	0.0709	0.0312
59	grad_bc m2	6	0.0708	0.0311
60	grad_bc sbt	6	0.0705	0.0308
61	average bc	6	0.0697	0.0300
62	average m1	6	0.0697	0.0300
63	average m2	6	0.0697	0.0300
64	average sbt	6	0.0697	0.0300
65	cal_prob_average m2	6	0.0683	0.0286
66	prob_average sbt	6	0.0670	0.0273
67	prob_average m2	6	0.0665	0.0268
68	prob_average m1	6	0.0664	0.0267
69	cal_prob_average sbt	6	0.0649	0.0252
70	cal_prob_average m1	6	0.0626	0.0229
71	logreg_sweep_C bc	6	0.0590	0.0193
72	logreg bc	6	0.0575	0.0178
73	logreg_no_interc_sweep_C bc	6	0.0574	0.0177
74	prob_average bc	6	0.0567	0.0170
75	logreg_no_interc bc	6	0.0566	0.0169
76	logreg_no_interc_sweep_C sbt	6	0.0549	0.0152
77	logreg_no_interc_sweep_C m2	6	0.0548	0.0151
78	logreg_sweep_C m1	6	0.0548	0.0151
79	logreg_no_interc_sweep_C m1	6	0.0547	0.0150
80	logreg_sweep_C sbt	6	0.0546	0.0149
81	logreg_sweep_C m2	6	0.0540	0.0143
82	logreg m1	6	0.0529	0.0132
83	logreg_no_interc m1	6	0.0528	0.0131
84	logreg sbt	6	0.0526	0.0129
85	logreg_no_interc m2	6	0.0526	0.0129
86	logreg_no_interc sbt	6	0.0524	0.0127
87	logreg m2	6	0.0523	0.0126
88	cal_prob_average bc	6	0.0458	0.0061
89	cal_average sbt	6	-0.2539	-0.2936

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Tue Jan 11 16:07:49 2022

	method	combination_size	imp_o_avg	imp_o_max
1	cal_average bc	6	0.0819	0.0422
2	cal_average m1	6	0.0819	0.0422
3	cal_average m2	6	0.0819	0.0422
4	cal_average.uncert sbt	6	0.0816	0.0419
5	cal_average.uncert m2	6	0.0807	0.0410
6	cal_average.uncert m1	6	0.0799	0.0402
7	cal_average.uncert bc	6	0.0796	0.0399
8	average of TemperatureScaling	6	0.0793	0.0396

9	prob_average.uncert m1	6	0.0790	0.0393
10	prob_average.uncert bc	6	0.0788	0.0391
11	prob_average.uncert sbt	6	0.0788	0.0391
12	average.uncert bc	6	0.0787	0.0390
13	grad_bc.uncert bc	6	0.0787	0.0390
14	grad_m1.uncert bc	6	0.0787	0.0390
15	prob_average.uncert m2	6	0.0787	0.0390
16	grad_m2.uncert bc	6	0.0786	0.0389
17	average.uncert m1	6	0.0785	0.0388
18	grad_bc.uncert m1	6	0.0785	0.0388
19	grad_m1.uncert m1	6	0.0785	0.0388
20	grad_m2.uncert m1	6	0.0785	0.0388
21	grad_bc.uncert m2	6	0.0784	0.0387
22	grad_m1.uncert m2	6	0.0784	0.0387
23	grad_m2 m2	6	0.0784	0.0387
24	grad_m2.uncert m2	6	0.0784	0.0387
25	average.uncert m2	6	0.0783	0.0386
26	average.uncert sbt	6	0.0783	0.0386
27	grad_bc.uncert sbt	6	0.0783	0.0386
28	grad_m1.uncert sbt	6	0.0783	0.0386
29	grad_m2.uncert sbt	6	0.0783	0.0386
30	cal_prob_average.uncert bc	6	0.0782	0.0385
31	cal_prob_average.uncert m1	6	0.0781	0.0384
32	cal_prob_average.uncert sbt	6	0.0781	0.0384
33	cal_prob_average.uncert m2	6	0.0778	0.0381
34	grad_m1 m2	6	0.0778	0.0381
35	grad_m1 m1	6	0.0776	0.0379
36	grad_m1 sbt	6	0.0775	0.0378
37	grad_m2 sbt	6	0.0774	0.0377
38	grad_m2 m1	6	0.0768	0.0371
39	logreg.uncert m1	6	0.0768	0.0371
40	logreg_no_interc.uncert bc	6	0.0767	0.0370
41	logreg.uncert bc	6	0.0766	0.0369
42	logreg_no_interc_sweep_C.uncert m1	6	0.0766	0.0369
43	logreg_sweep_C.uncert m1	6	0.0766	0.0369
44	logreg_no_interc.uncert m1	6	0.0765	0.0368
45	logreg_sweep_C.uncert bc	6	0.0764	0.0367
46	logreg_no_interc_sweep_C.uncert bc	6	0.0763	0.0366
47	logreg_no_interc.uncert sbt	6	0.0762	0.0365
48	logreg.uncert sbt	6	0.0761	0.0364
49	logreg_sweep_C.uncert m2	6	0.0761	0.0364
50	logreg_no_interc_sweep_C.uncert m2	6	0.0760	0.0363
51	logreg_no_interc_sweep_C.uncert sbt	6	0.0760	0.0363
52	logreg.uncert m2	6	0.0759	0.0362
53	logreg_no_interc.uncert m2	6	0.0759	0.0362
54	logreg_sweep_C.uncert sbt	6	0.0758	0.0361
55	grad_bc bc	6	0.0721	0.0324
56	grad_bc m1	6	0.0713	0.0316
57	grad_m2 bc	6	0.0710	0.0313
58	grad_m1 bc	6	0.0709	0.0312
59	grad_bc m2	6	0.0708	0.0311
60	grad_bc sbt	6	0.0705	0.0308
61	average bc	6	0.0697	0.0300
62	average m1	6	0.0697	0.0300

63	average m2	6	0.0697	0.0300
64	average sbt	6	0.0697	0.0300
65	cal_prob_average m2	6	0.0683	0.0286
66	prob_average sbt	6	0.0670	0.0273
67	prob_average m2	6	0.0665	0.0268
68	prob_average m1	6	0.0664	0.0267
69	cal_prob_average sbt	6	0.0649	0.0252
70	cal_prob_average m1	6	0.0626	0.0229
71	logreg_sweep_C bc	6	0.0590	0.0193
72	logreg bc	6	0.0575	0.0178
73	logreg_no_interc_sweep_C bc	6	0.0574	0.0177
74	prob_average bc	6	0.0567	0.0170
75	logreg_no_interc bc	6	0.0566	0.0169
76	logreg_no_interc_sweep_C sbt	6	0.0549	0.0152
77	logreg_no_interc_sweep_C m2	6	0.0548	0.0151
78	logreg_sweep_C m1	6	0.0548	0.0151
79	logreg_no_interc_sweep_C m1	6	0.0547	0.0150
80	logreg_sweep_C sbt	6	0.0546	0.0149
81	logreg_sweep_C m2	6	0.0540	0.0143
82	logreg m1	6	0.0529	0.0132
83	logreg_no_interc m1	6	0.0528	0.0131
84	logreg sbt	6	0.0526	0.0129
85	logreg_no_interc m2	6	0.0526	0.0129
86	logreg_no_interc sbt	6	0.0524	0.0127
87	logreg m2	6	0.0523	0.0126
88	cal_prob_average bc	6	0.0458	0.0061
89	cal_average sbt	6	-0.2539	-0.2936