WeightedLinearEnsemble evaluation validation training

```
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(comprehenr)
```

```
## Warning: package 'comprehenr' 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
```

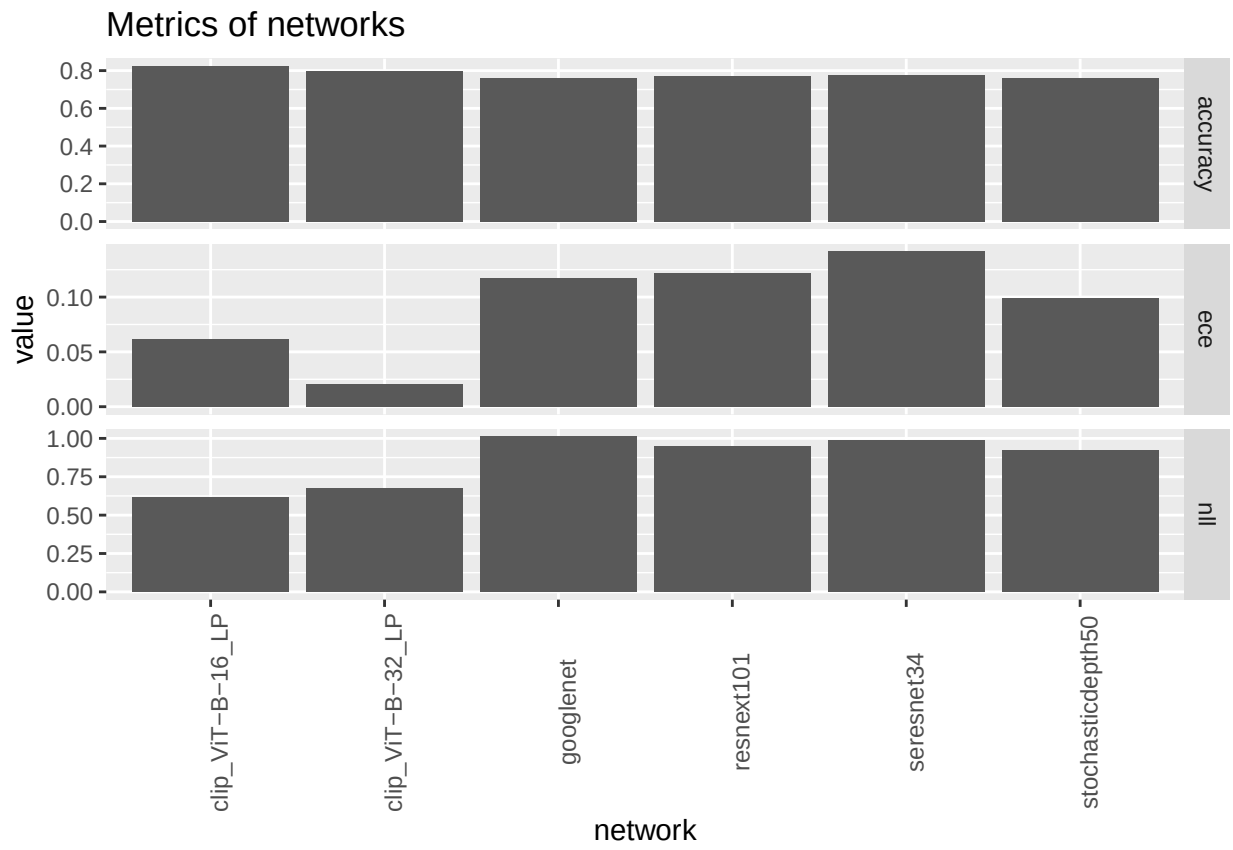
```
library(xtable)
```

```
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/evaluation_va
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
```

Metrics for networks.

```
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
```



Plots for individual combinations of networks.

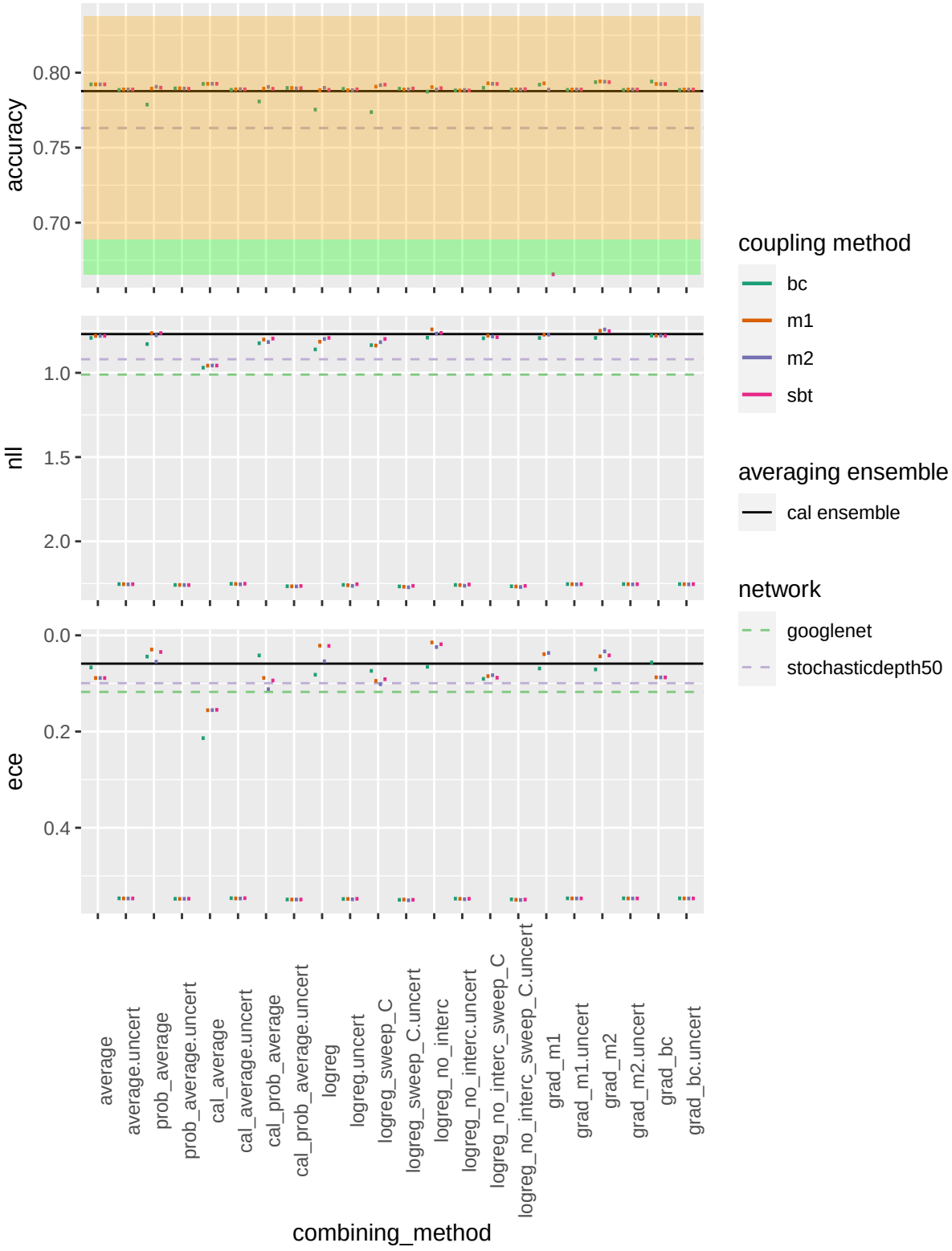
```r
ens_pwc_plt_df <- ens_df_pwc %>% filter(combining_method != "lda")
ens_cal_plt_df <- ens_df_cal

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(
```

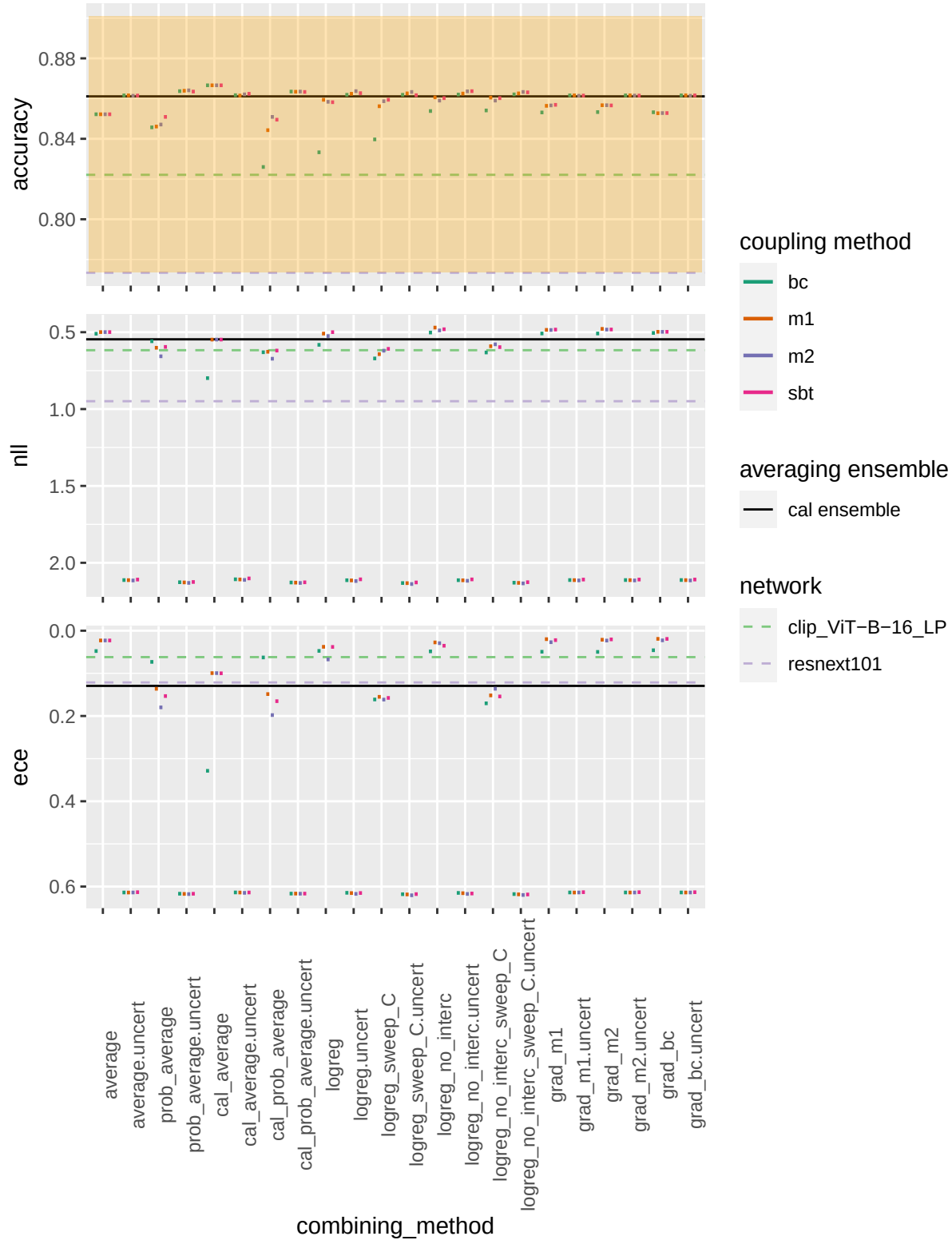
```
      data = cur_nets,
      mapping = aes(yintercept = nll, colour1 = network),
      linetype = "dashed"
    ) %>%
      rename_geom_aes(new_aes = c("colour" = "colour1"))
  ) +
  geom_hline(
    data = cur_ens_cal,
    mapping = aes(yintercept = nll, color = "cal ensemble")
  ) +
  (
    geom_hpline(
      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$err_incons[[1]]),
          collapse = " "
        ),
        paste(
          c("Average pairwise accuracy variance", cur_ens_cal$mean_pwa_var[[1]]),
          collapse = " "
        ),
        sep = "\n"
      )))
  }
}
```

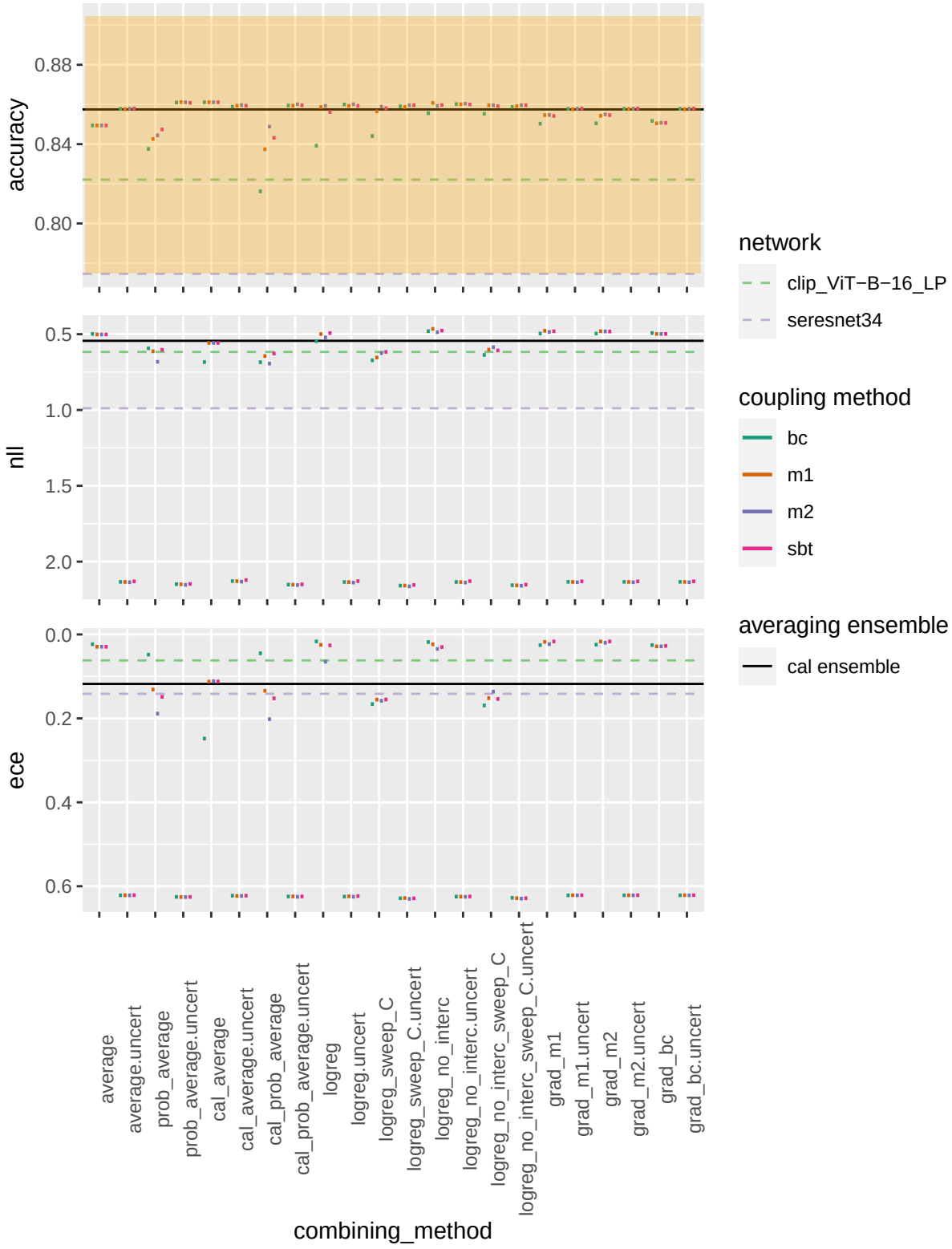
Ensemble metrics
Error inconsistency 0.148800000548363
Average pairwise accuracy variance 2.3084990971256e−05



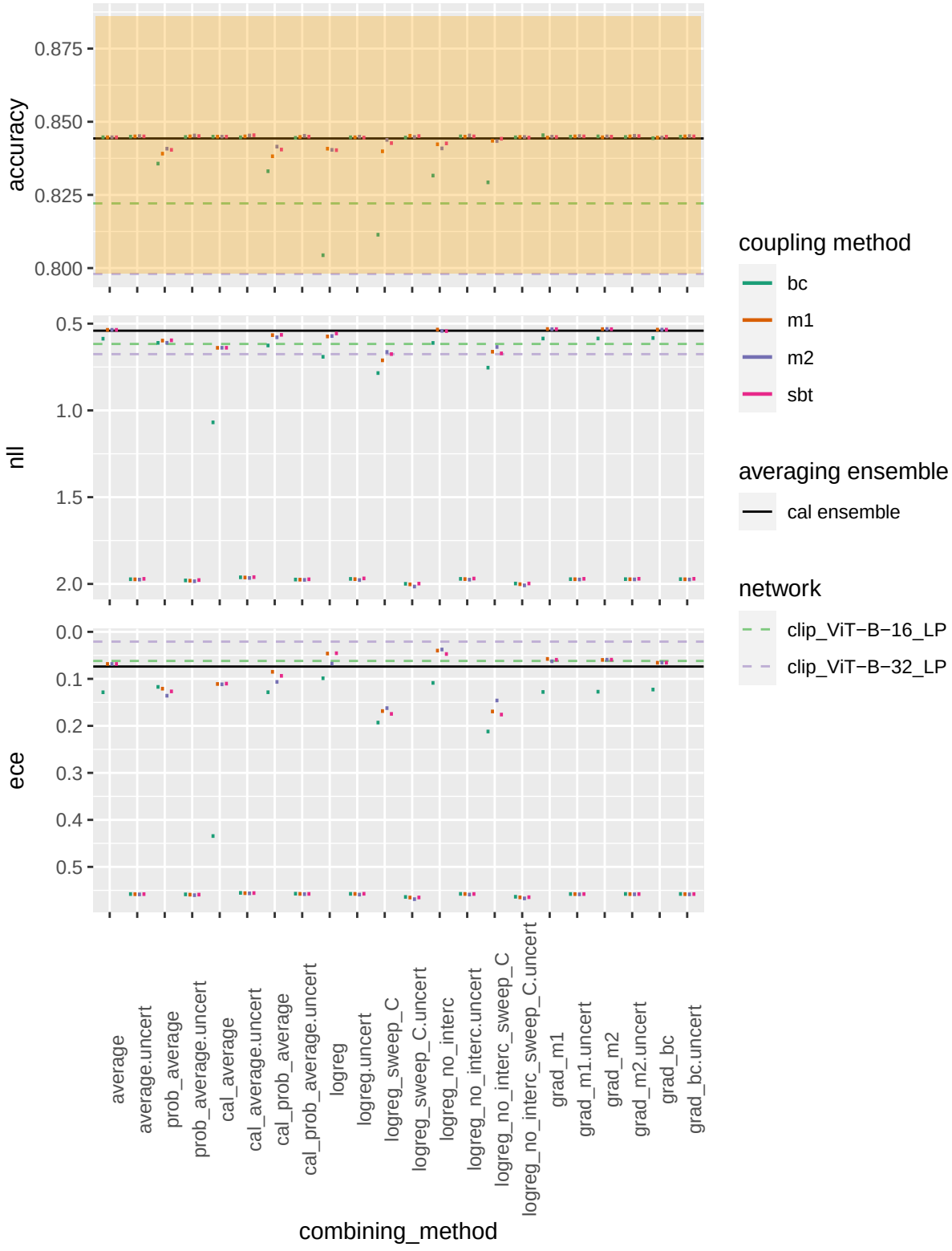
Ensemble metrics
Error inconsistency 0.206299990415573
Average pairwise accuracy variance 3.54999865521677e−05



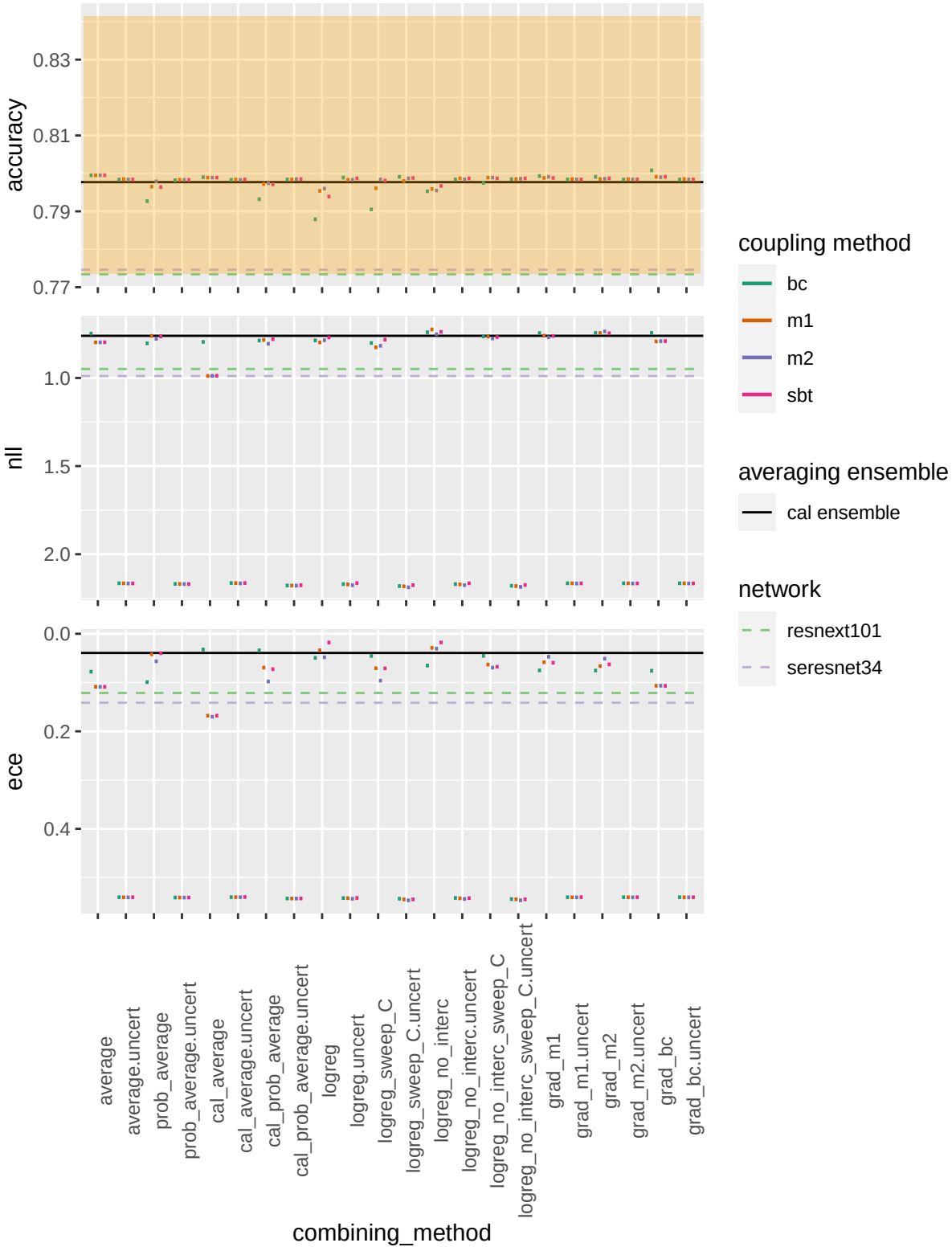
Ensemble metrics
Error inconsistency 0.212099999189377
Average pairwise accuracy variance 3.87824838981032e−05



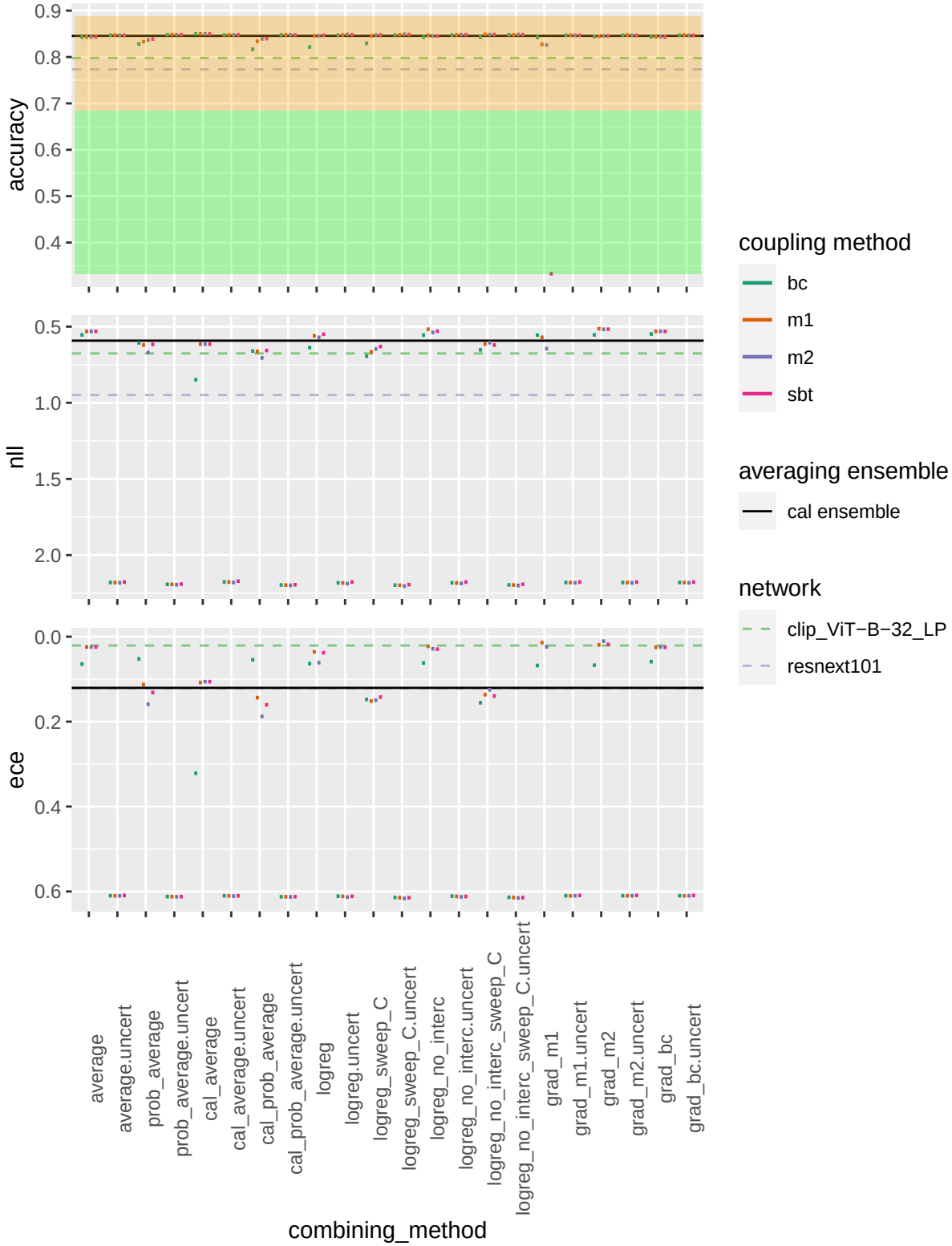
Ensemble metrics
Error inconsistency 0.151899993419647
Average pairwise accuracy variance 1.15449938675738e−05



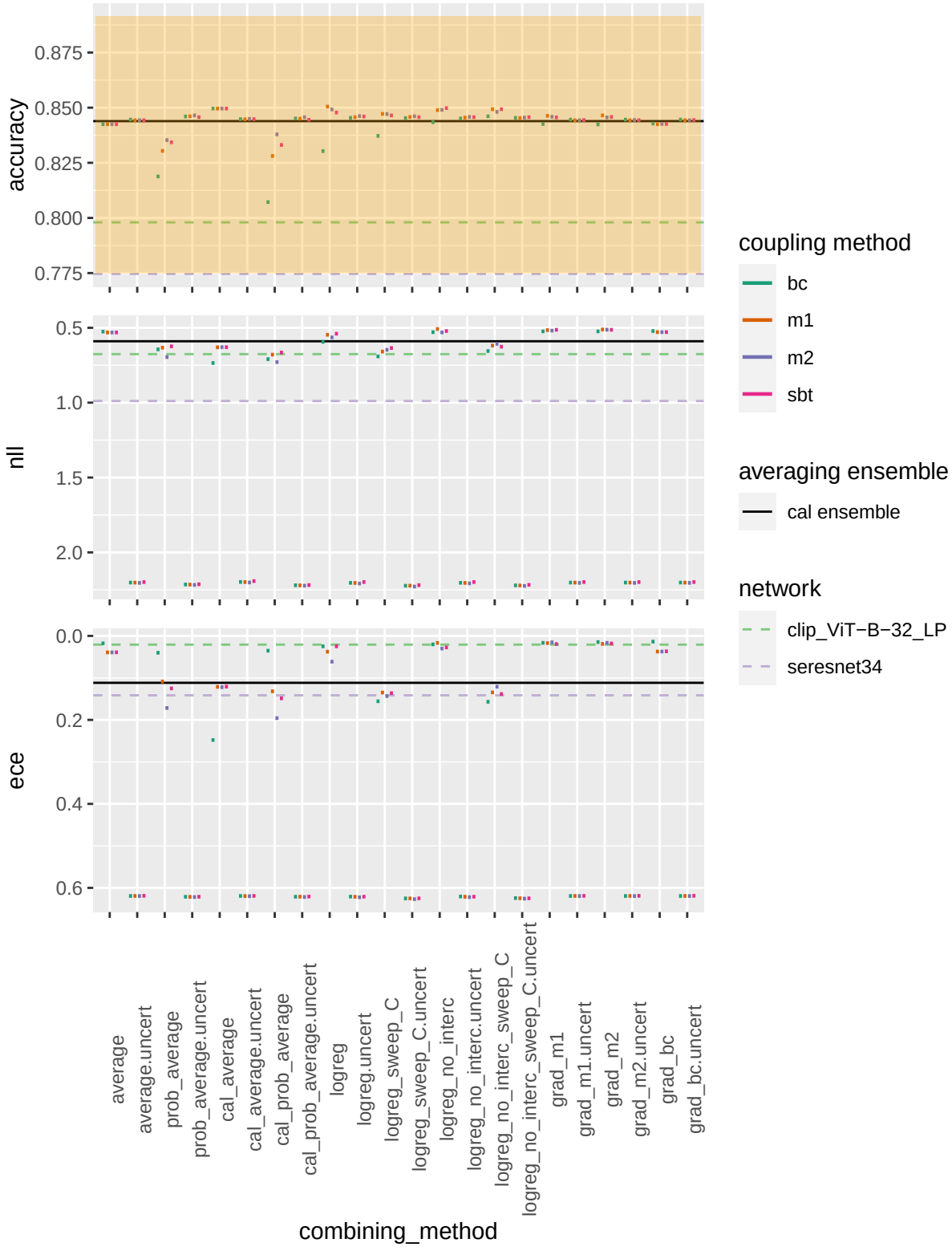
Ensemble metrics
Error inconsistency 0.134800001978874
Average pairwise accuracy variance 1.62699925567722e−05



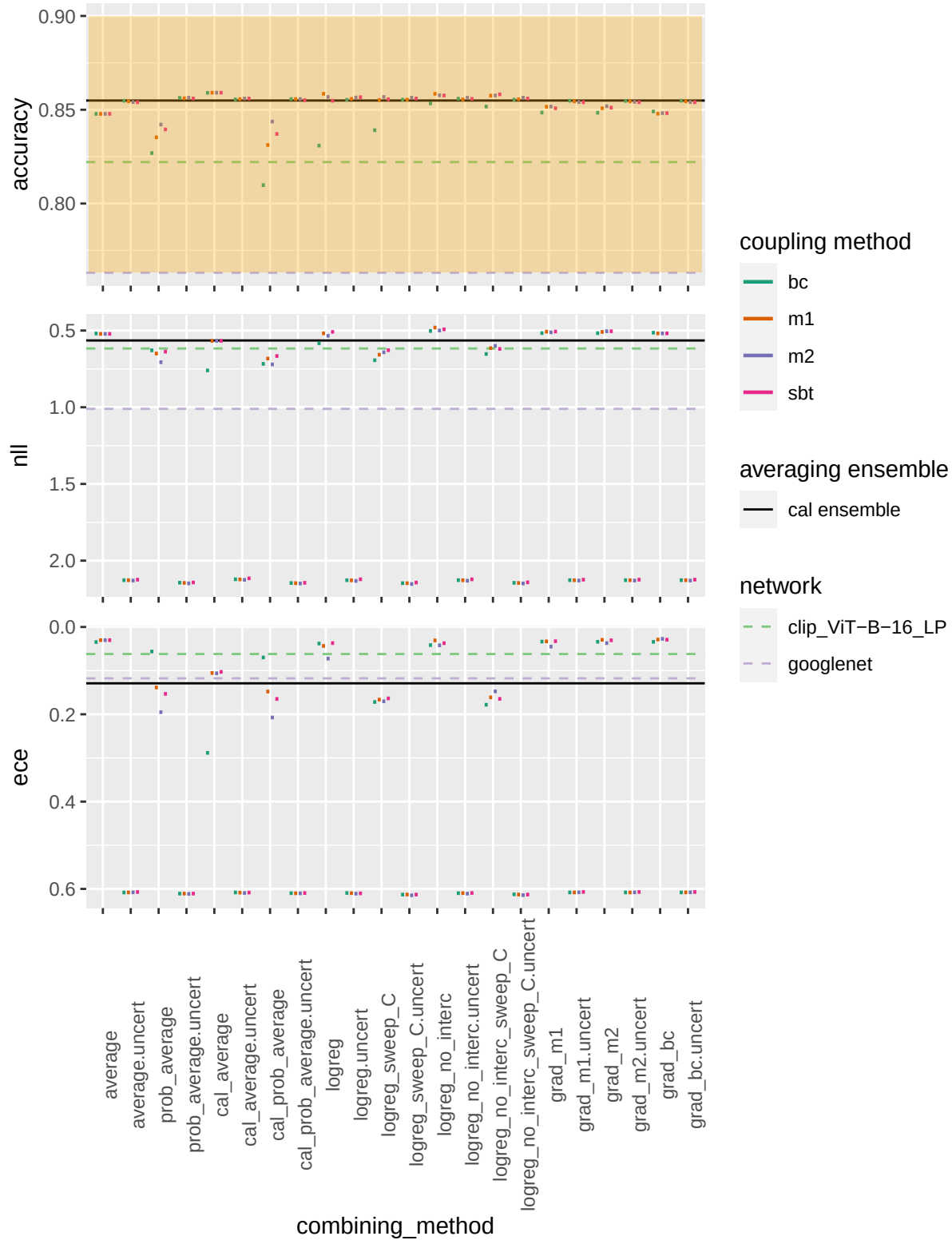
Ensemble metrics
Error inconsistency 0.203799992799759
Average pairwise accuracy variance 3.10324903693981e−05



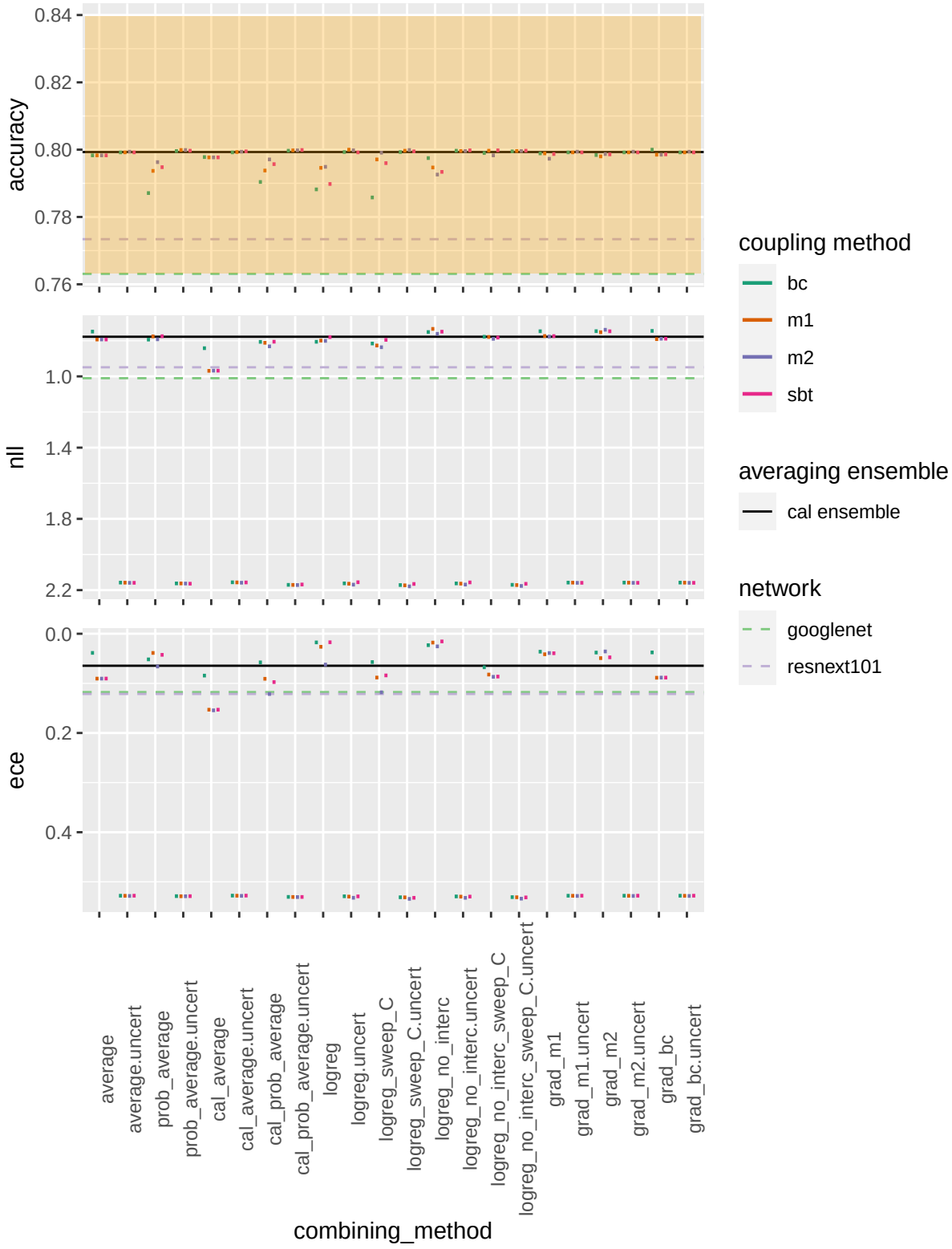
Ensemble metrics
Error inconsistency 0.210199996829033
Average pairwise accuracy variance 3.34249853040092e−05



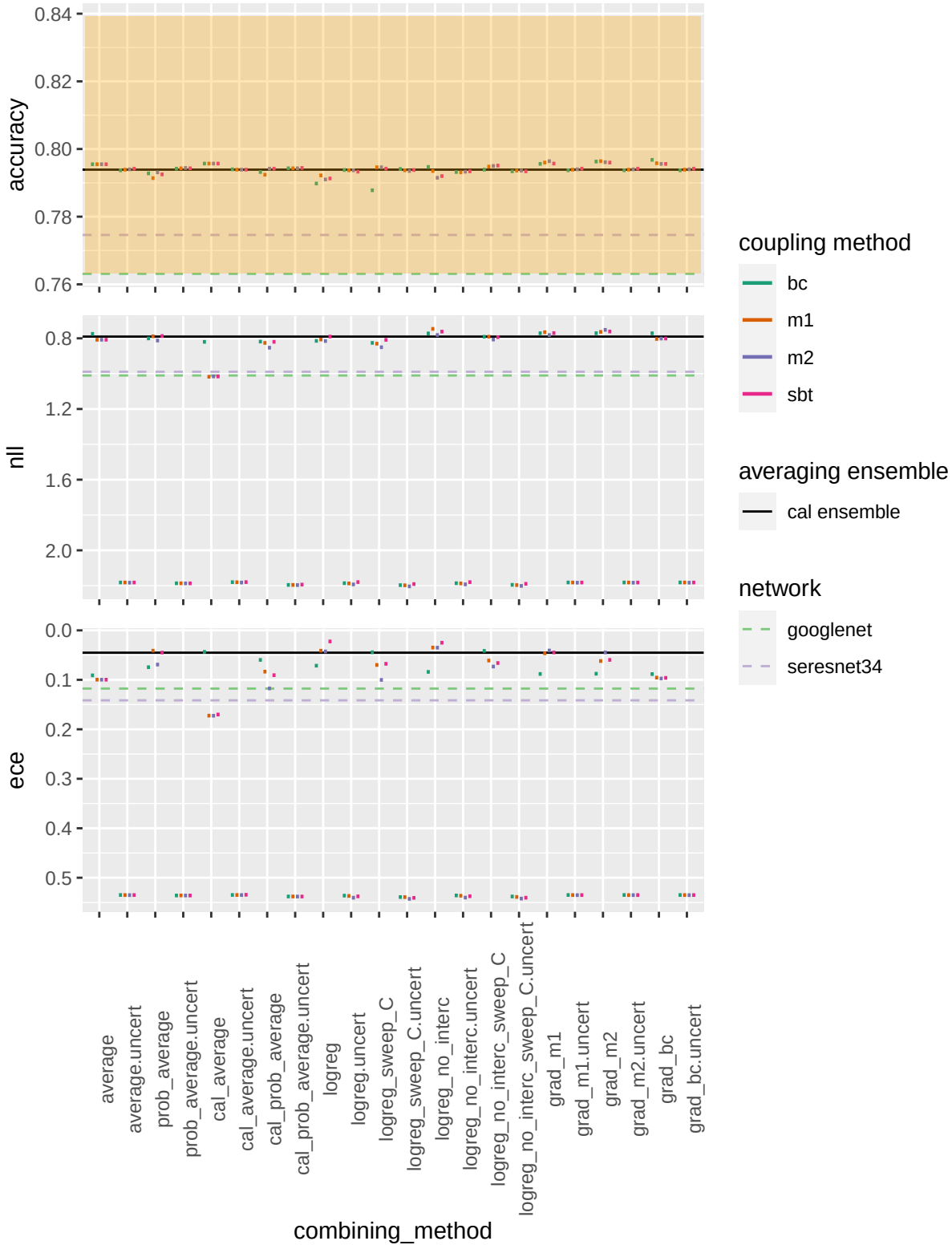
Ensemble metrics
Error inconsistency 0.214399993419647
Average pairwise accuracy variance 5.38337335456163e−05



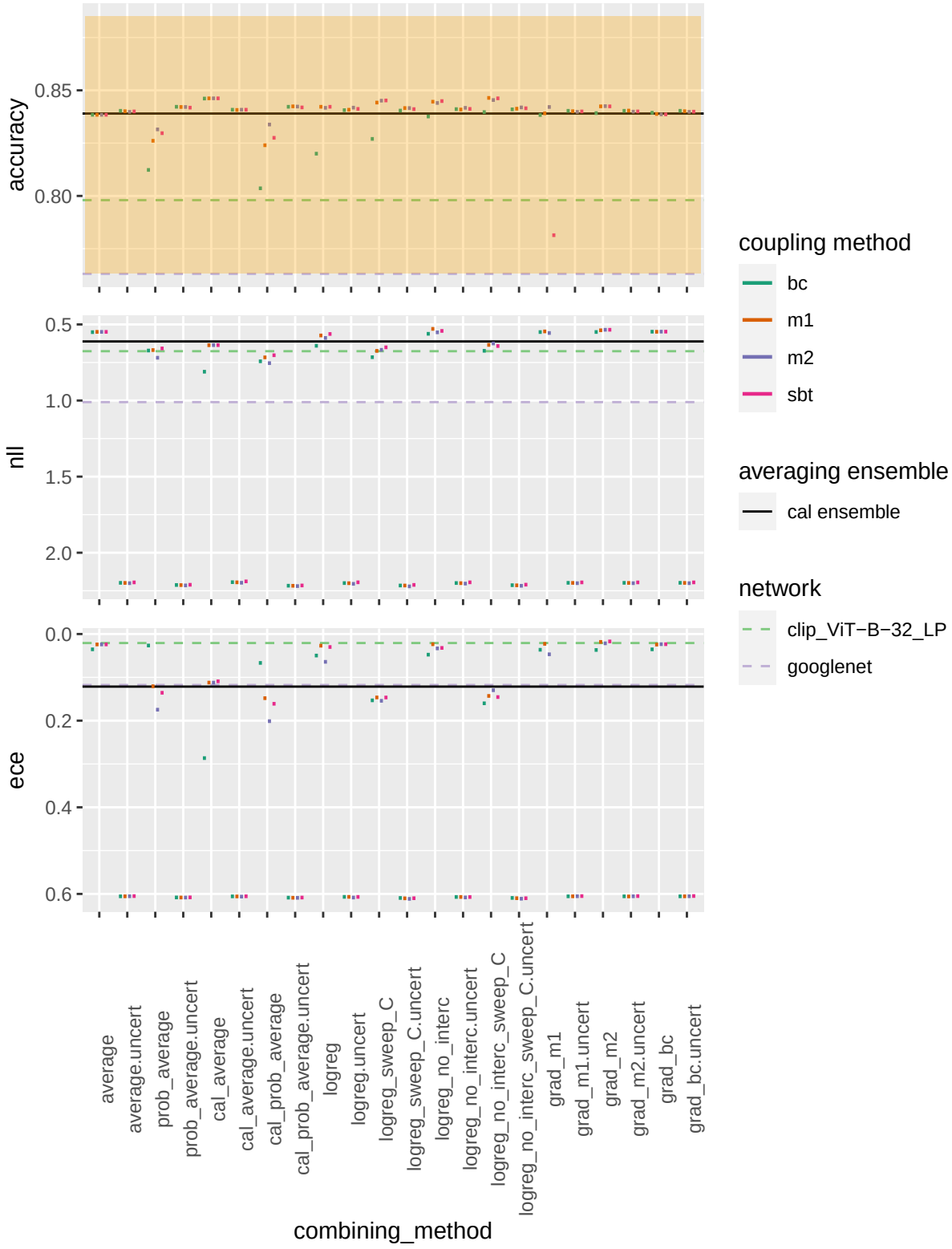
Ensemble metrics
Error inconsistency 0.142700001597404
Average pairwise accuracy variance 2.15937434404623e−05



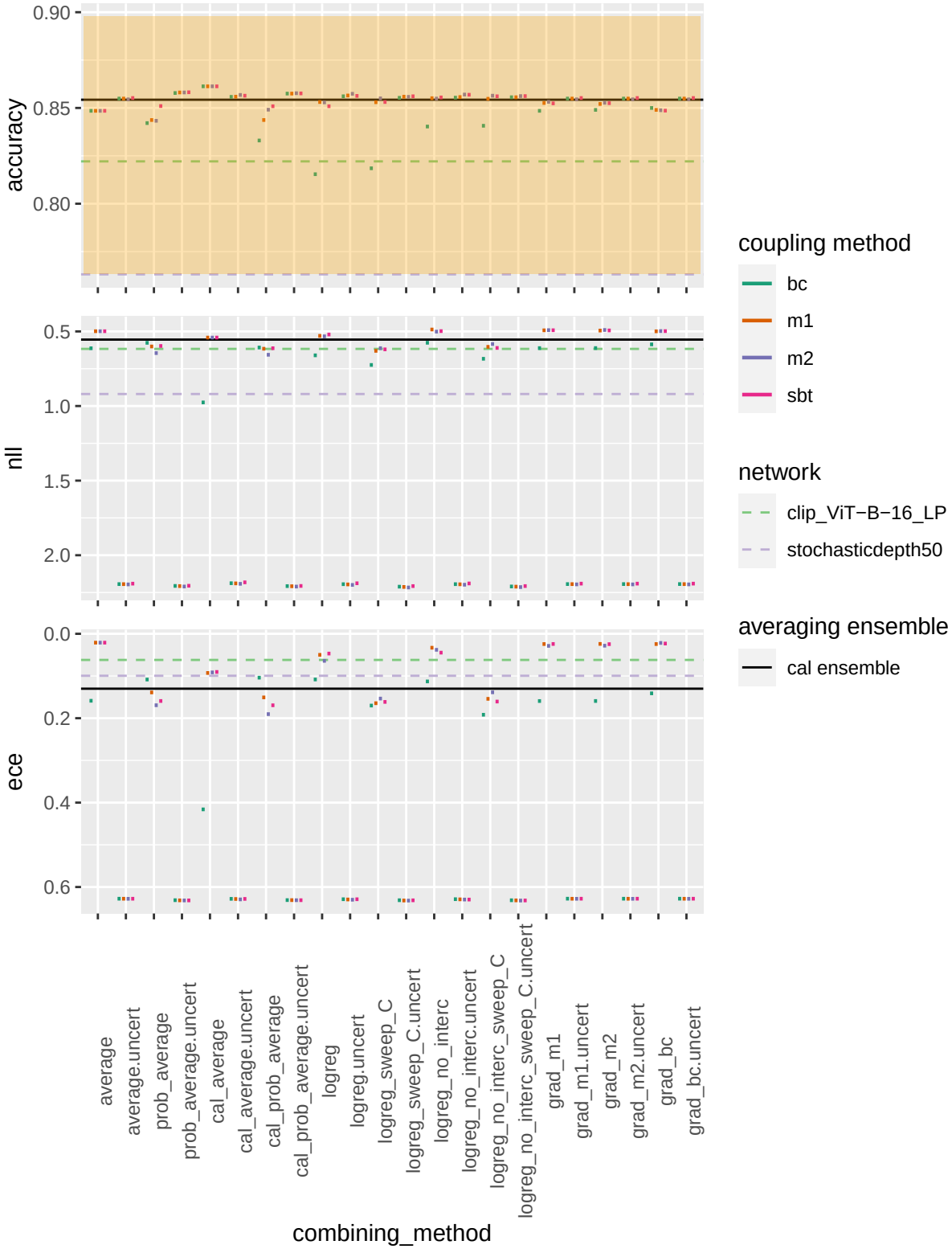
Ensemble metrics
Error inconsistency 0.140699997544289
Average pairwise accuracy variance 2.17712404264603e−05



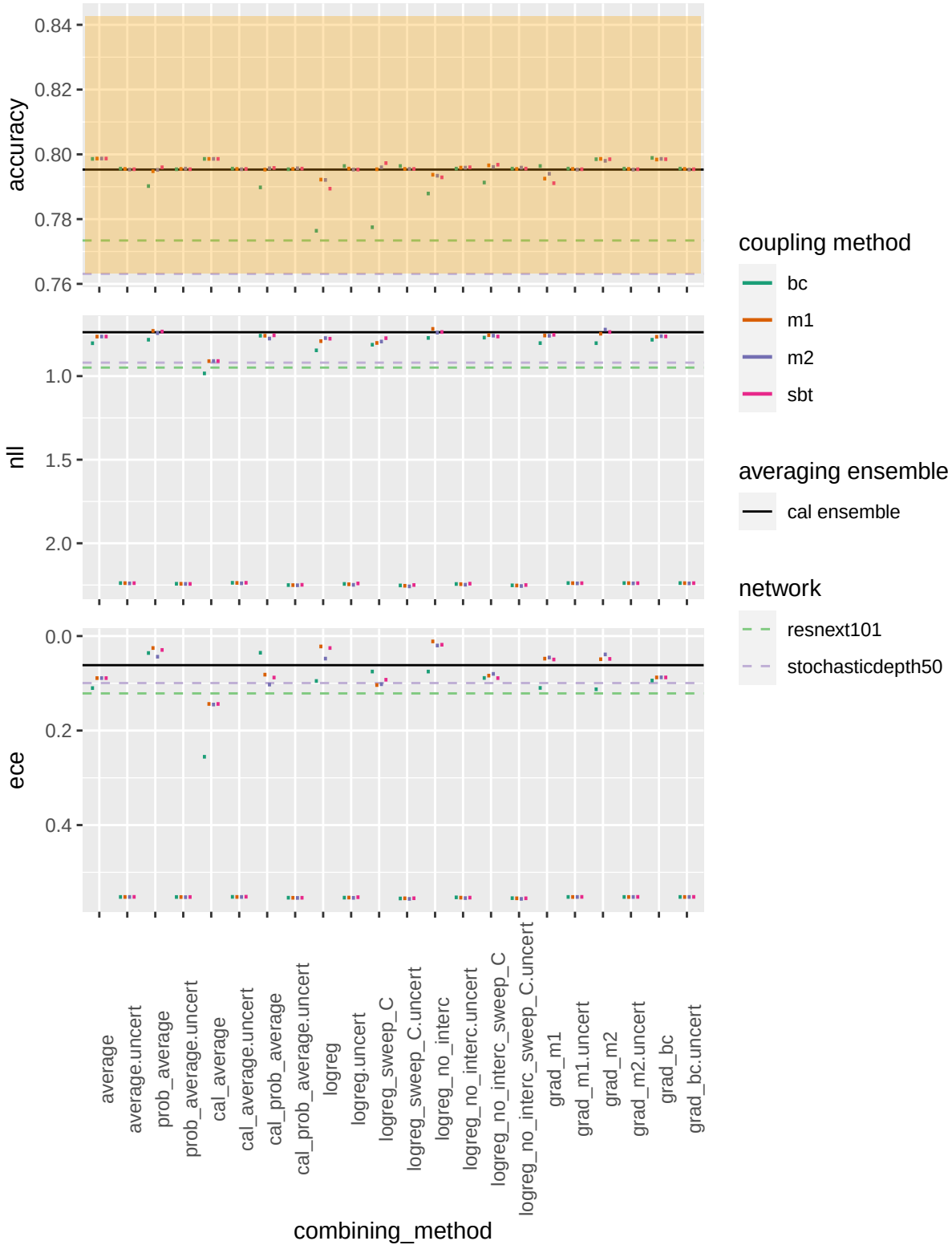
Ensemble metrics
Error inconsistency 0.208899989724159
Average pairwise accuracy variance 4.53812390333042e−05



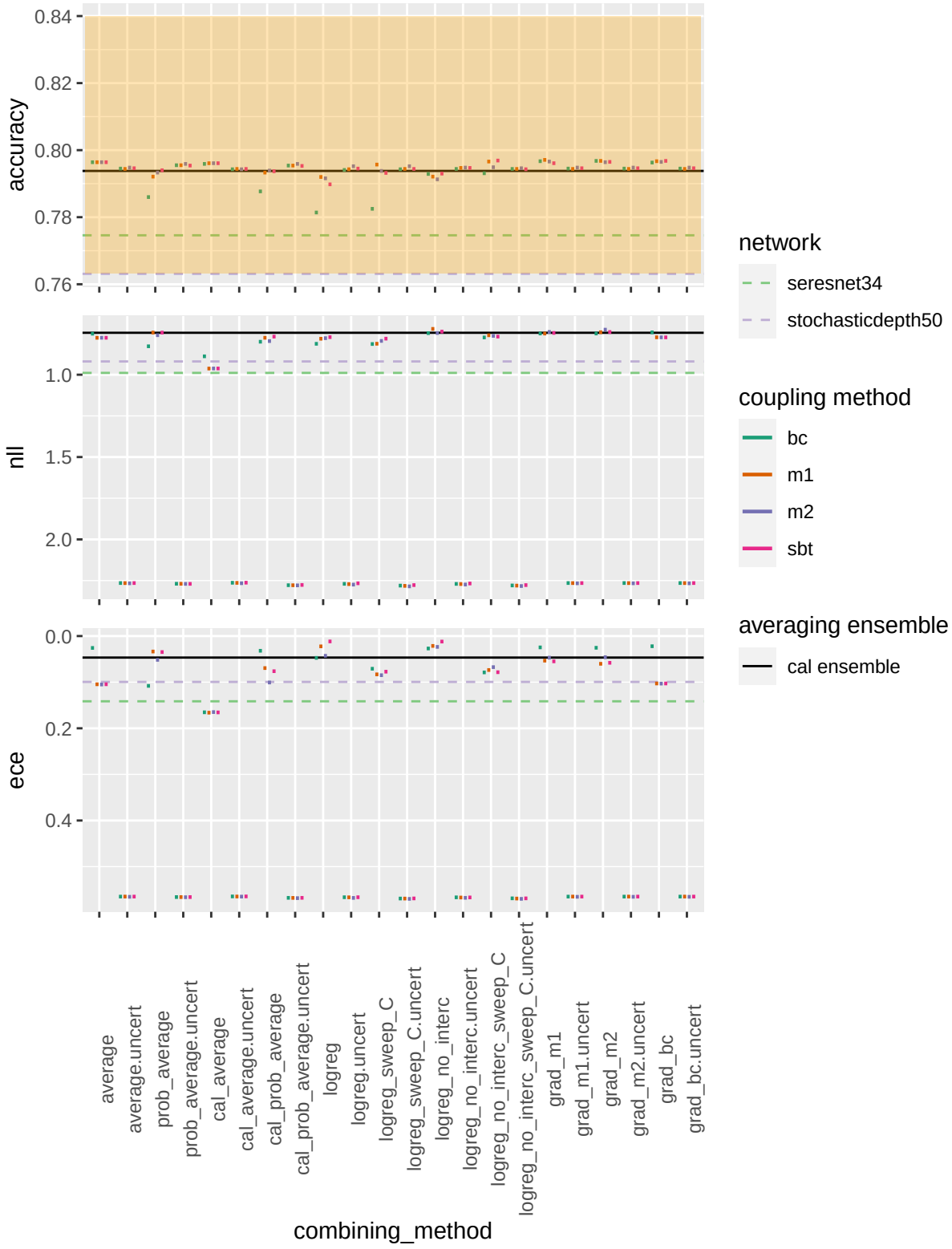
Ensemble metrics
Error inconsistency 0.210400000214577
Average pairwise accuracy variance 3.26337321894243e−05



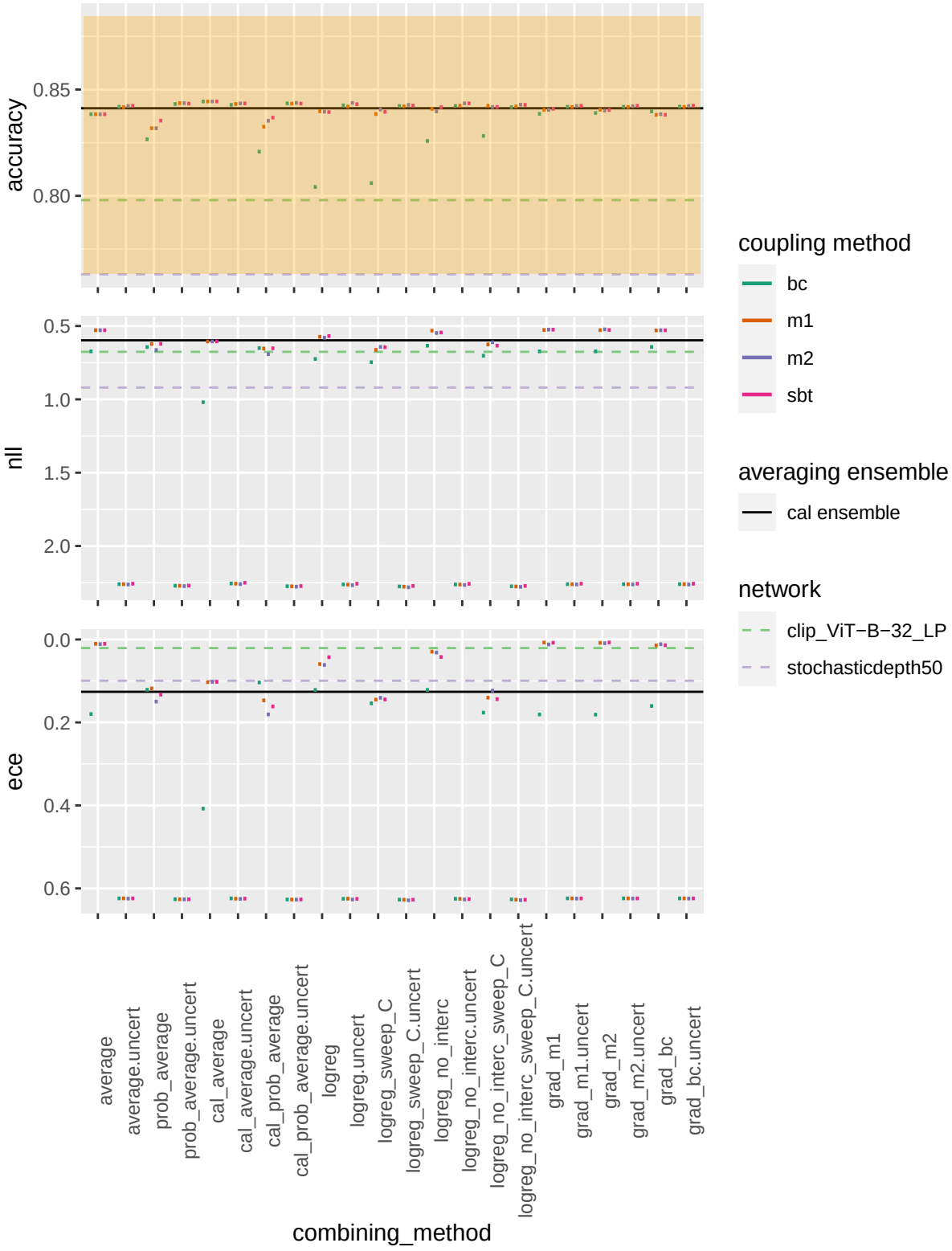
Ensemble metrics
Error inconsistency 0.148699998855591
Average pairwise accuracy variance 1.73487424035557e−05



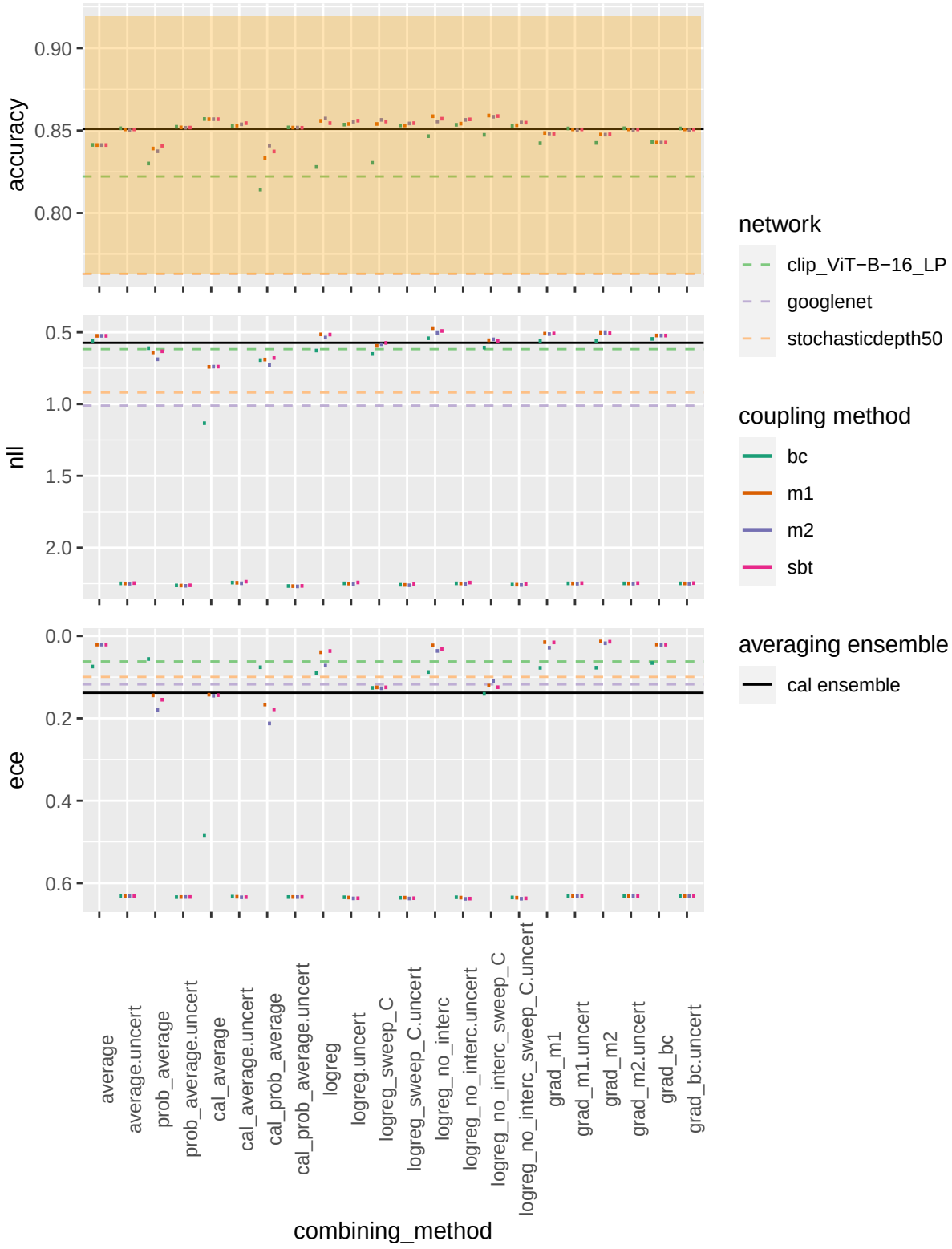
Ensemble metrics
Error inconsistency 0.142099991440773
Average pairwise accuracy variance 1.83112406375585e−05



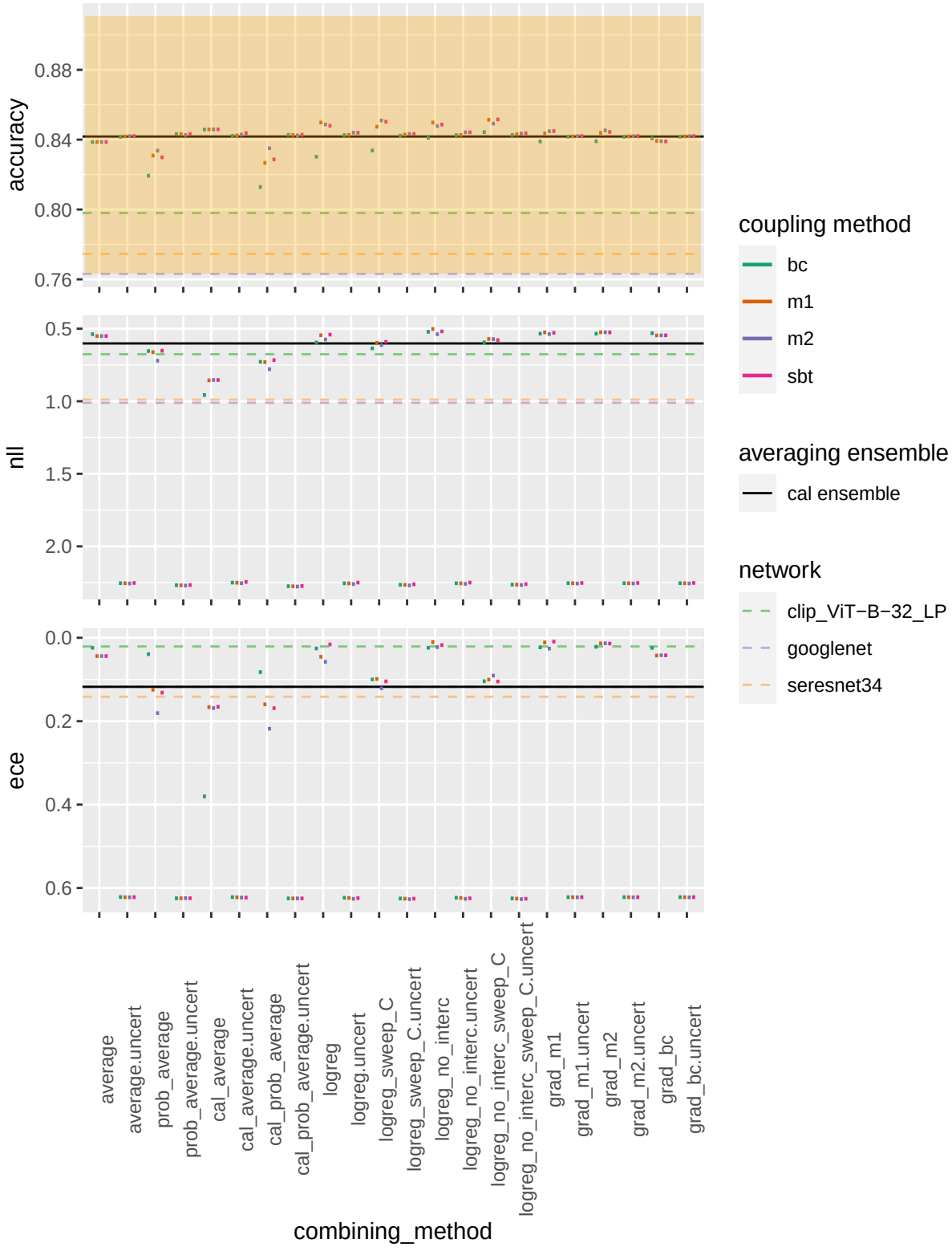
Ensemble metrics
Error inconsistency 0.207699999213219
Average pairwise accuracy variance 2.9091237593093e−05



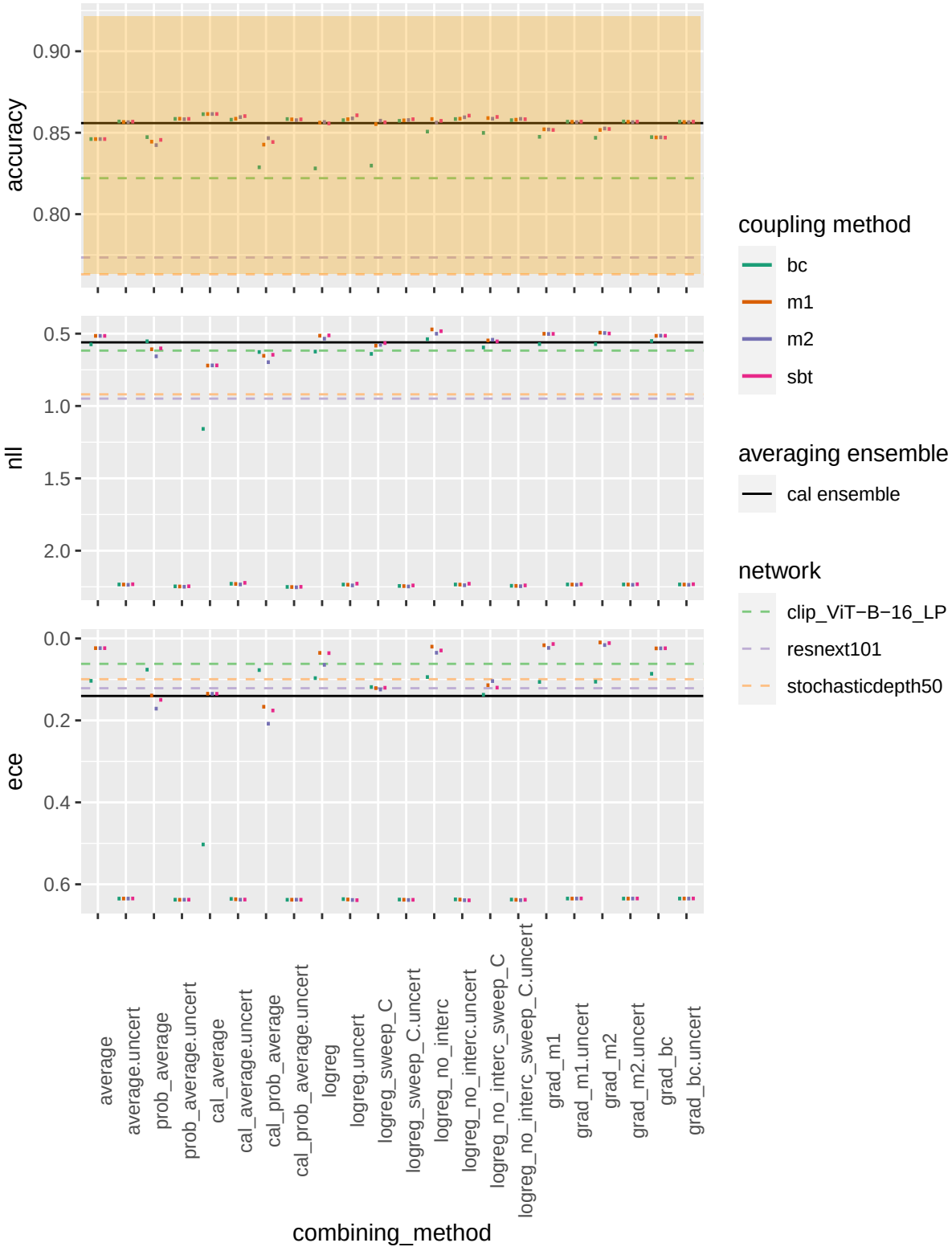
Ensemble metrics
Error inconsistency 0.286799997091293
Average pairwise accuracy variance 4.86899771203753e−05



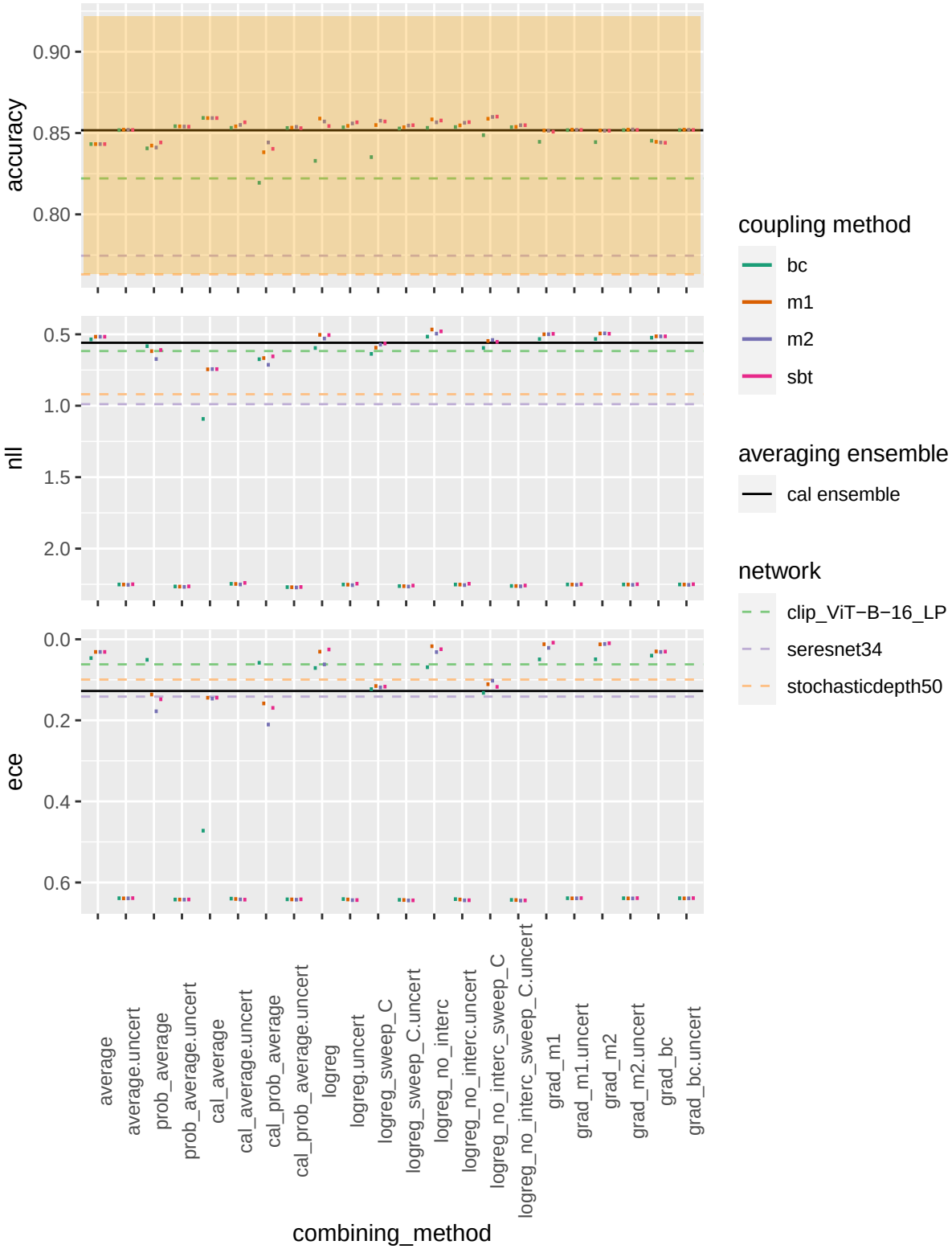
Ensemble metrics
Error inconsistency 0.27989998459816
Average pairwise accuracy variance 4.47010970674455e−05



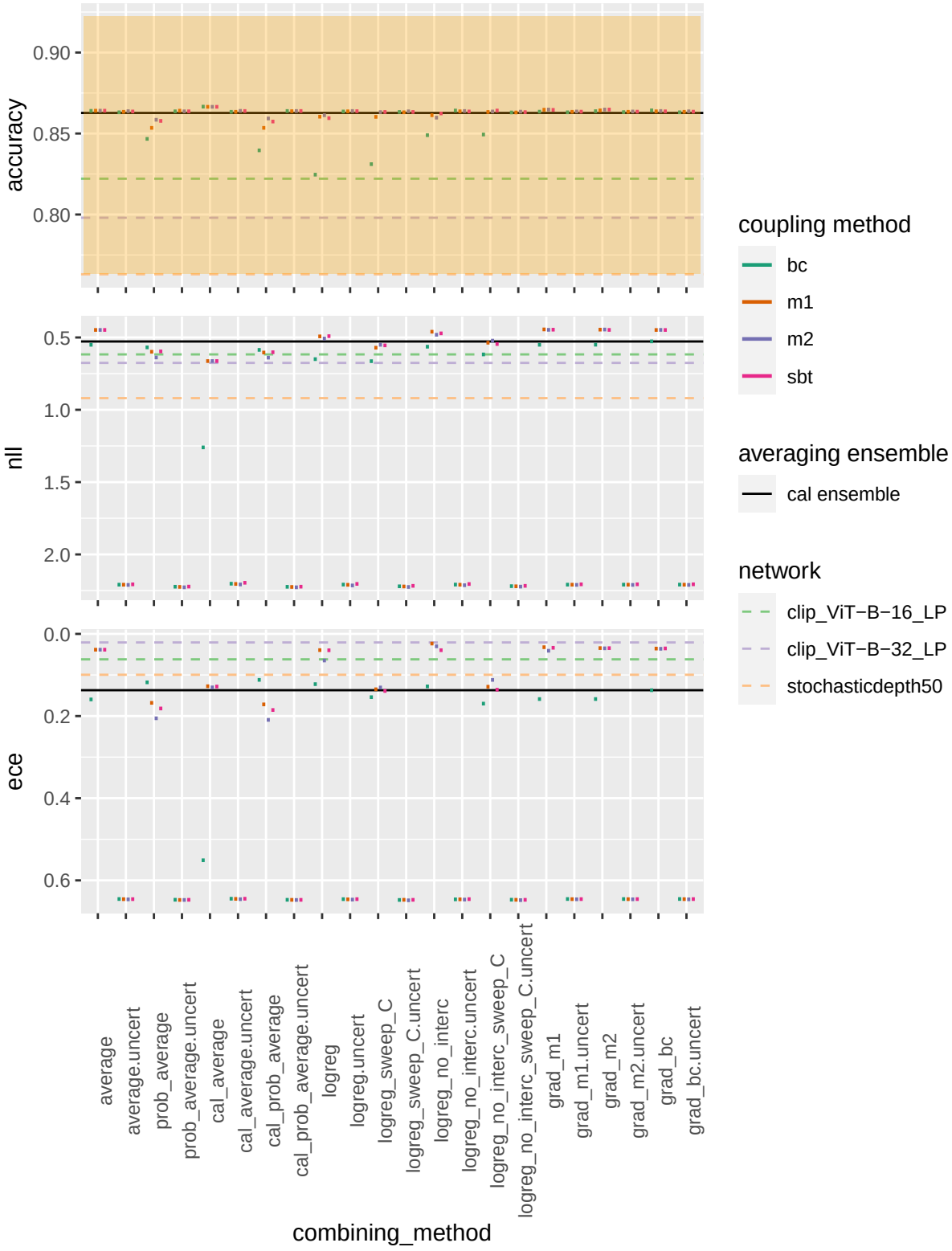
Ensemble metrics
Error inconsistency 0.282700002193451
Average pairwise accuracy variance 3.79922021238599e−05



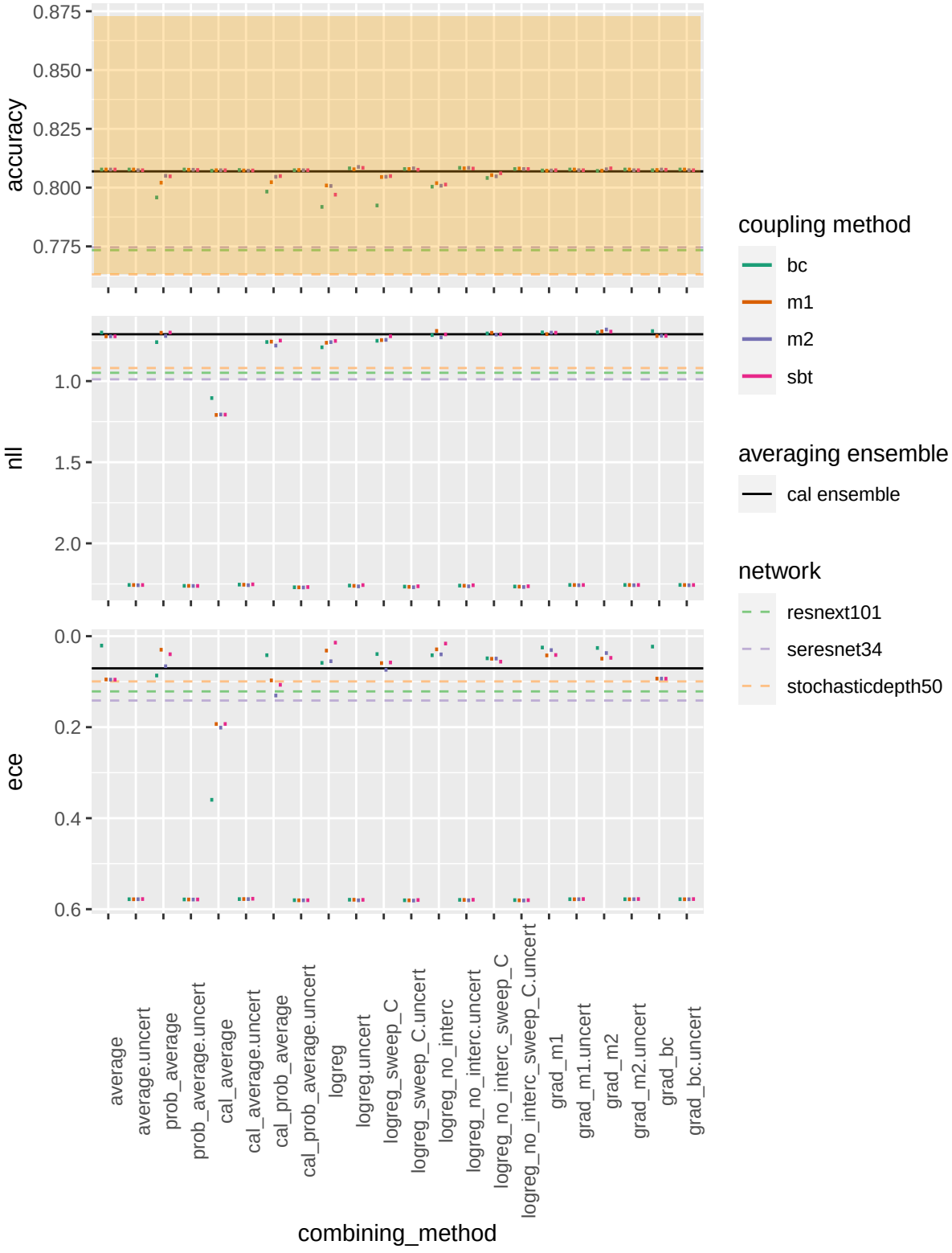
Ensemble metrics
Error inconsistency 0.282299995422363
Average pairwise accuracy variance 3.98788724851329e−05



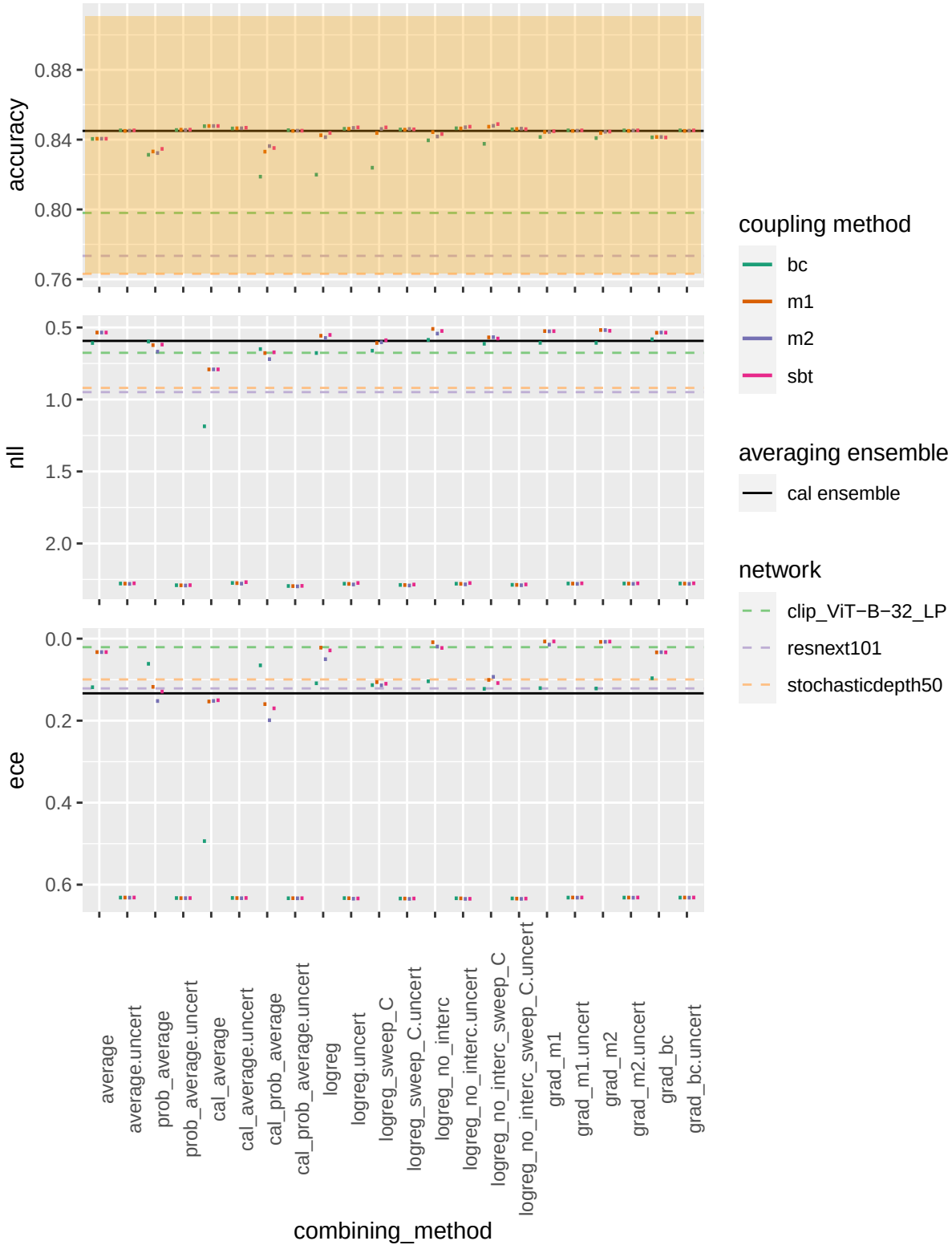
Ensemble metrics
Error inconsistency 0.284999996423721
Average pairwise accuracy variance 3.25644286931492e−05



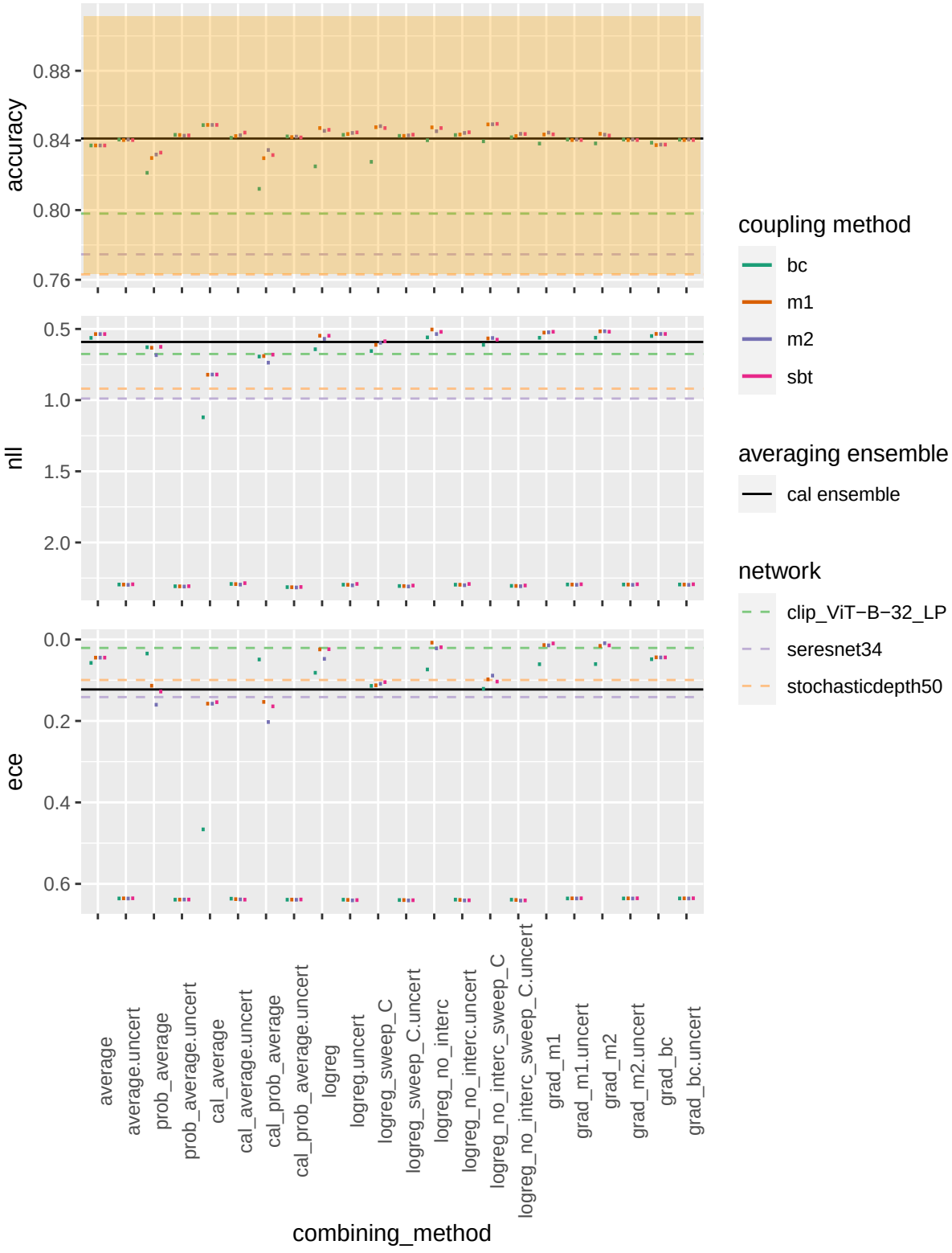
Ensemble metrics
Error inconsistency 0.212799996137619
Average pairwise accuracy variance 2.30799869314069e−05



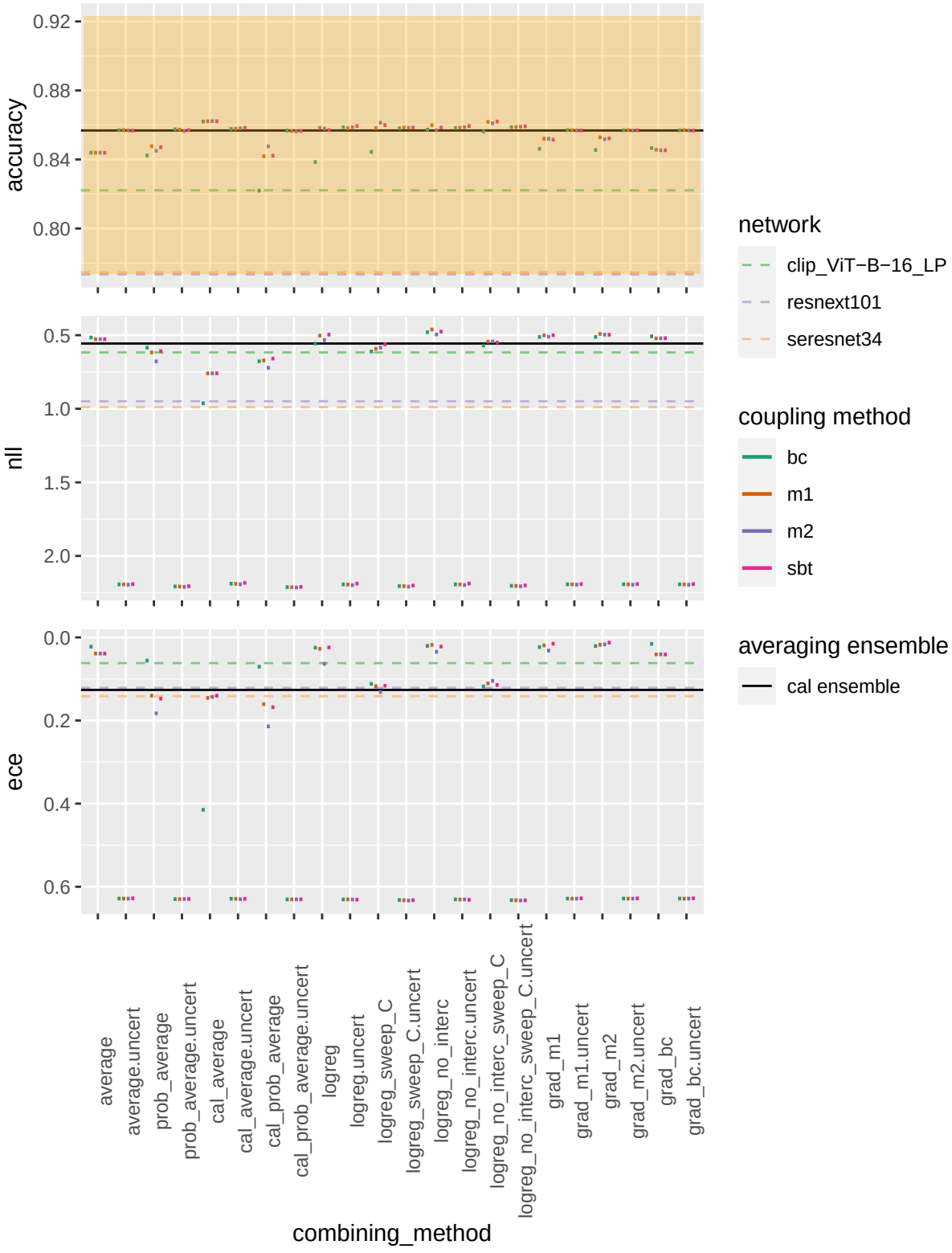
Ensemble metrics
Error inconsistency 0.280099987983704
Average pairwise accuracy variance 3.4432207030477e−05



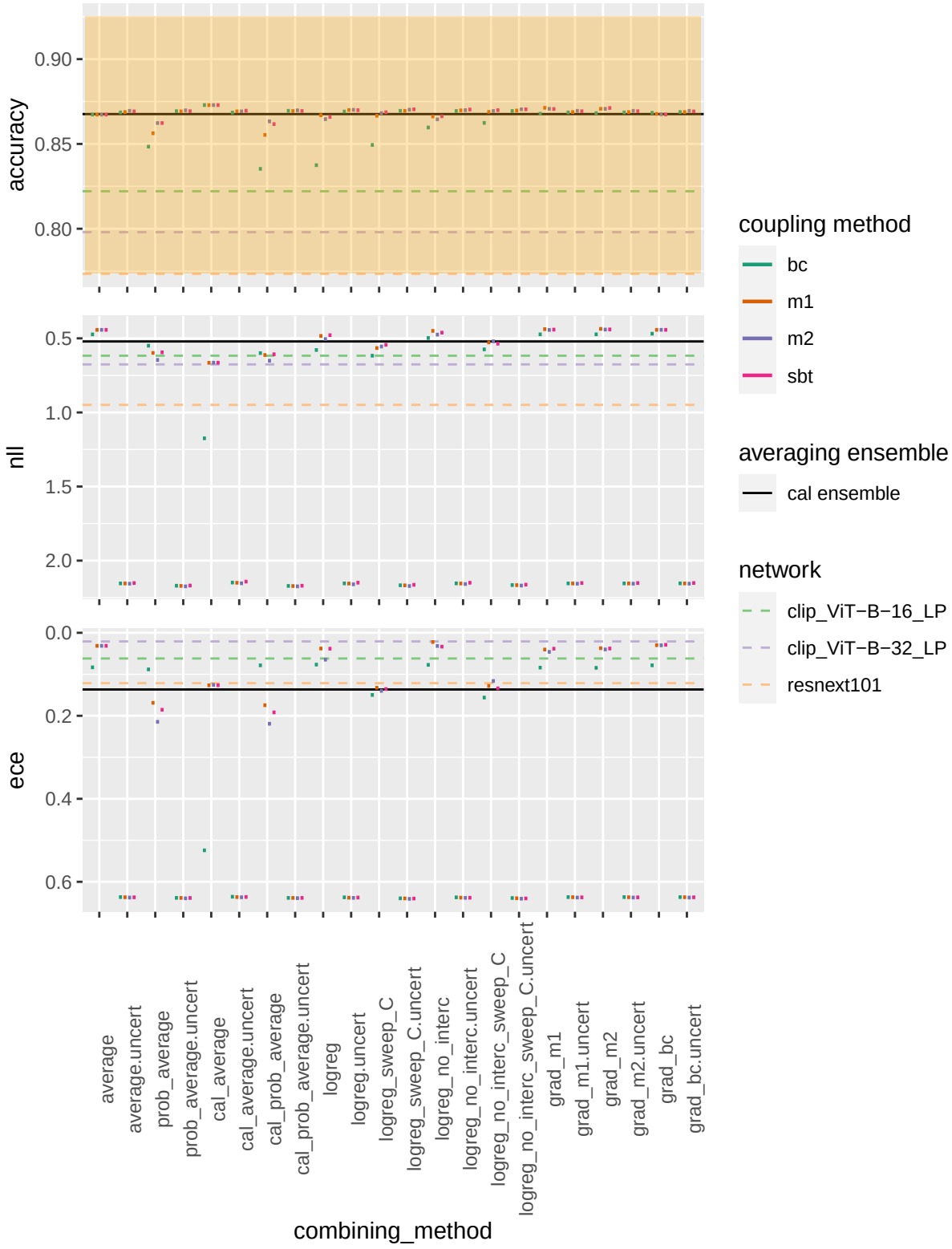
Ensemble metrics
Error inconsistency 0.280000001192093
Average pairwise accuracy variance 3.59233163180761e−05



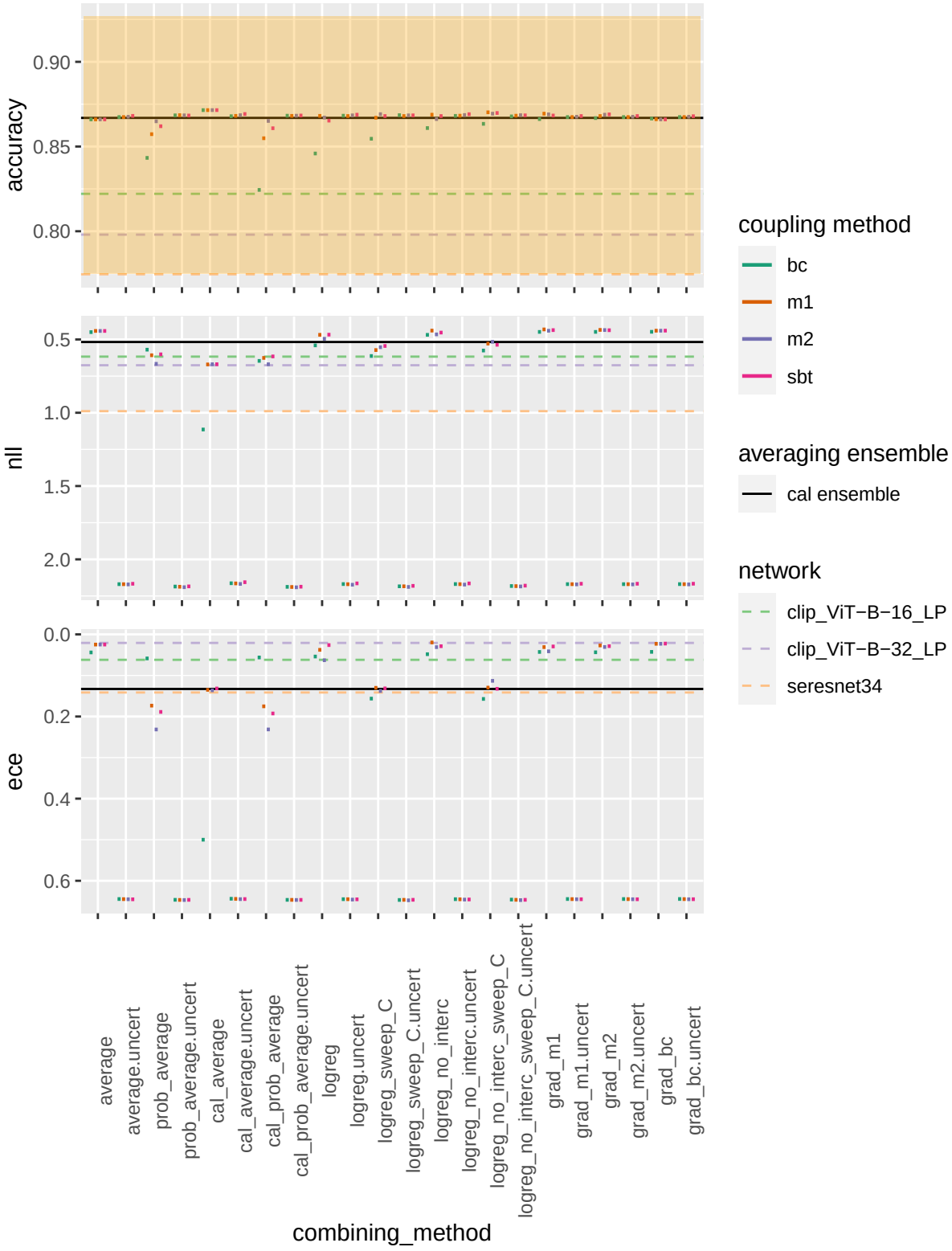
Ensemble metrics
Error inconsistency 0.276600003242493
Average pairwise accuracy variance 4.0245537093142e−05



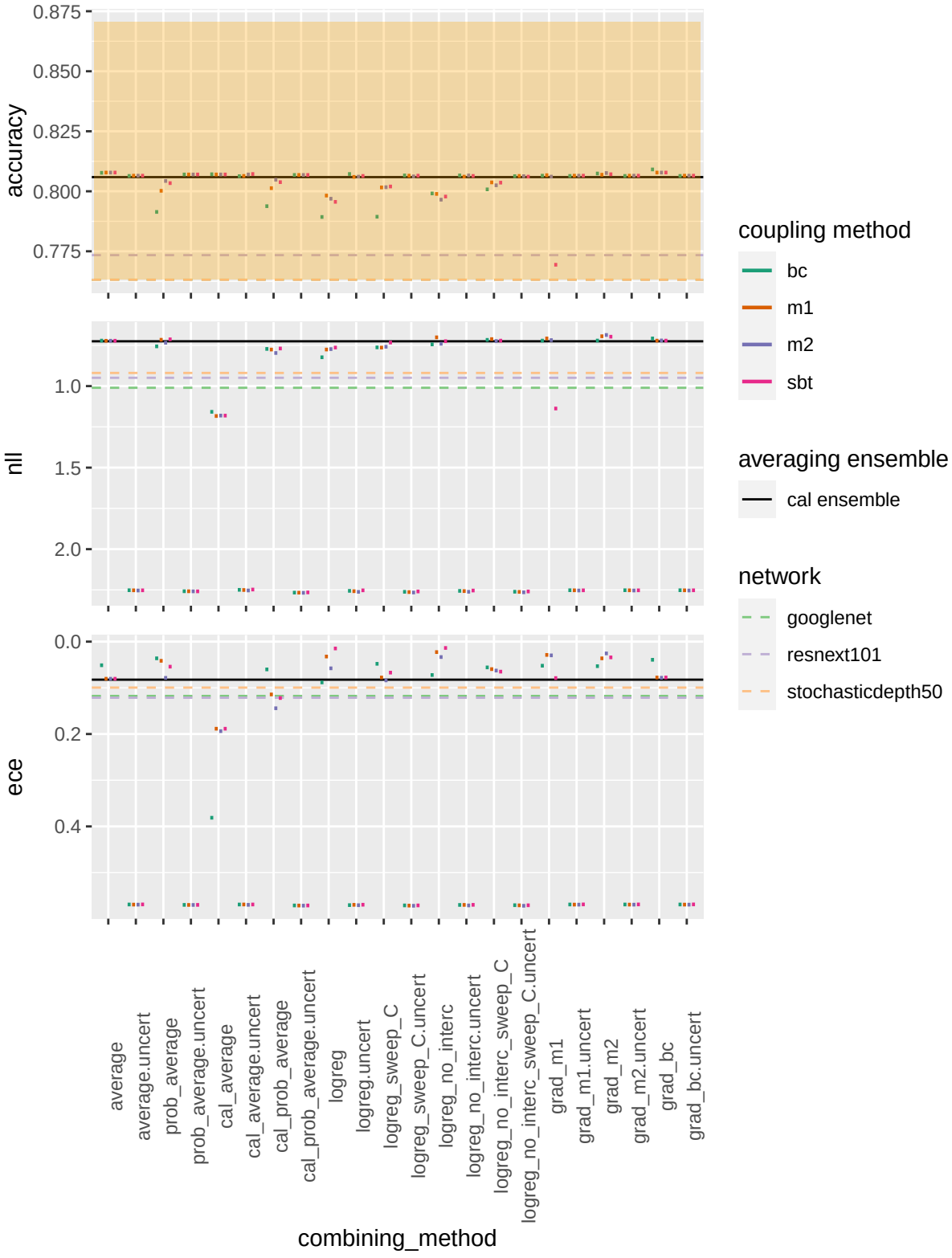
Ensemble metrics
Error inconsistency 0.28099998831749
Average pairwise accuracy variance 3.47010973200668e−05



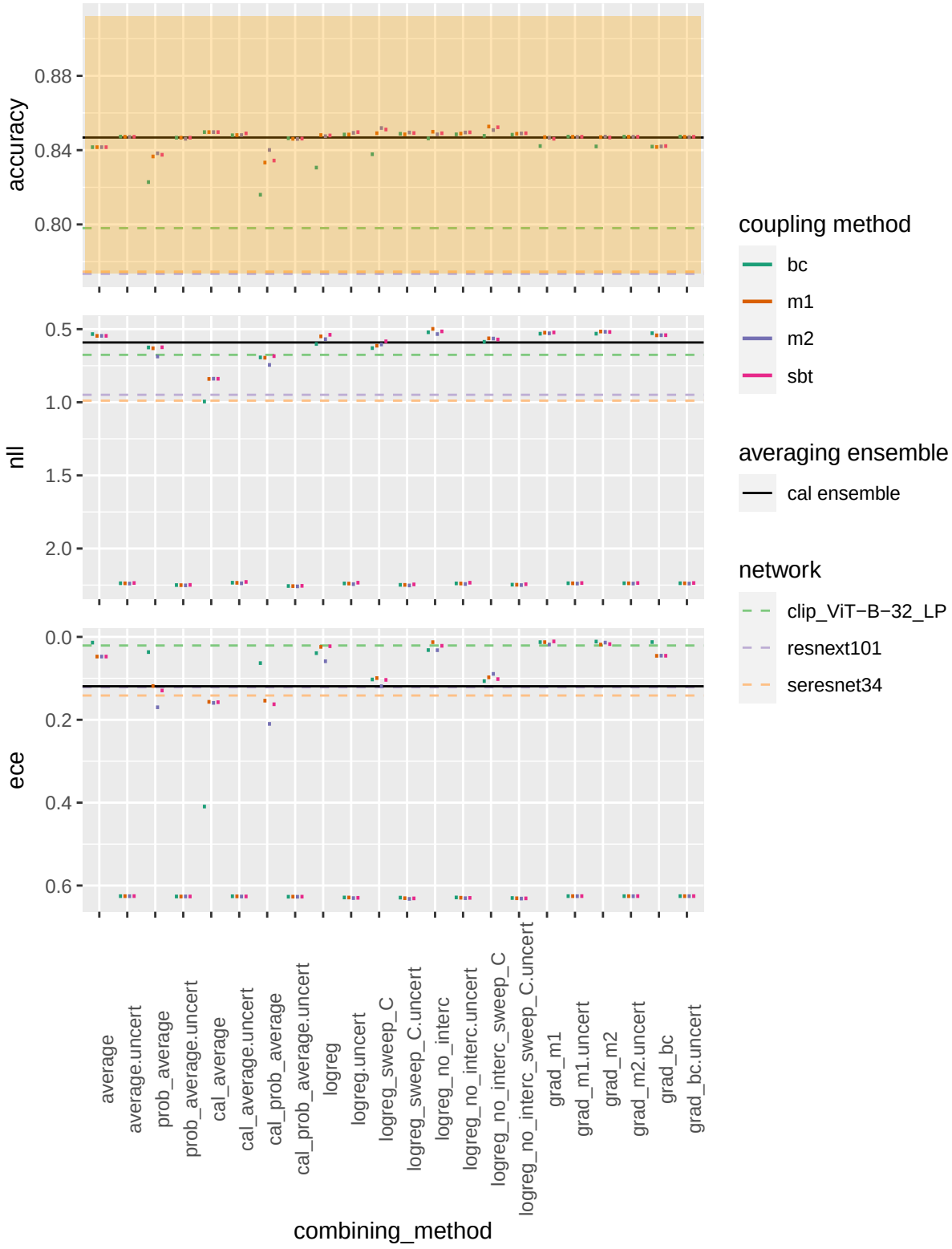
Ensemble metrics
Error inconsistency 0.287099987268448
Average pairwise accuracy variance 3.72233153029811e−05



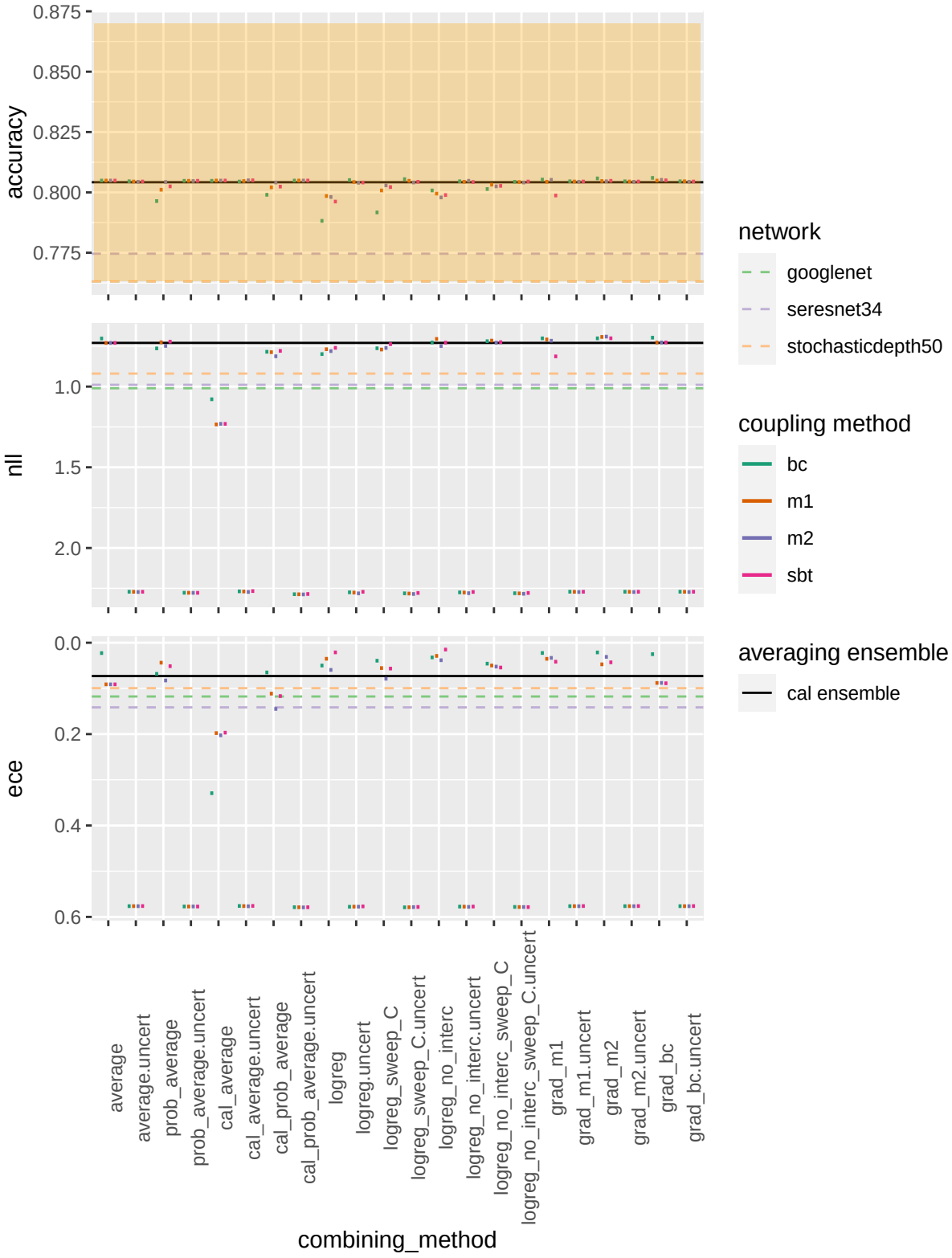
Ensemble metrics
Error inconsistency 0.220100000500679
Average pairwise accuracy variance 2.75677666650154e−05



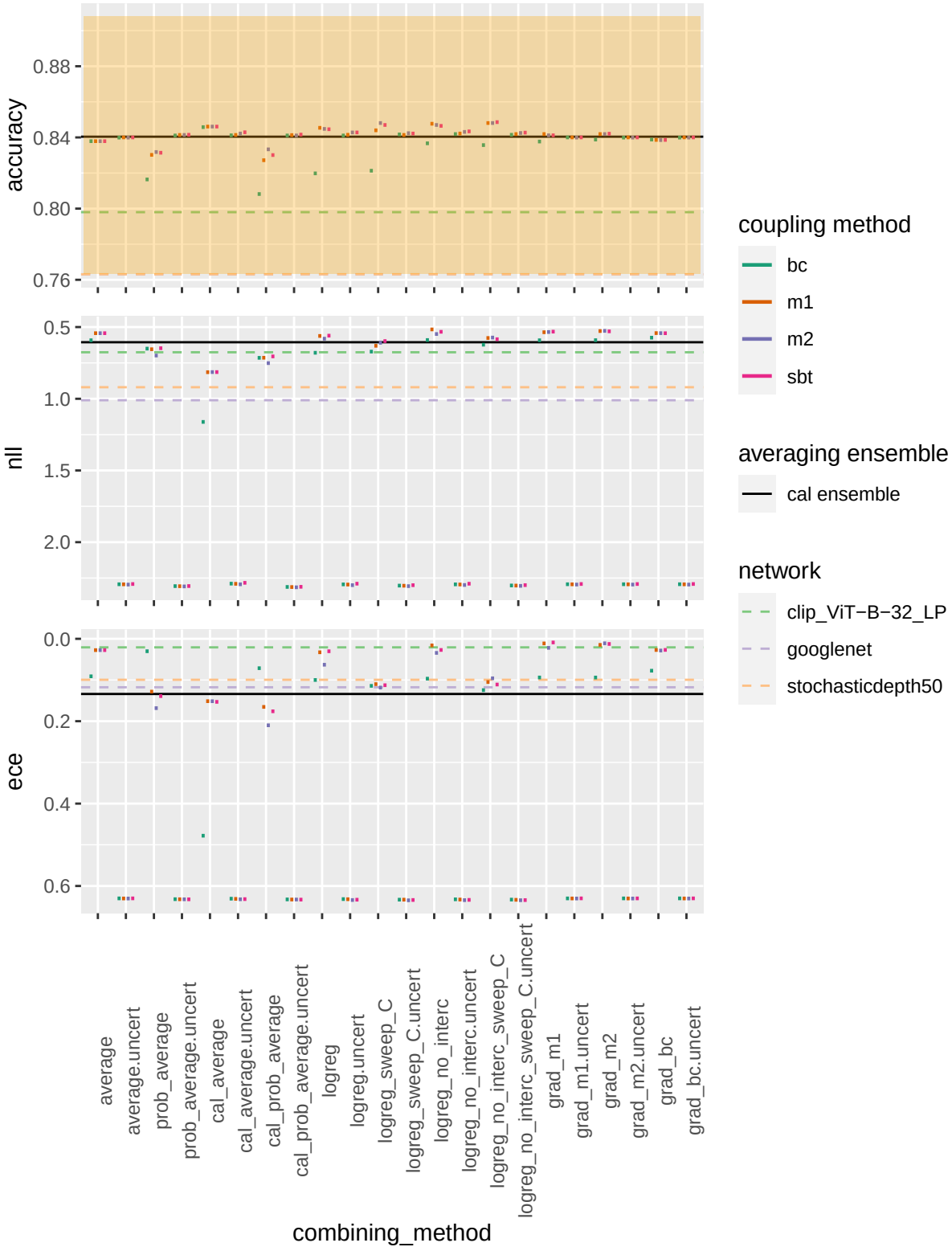
Ensemble metrics
Error inconsistency 0.274399995803833
Average pairwise accuracy variance 3.58788711309899e−05



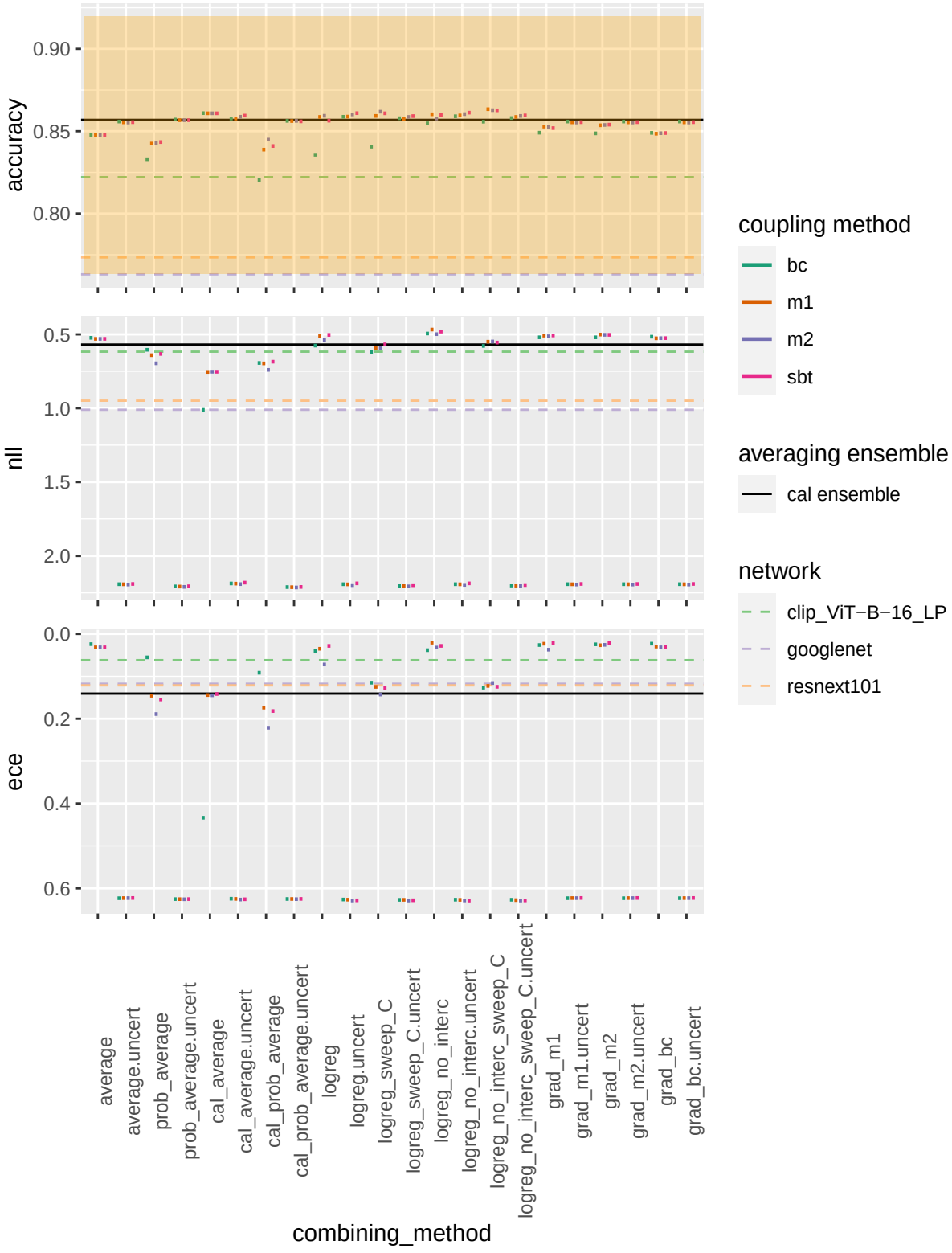
Ensemble metrics
Error inconsistency 0.215799987316132
Average pairwise accuracy variance 2.80744316114578e−05



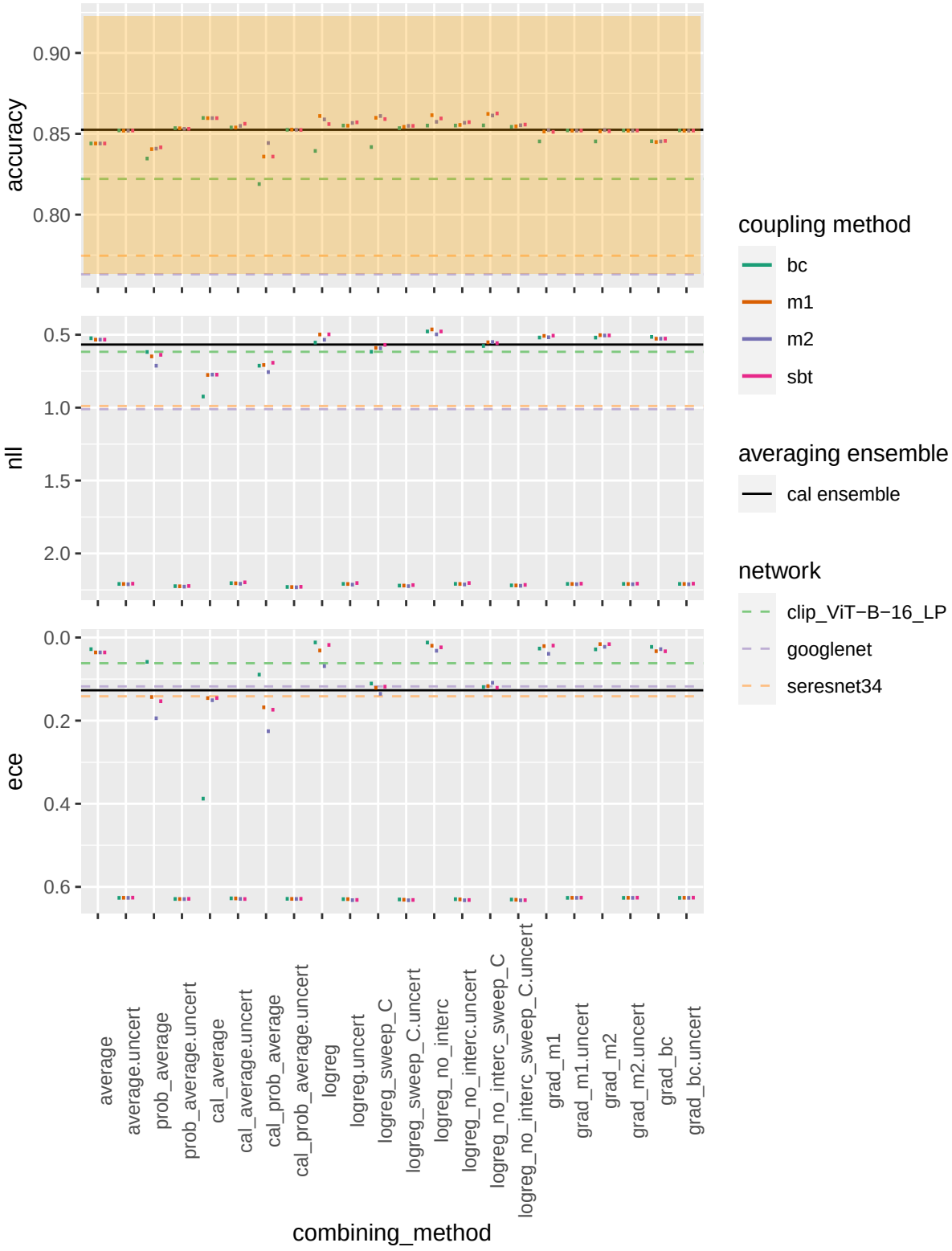
Ensemble metrics
Error inconsistency 0.282700002193451
Average pairwise accuracy variance 4.33588720625266e−05



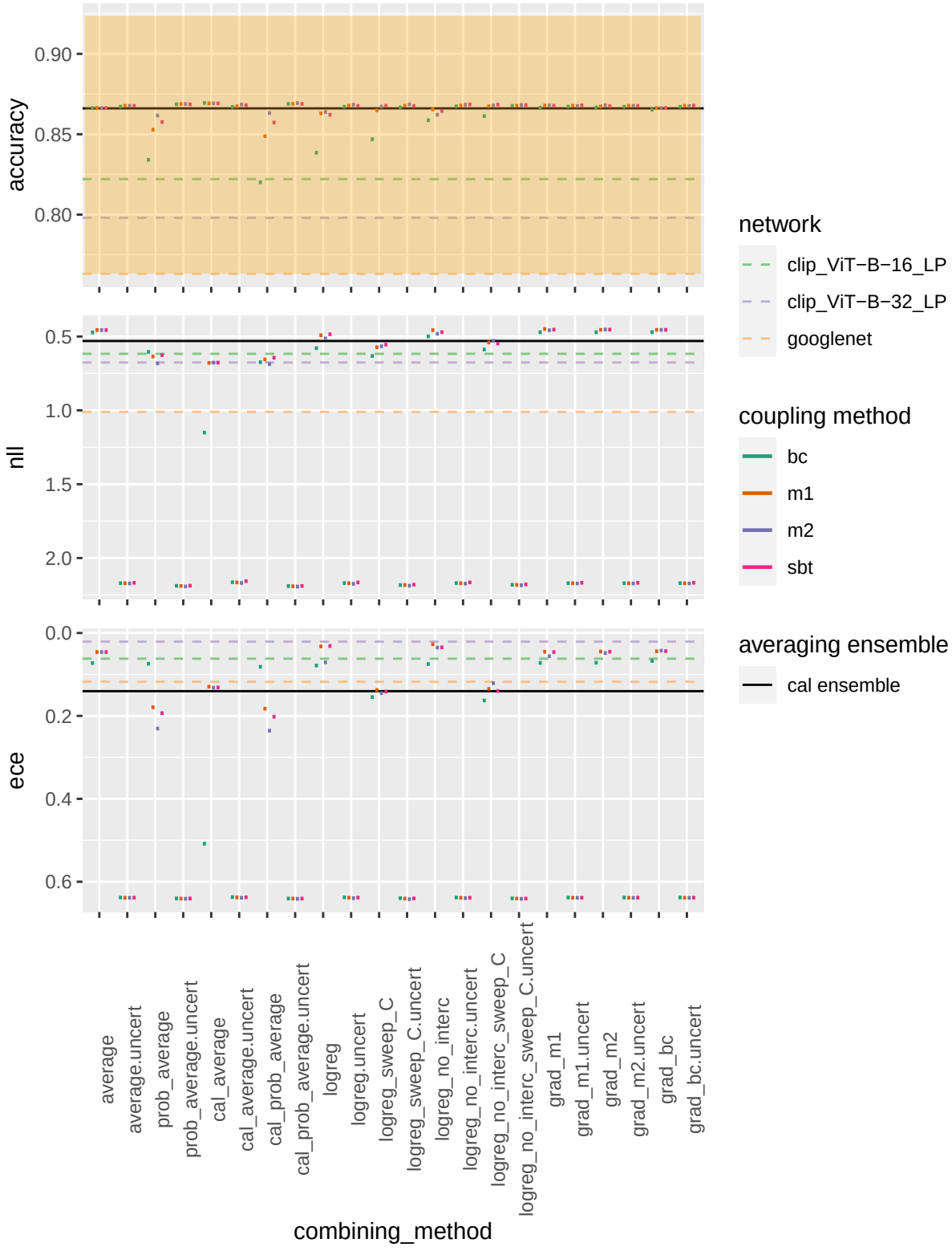
Ensemble metrics
Error inconsistency 0.281699985265732
Average pairwise accuracy variance 4.93010920763481e−05



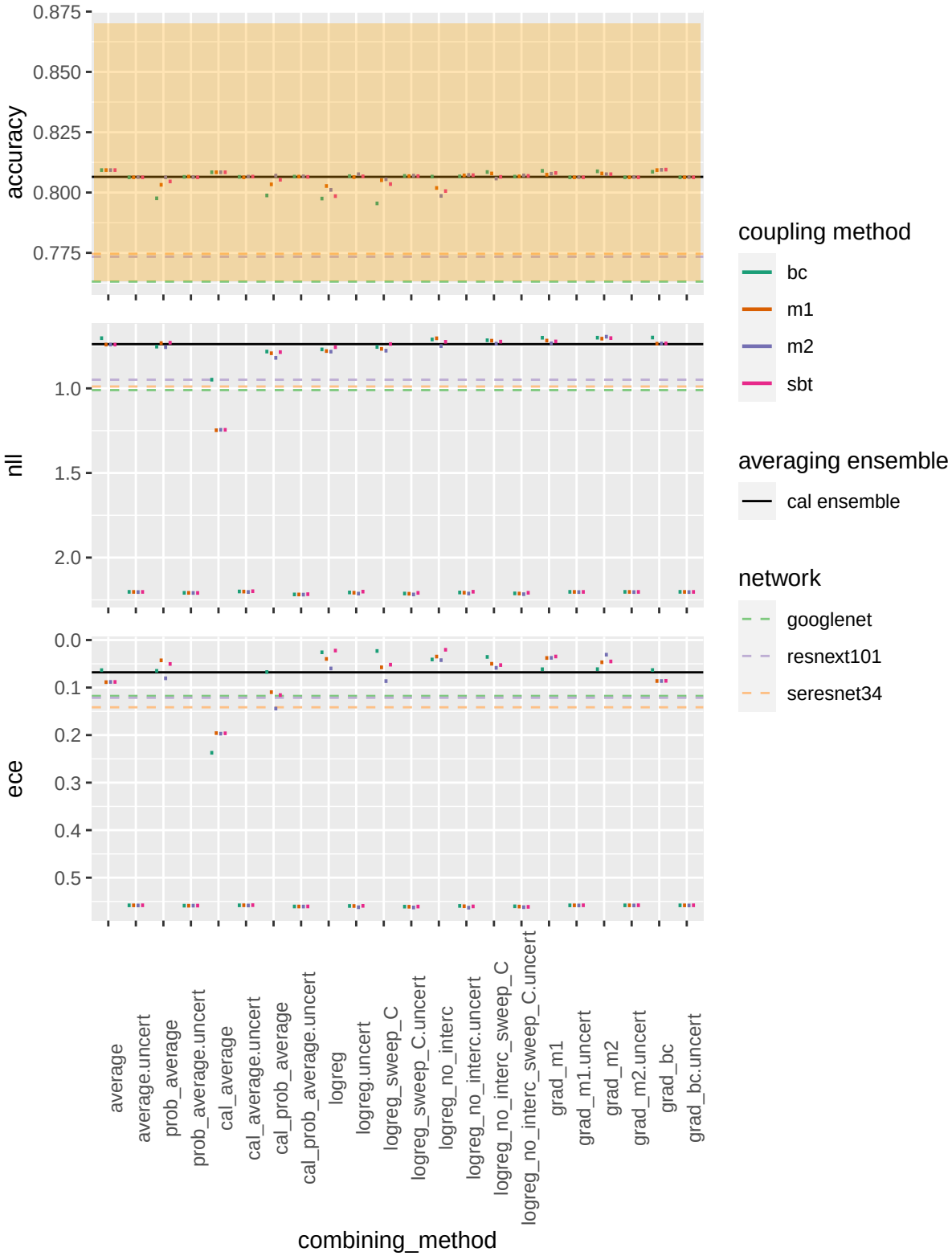
Ensemble metrics
Error inconsistency 0.283600002527237
Average pairwise accuracy variance 5.08388693560846e−05



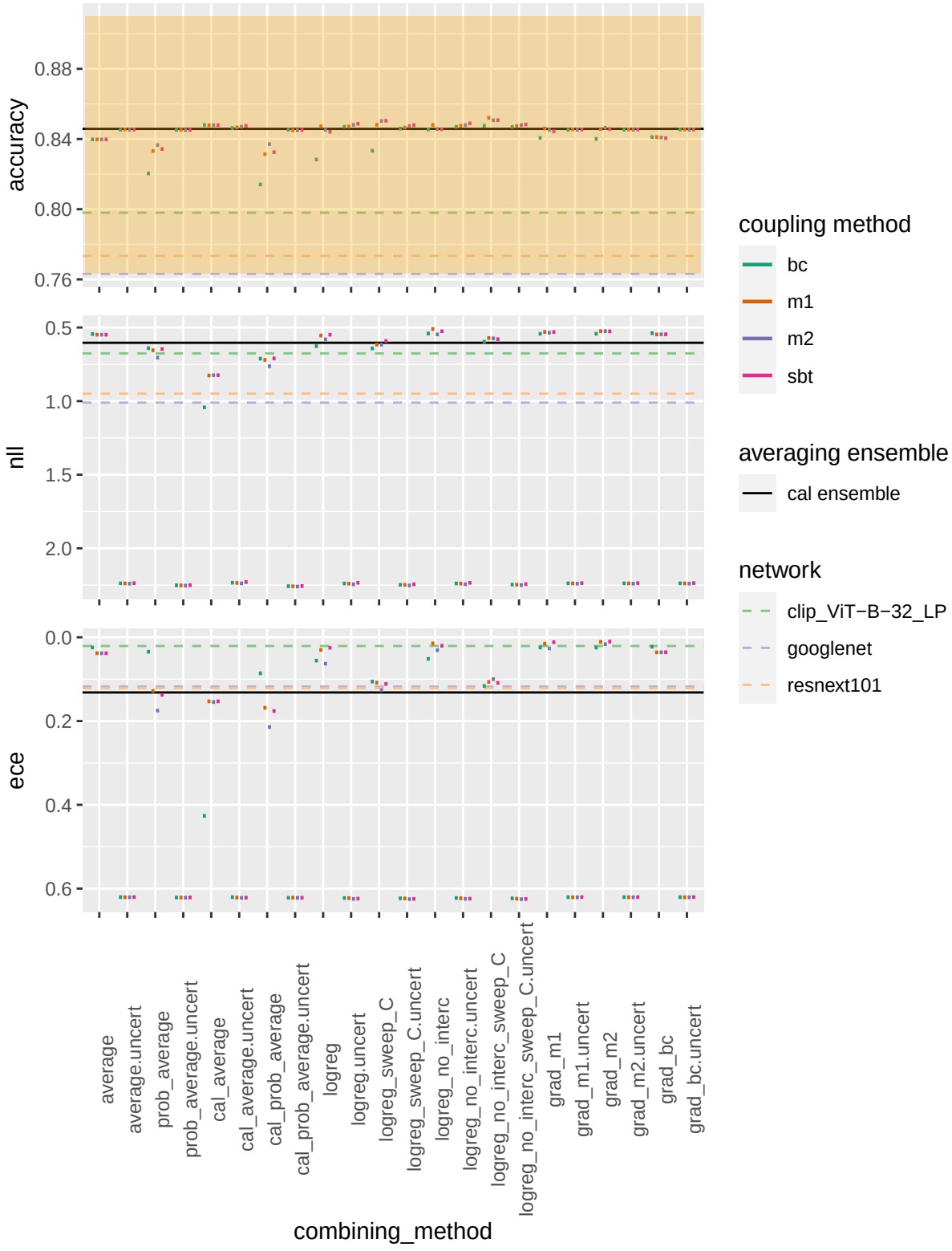
Ensemble metrics
Error inconsistency 0.287599980831146
Average pairwise accuracy variance 4.92266481160186e−05



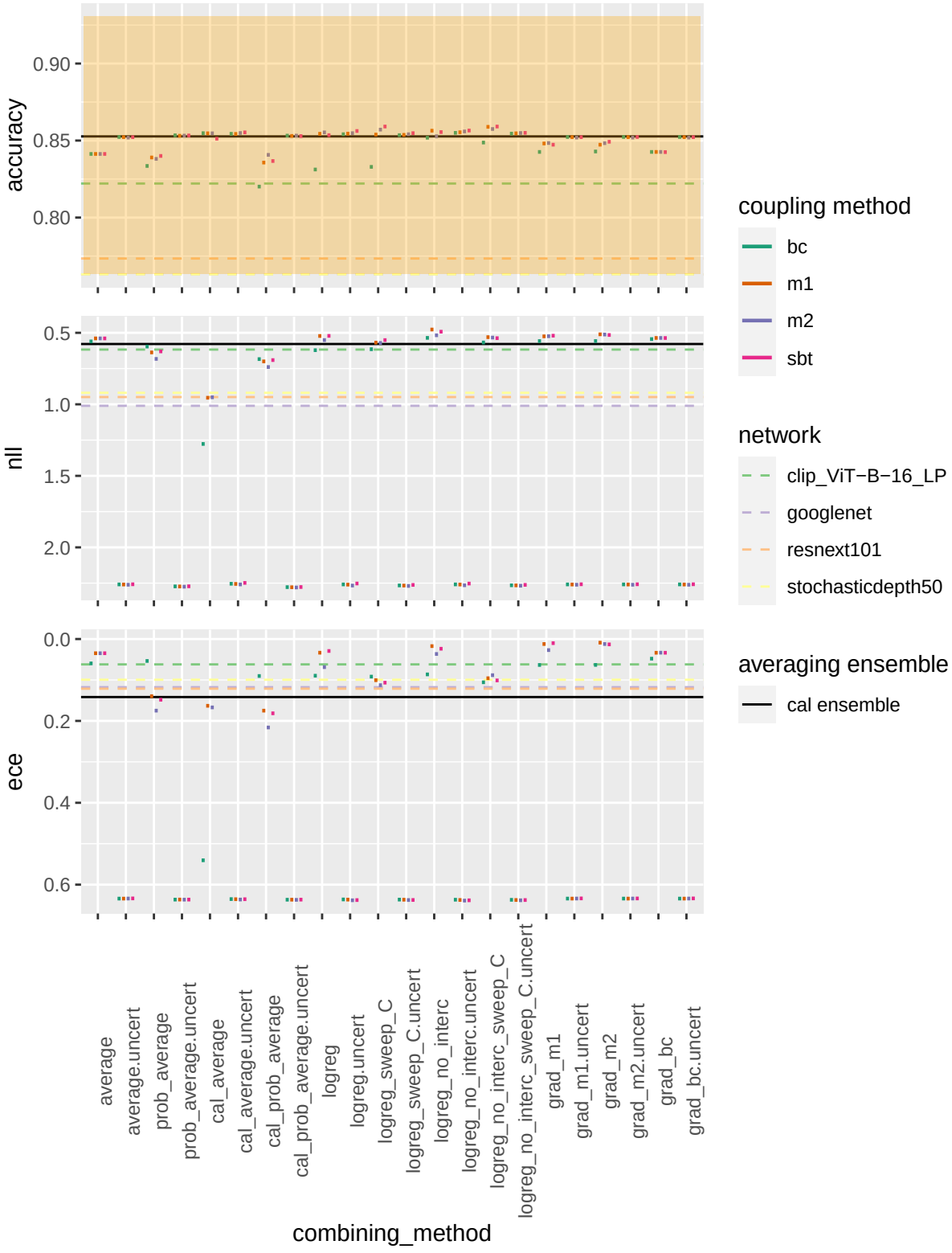
Ensemble metrics
Error inconsistency 0.209099993109703
Average pairwise accuracy variance 2.65044345724164e−05



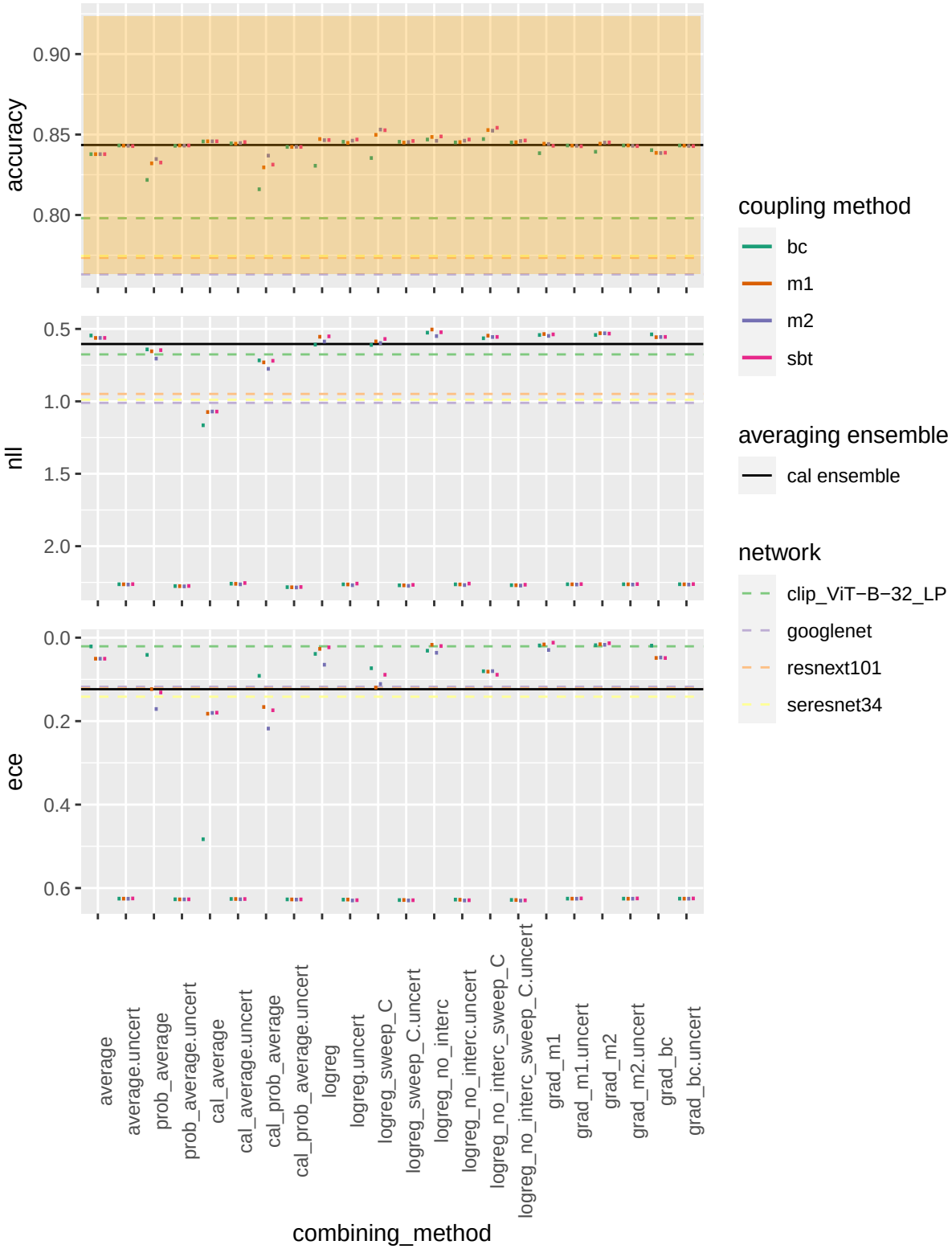
Ensemble metrics
Error inconsistency 0.277700006961823
Average pairwise accuracy variance 4.35588772234041e−05



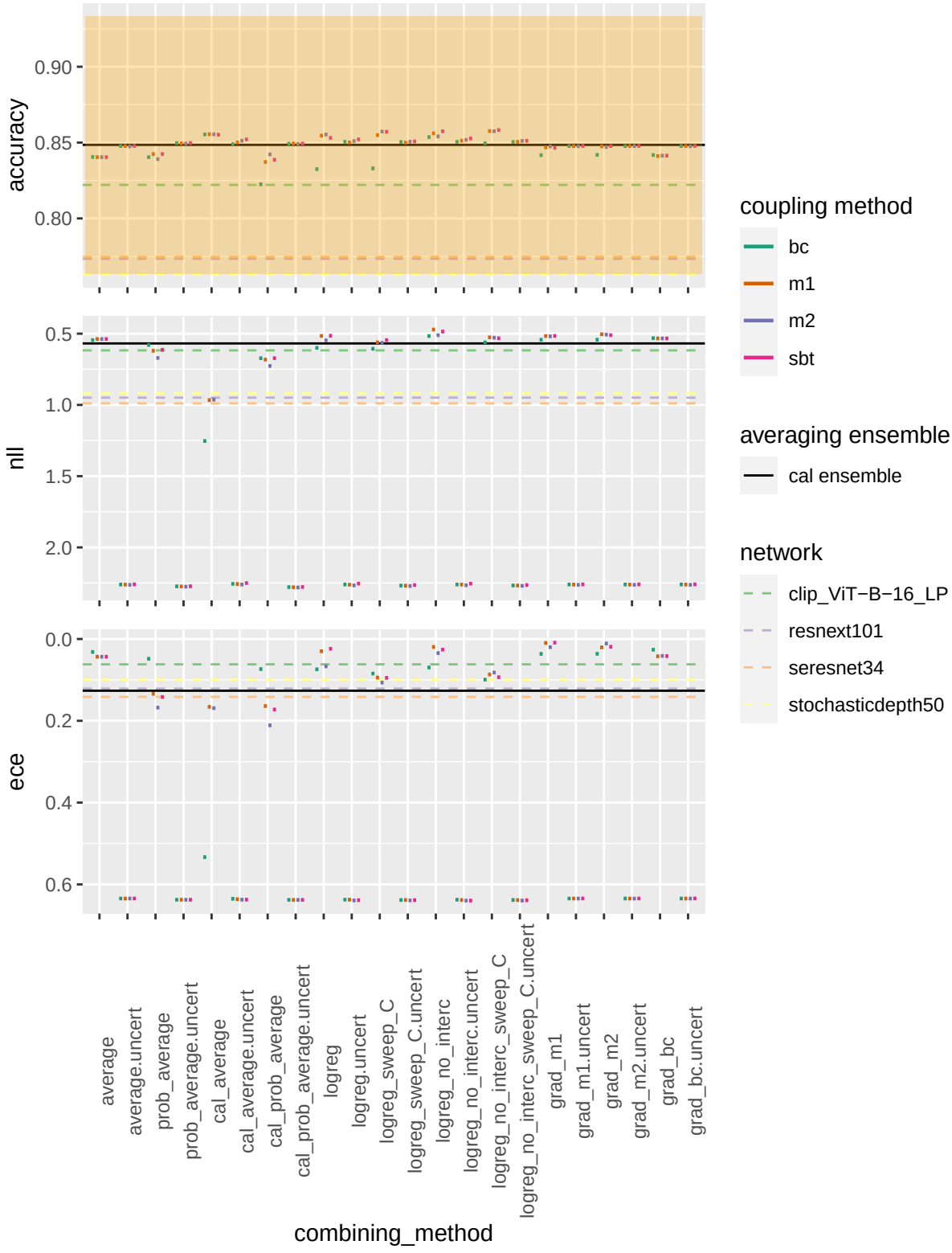
Ensemble metrics
Error inconsistency 0.326900005340576
Average pairwise accuracy variance 4.59987313661259e−05



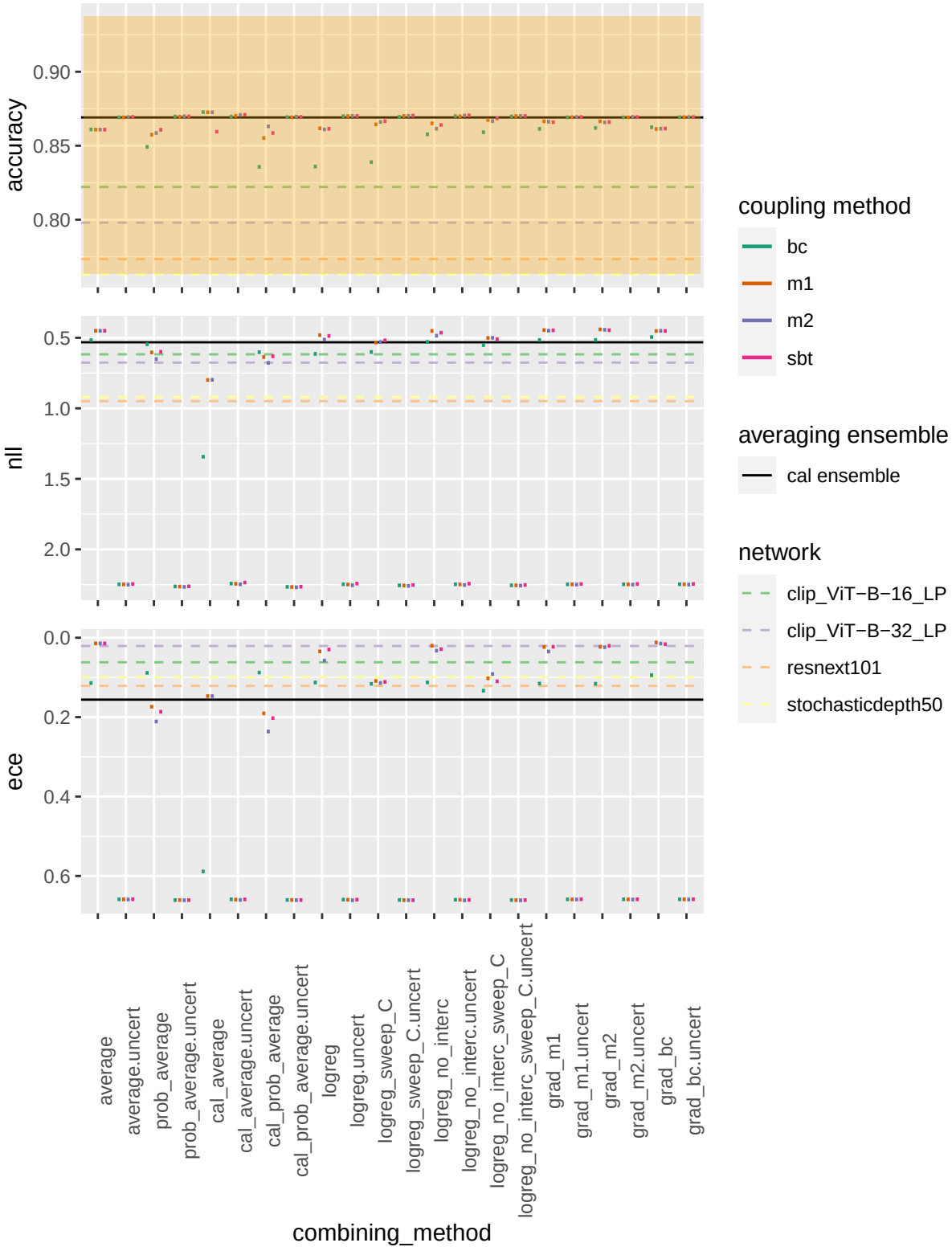
Ensemble metrics
Error inconsistency 0.319399982690811
Average pairwise accuracy variance 4.23684214183595e−05



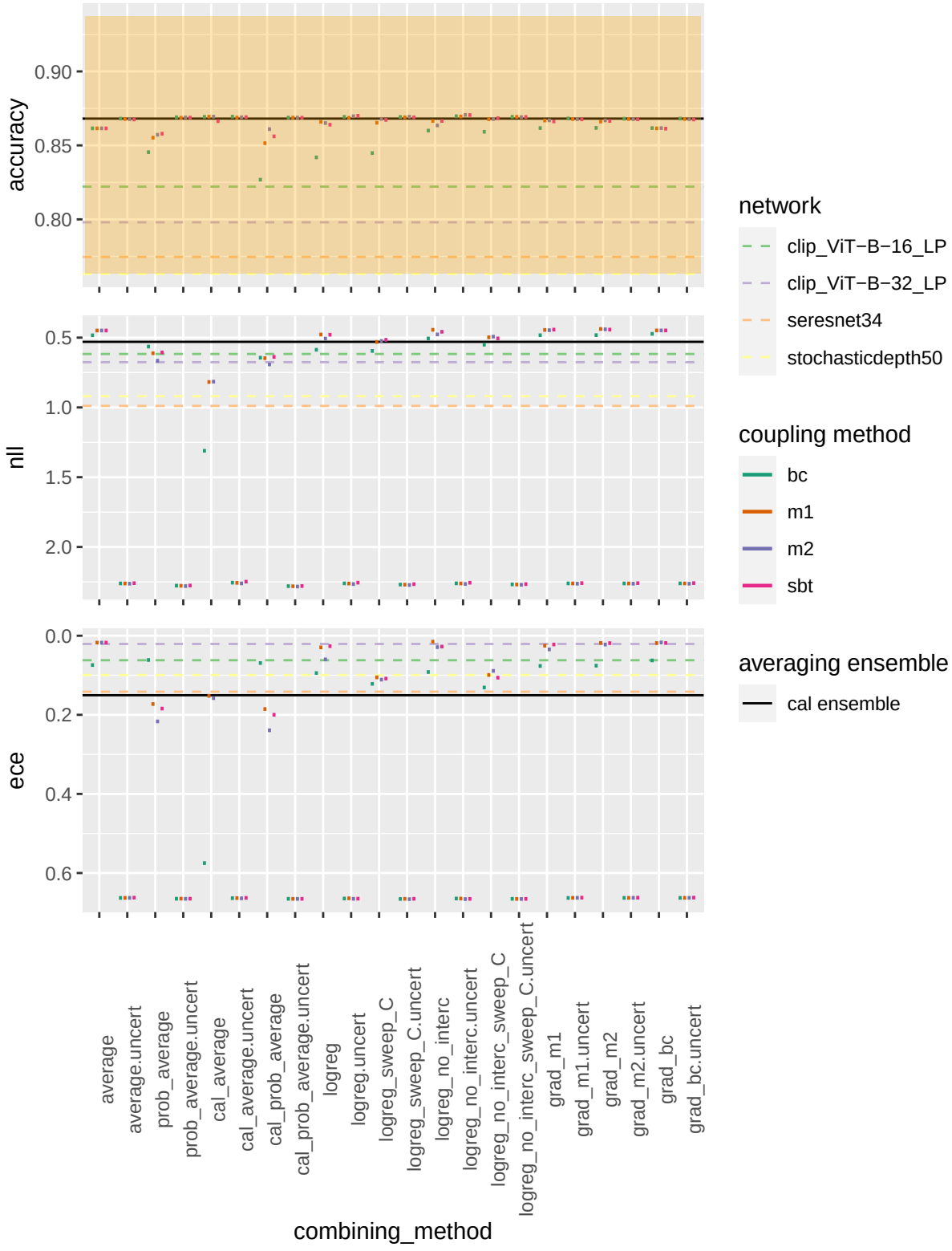
Ensemble metrics
Error inconsistency 0.321399986743927
Average pairwise accuracy variance 3.97115436499007e−05



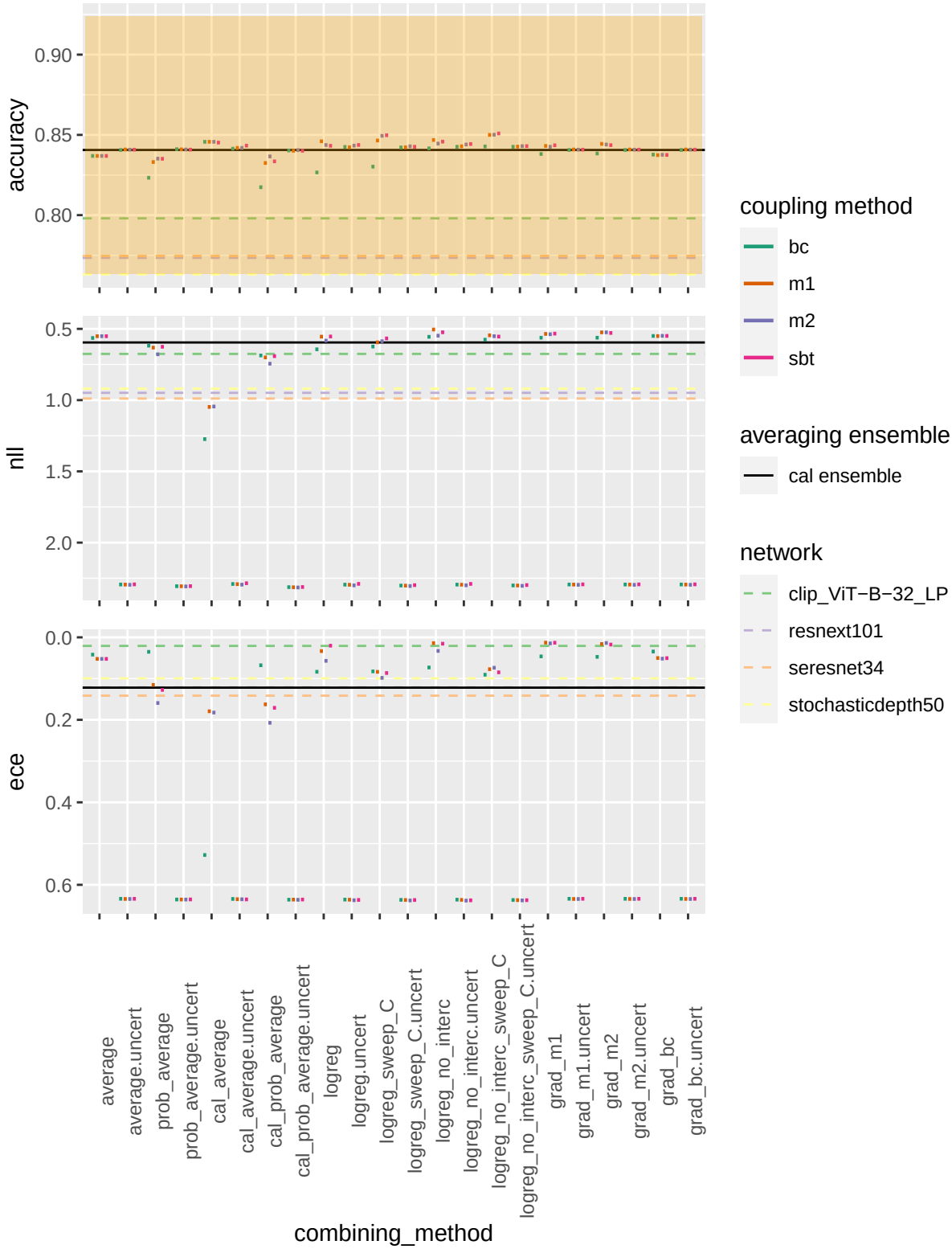
Ensemble metrics
Error inconsistency 0.337999999523163
Average pairwise accuracy variance 3.92877918784507e−05



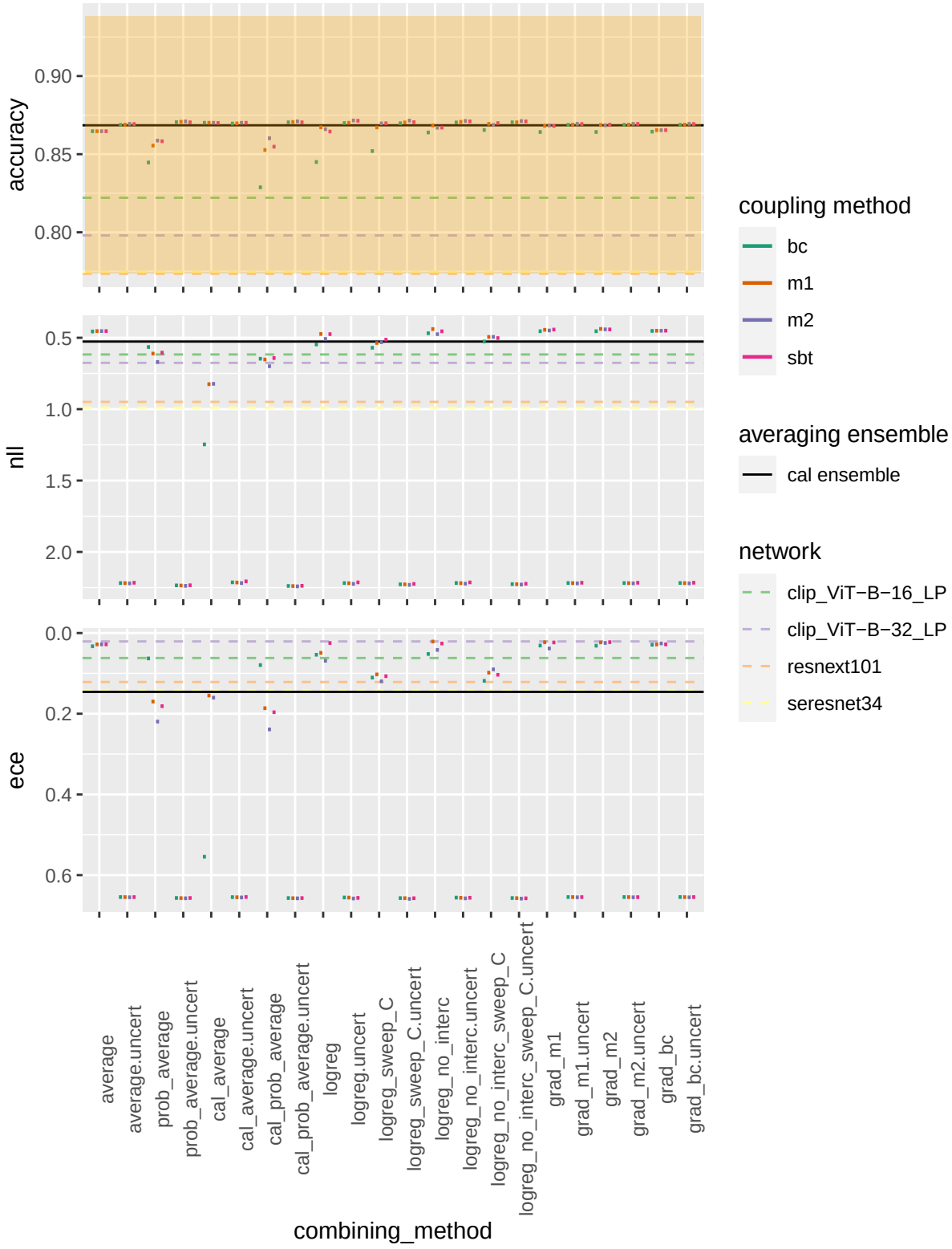
Ensemble metrics
Error inconsistency 0.338499993085861
Average pairwise accuracy variance 4.09471685998142e−05



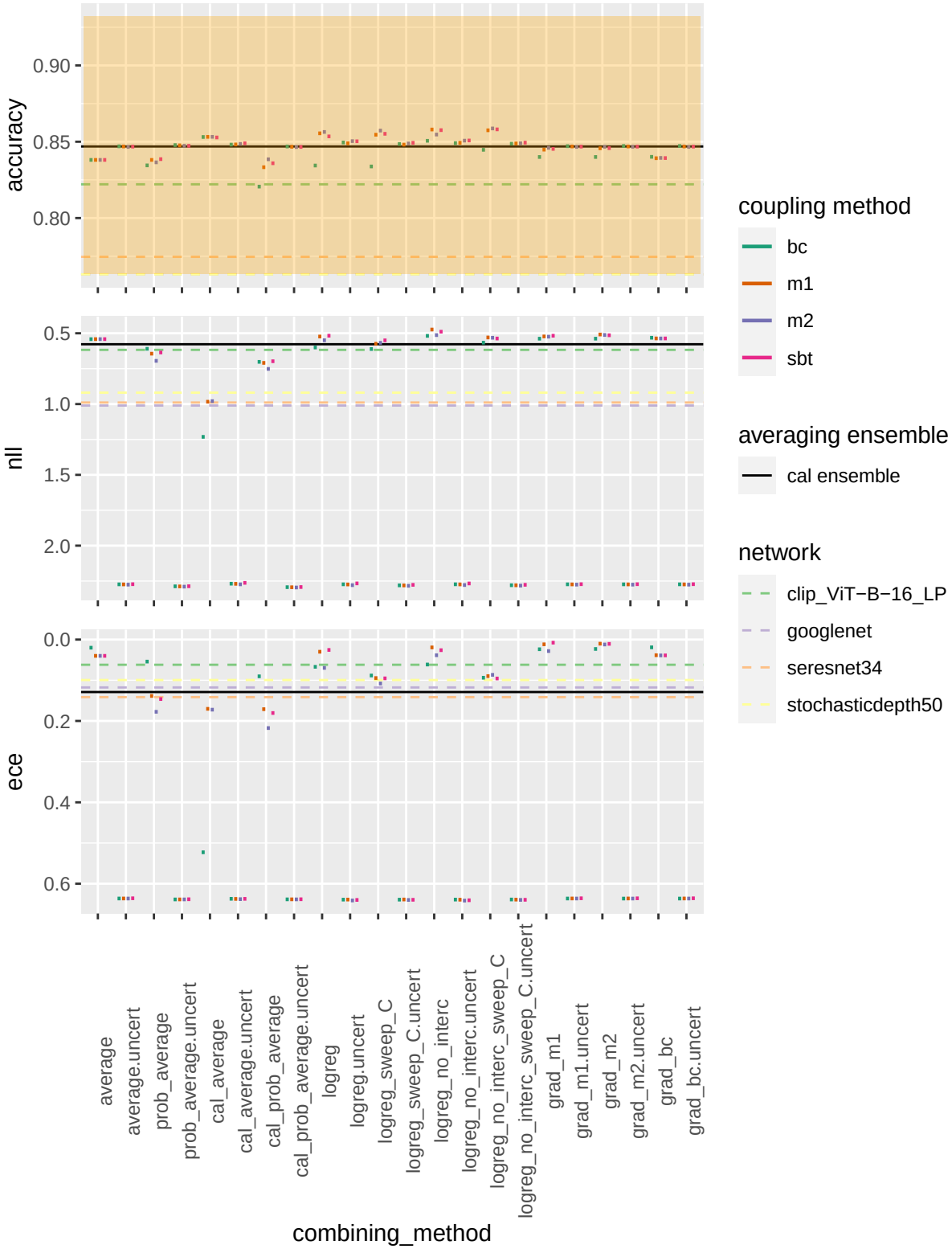
Ensemble metrics
Error inconsistency 0.319199979305267
Average pairwise accuracy variance 3.63696744898334e−05



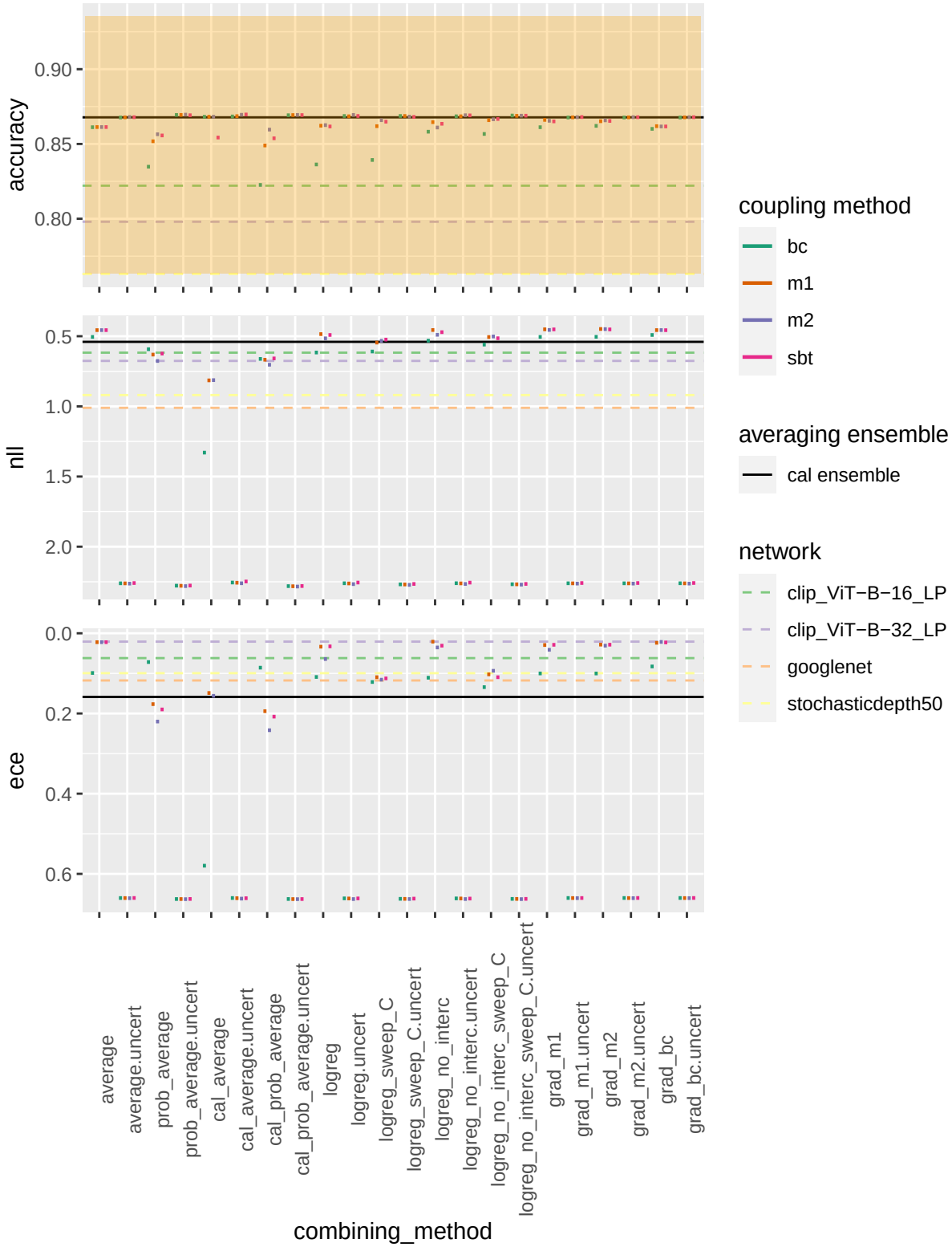
Ensemble metrics
Error inconsistency 0.333099991083145
Average pairwise accuracy variance 4.16387301811483e−05



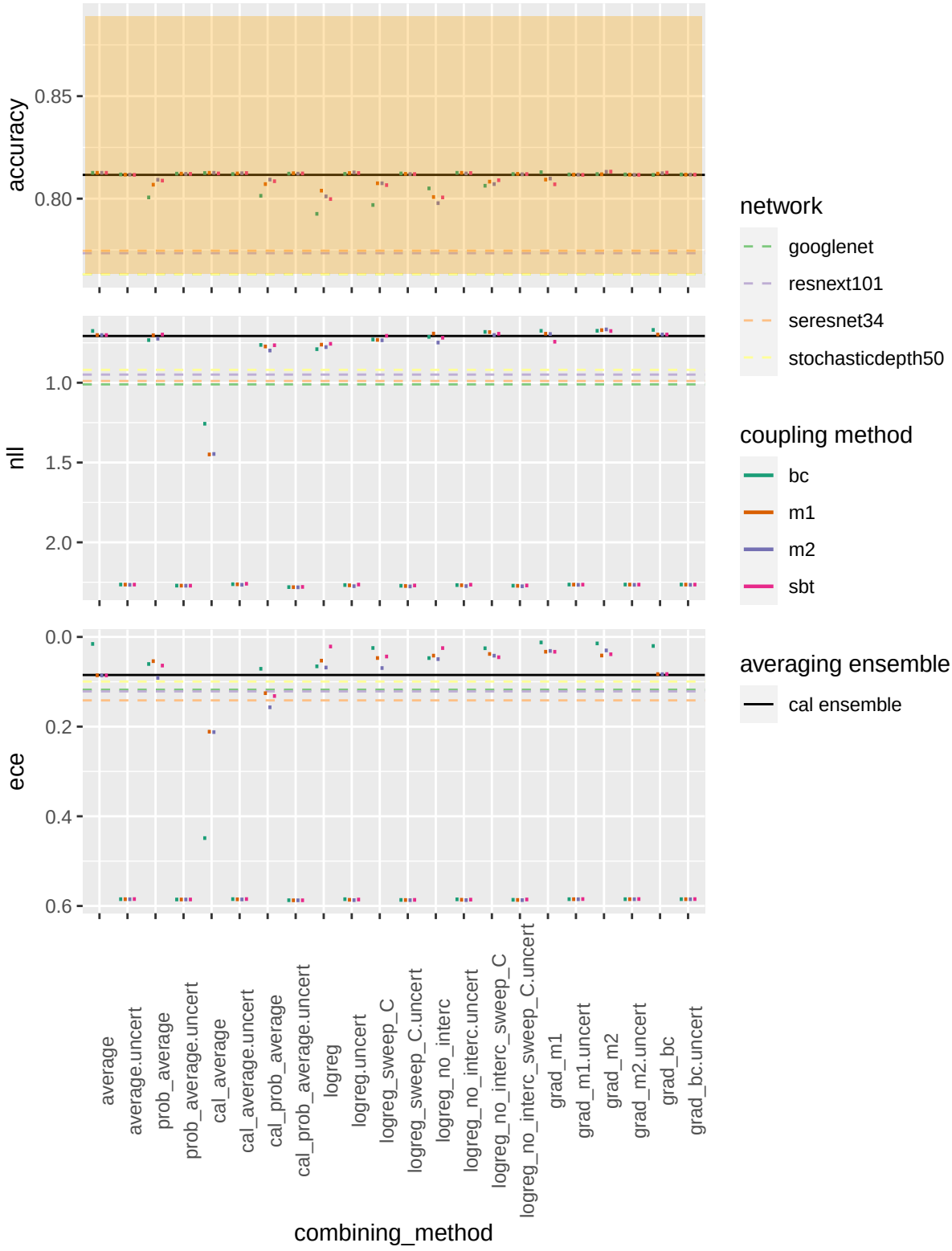
Ensemble metrics
Error inconsistency 0.326499998569489
Average pairwise accuracy variance 4.71043567813467e−05



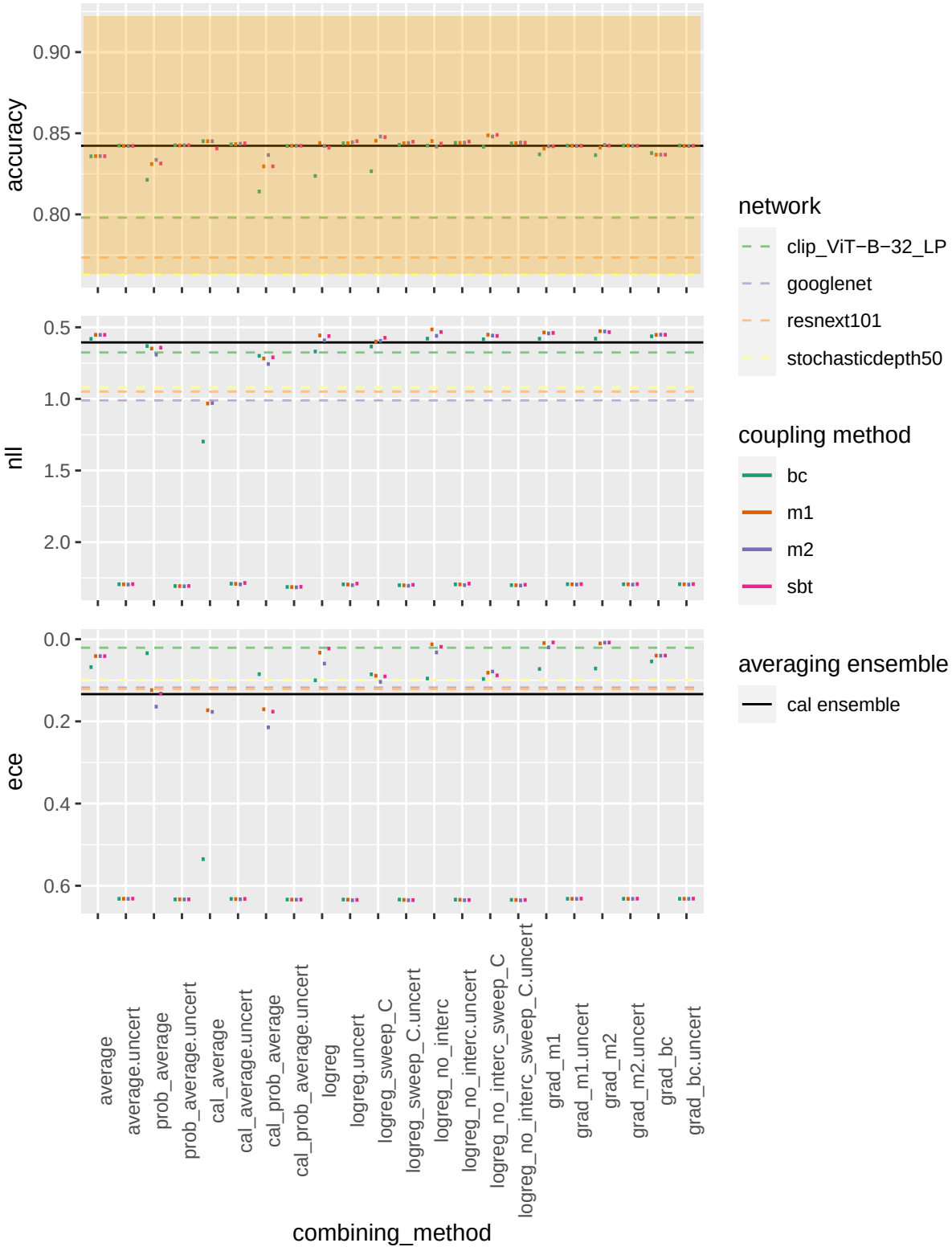
Ensemble metrics
Error inconsistency 0.341999977827072
Average pairwise accuracy variance 4.88924815726932e−05



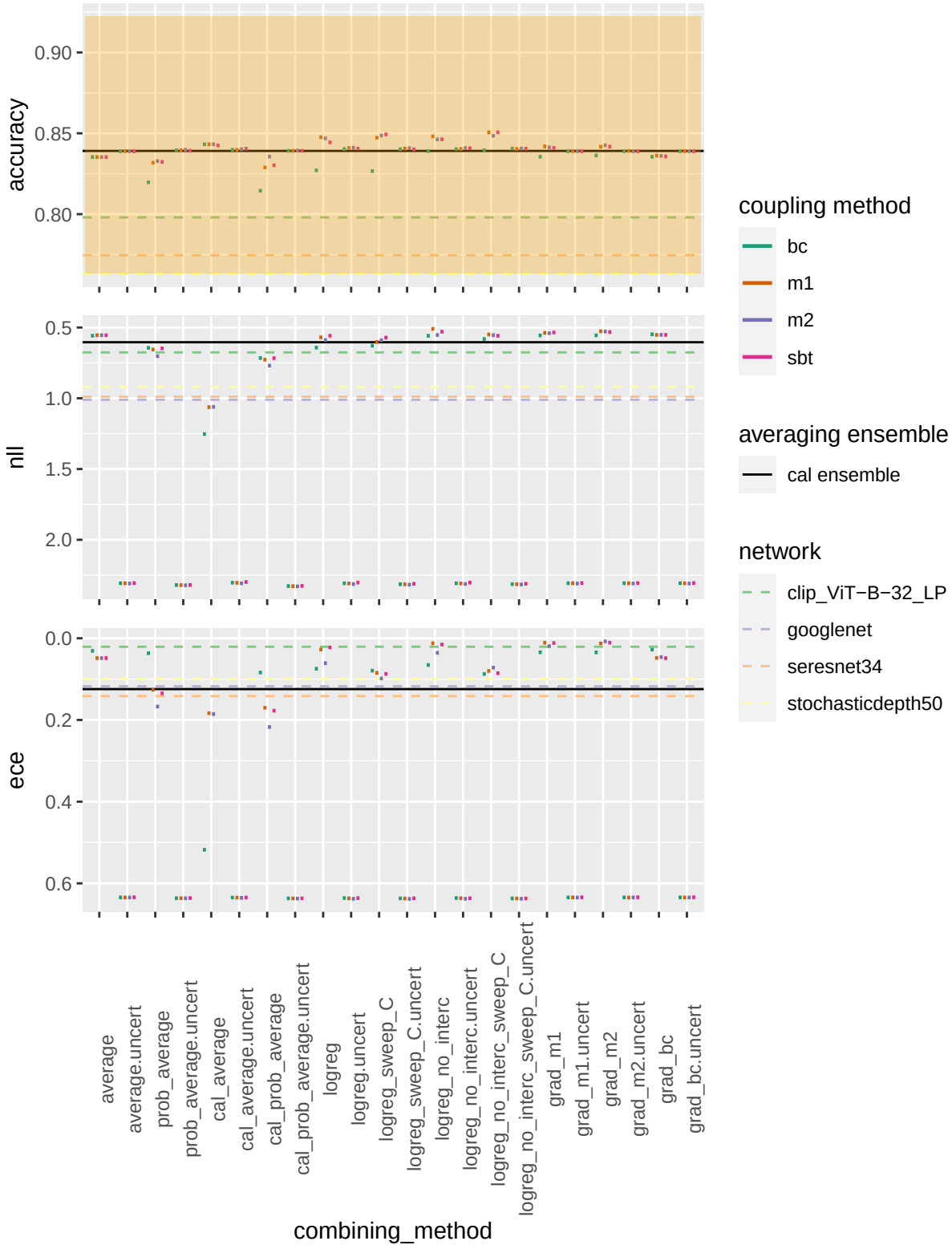
Ensemble metrics
Error inconsistency 0.260600000619888
Average pairwise accuracy variance 2.95949866995215e−05



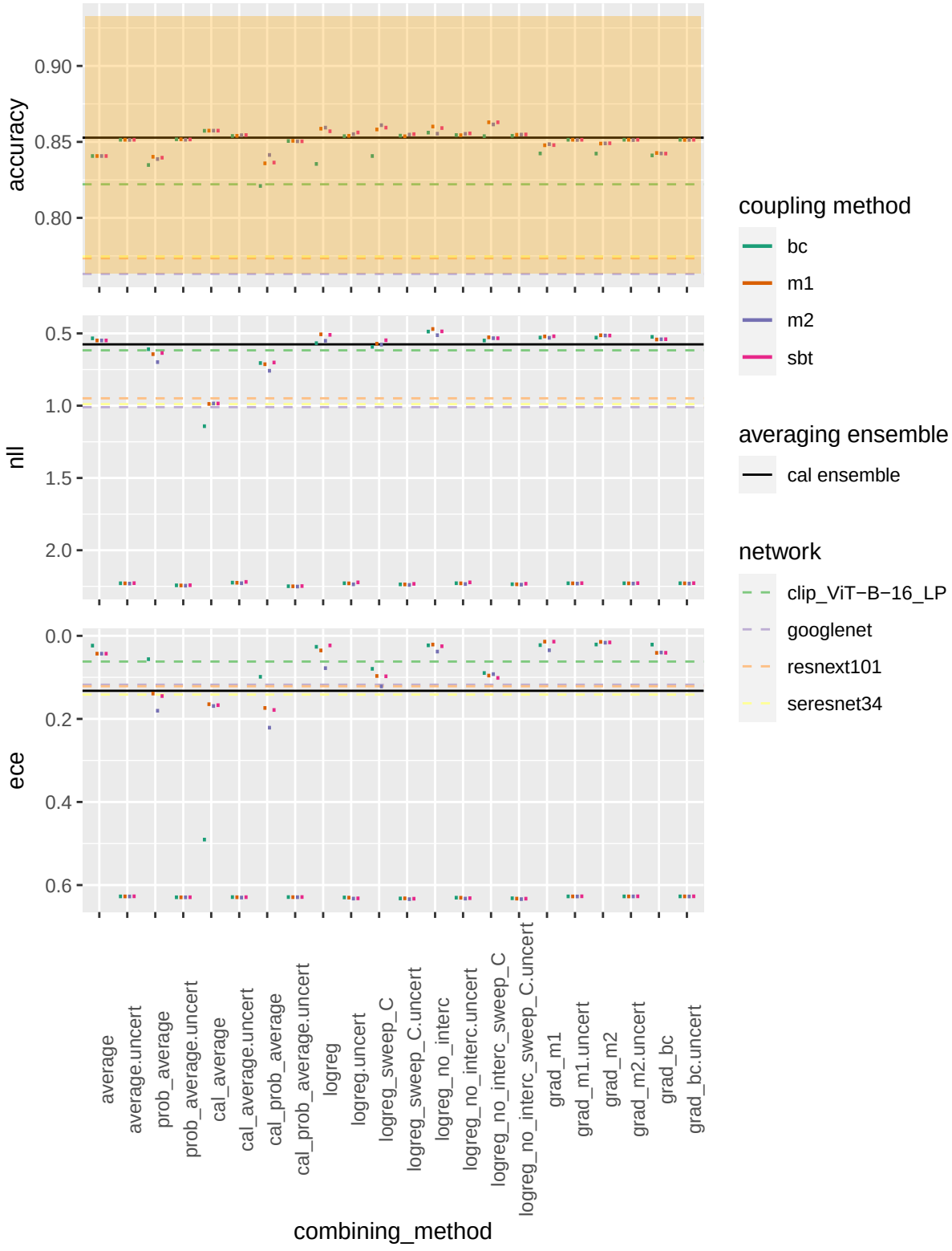
Ensemble metrics
Error inconsistency 0.324099987745285
Average pairwise accuracy variance 4.18831114075147e−05



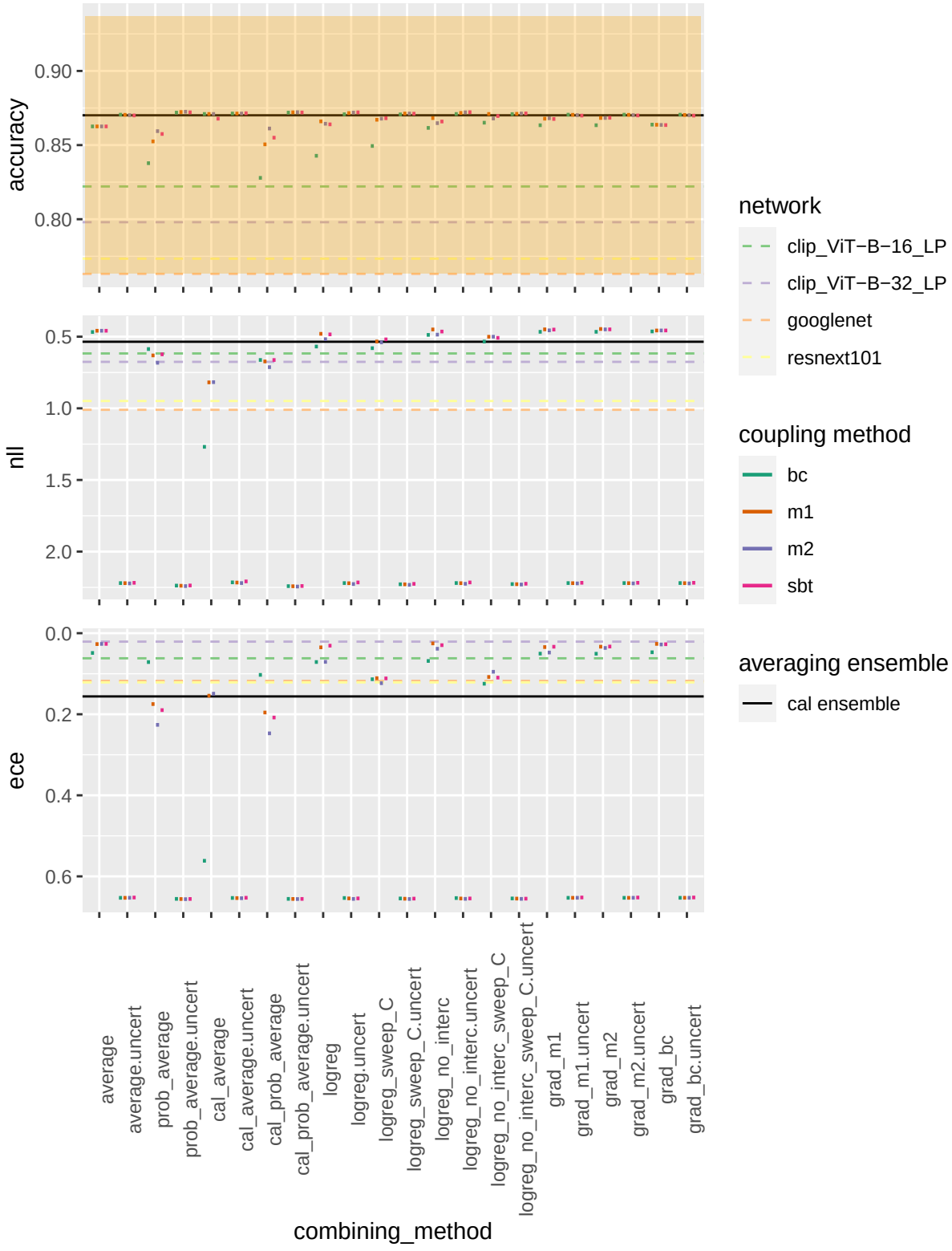
Ensemble metrics
Error inconsistency 0.323499977588654
Average pairwise accuracy variance 4.2766234400915e−05



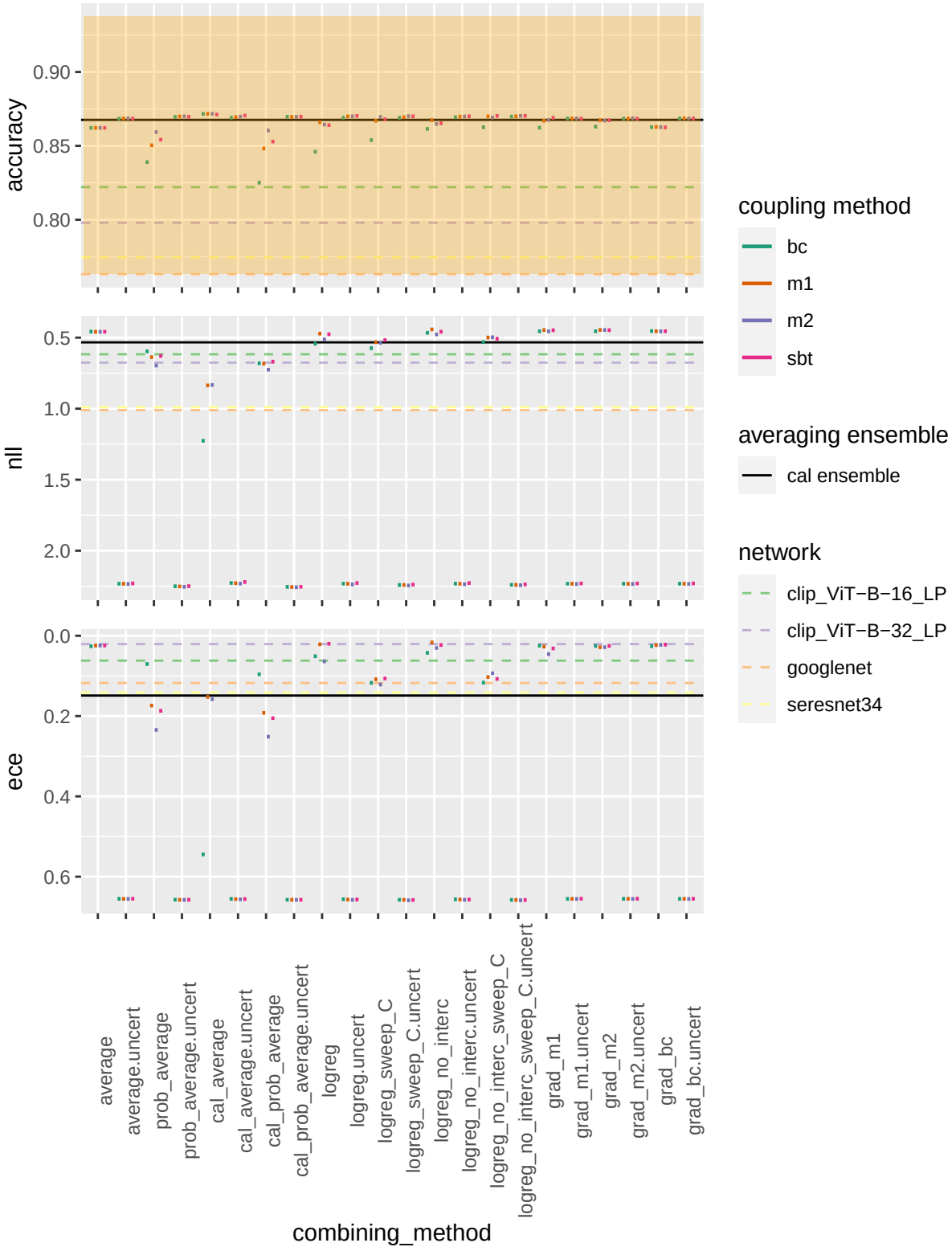
Ensemble metrics
Error inconsistency 0.322099983692169
Average pairwise accuracy variance 4.6937791921664e−05



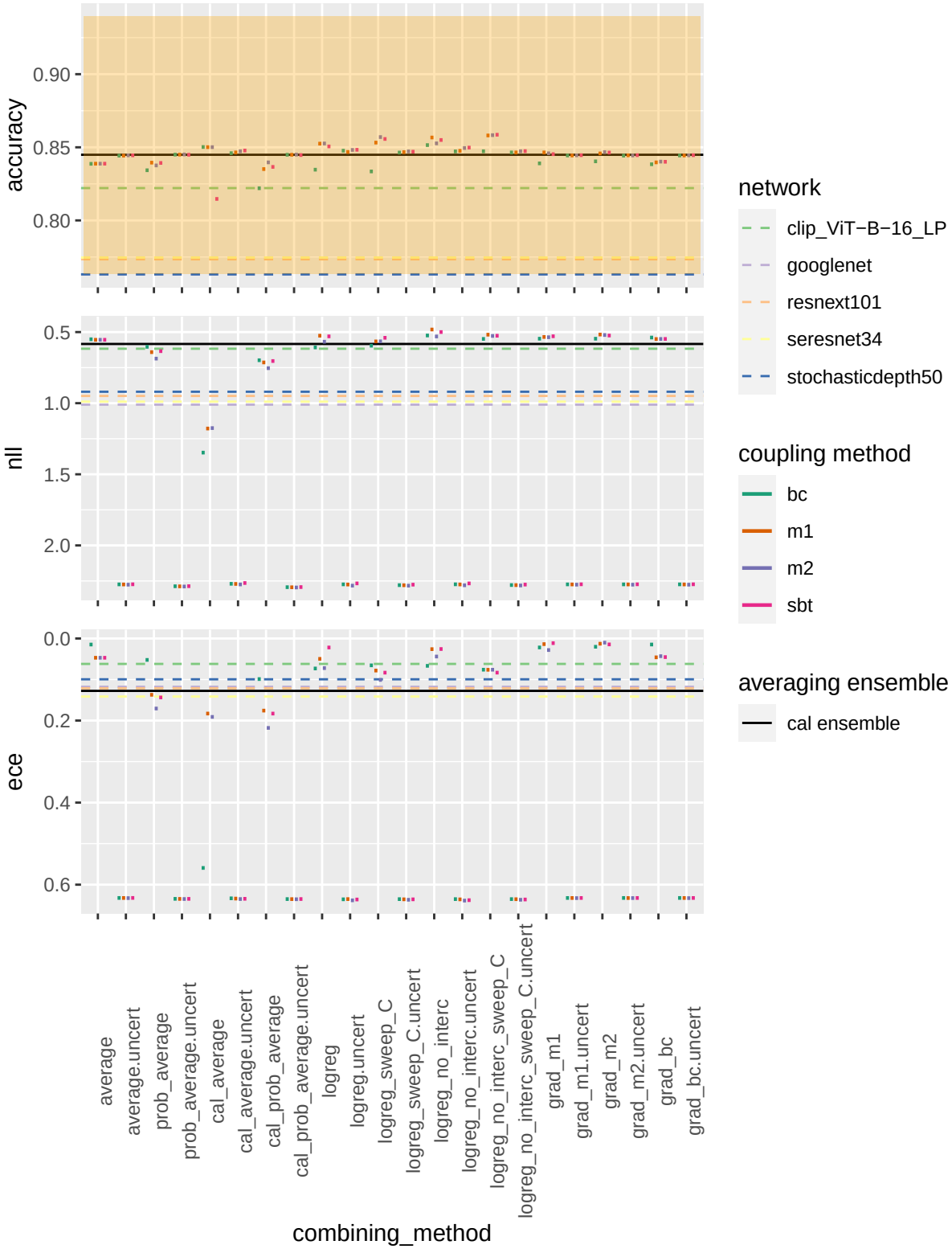
Ensemble metrics
Error inconsistency 0.337799996137619
Average pairwise accuracy variance 4.97215442010202e−05



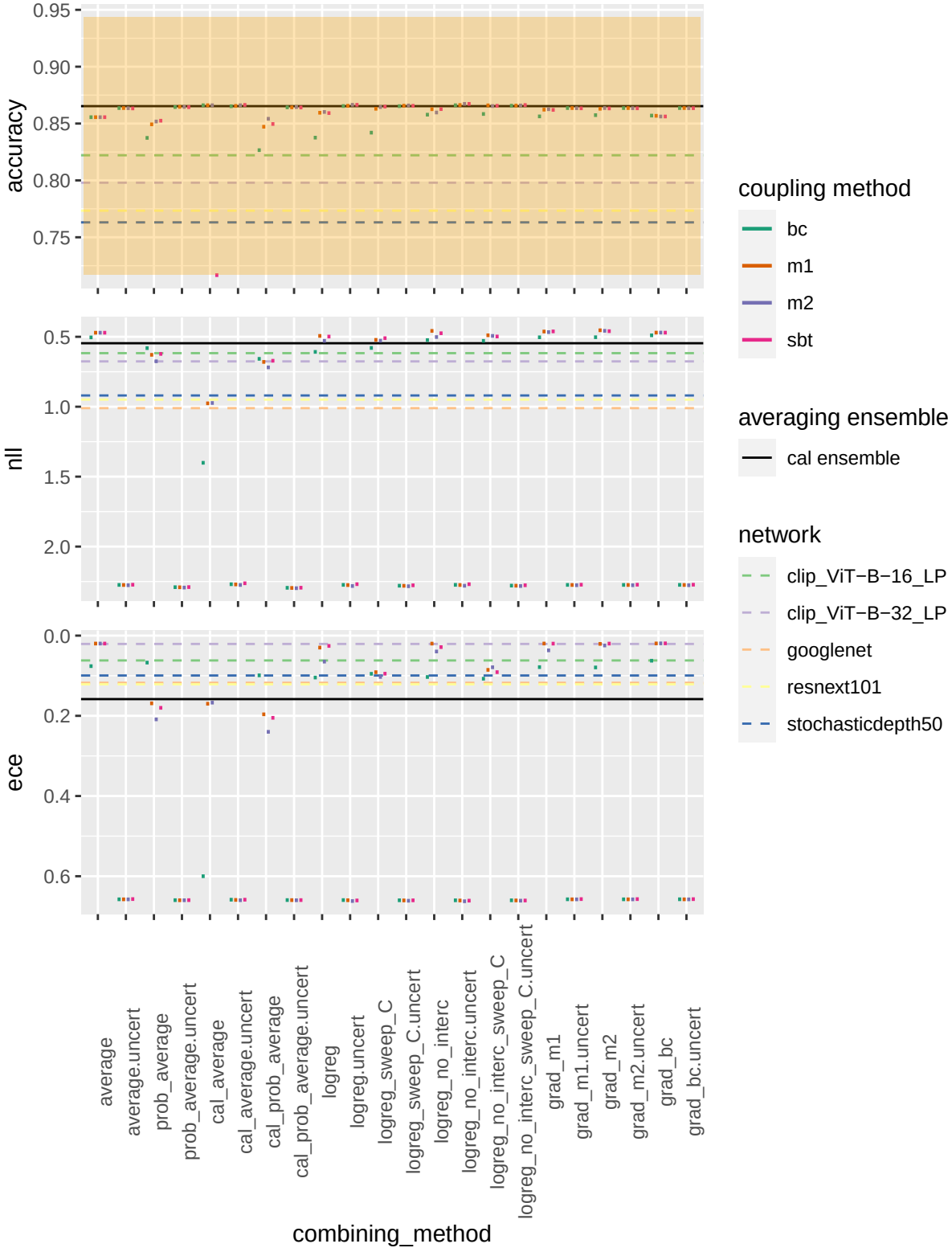
Ensemble metrics
Error inconsistency 0.339899986982346
Average pairwise accuracy variance 5.11846701556351e−05



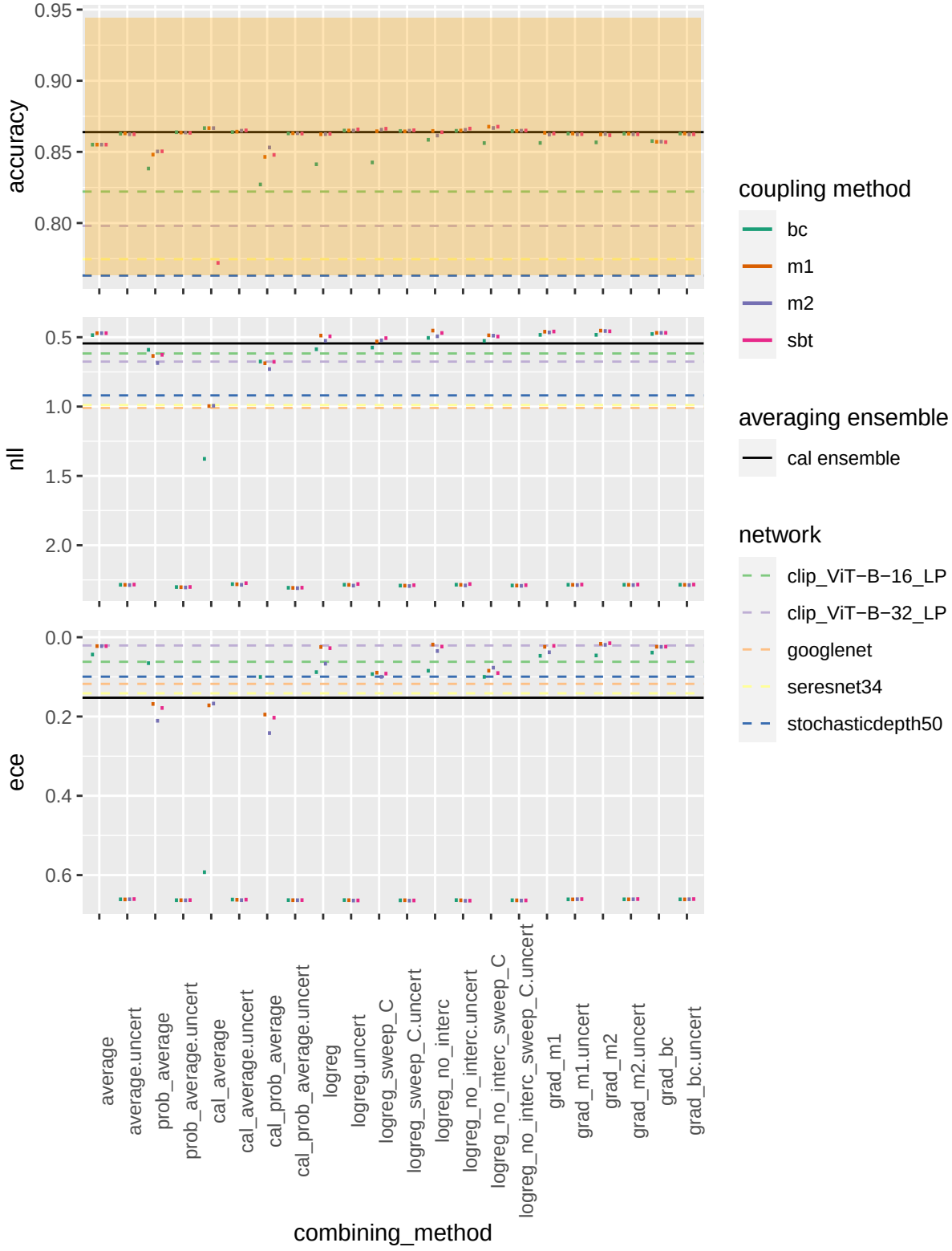
Ensemble metrics
Error inconsistency 0.353499978780746
Average pairwise accuracy variance 4.46607882622629e−05



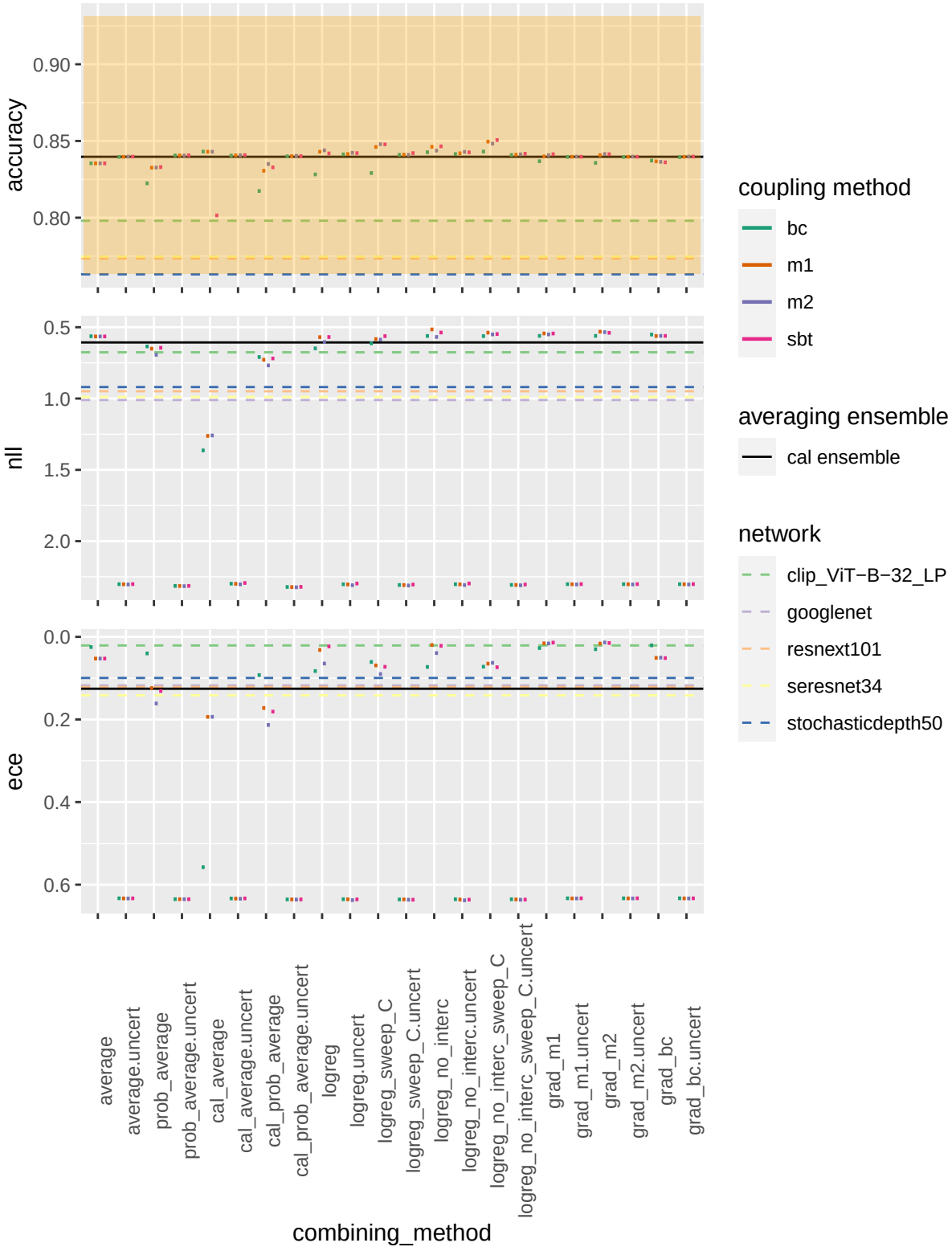
Ensemble metrics
Error inconsistency 0.373499989509583
Average pairwise accuracy variance 4.81671813759021e−05



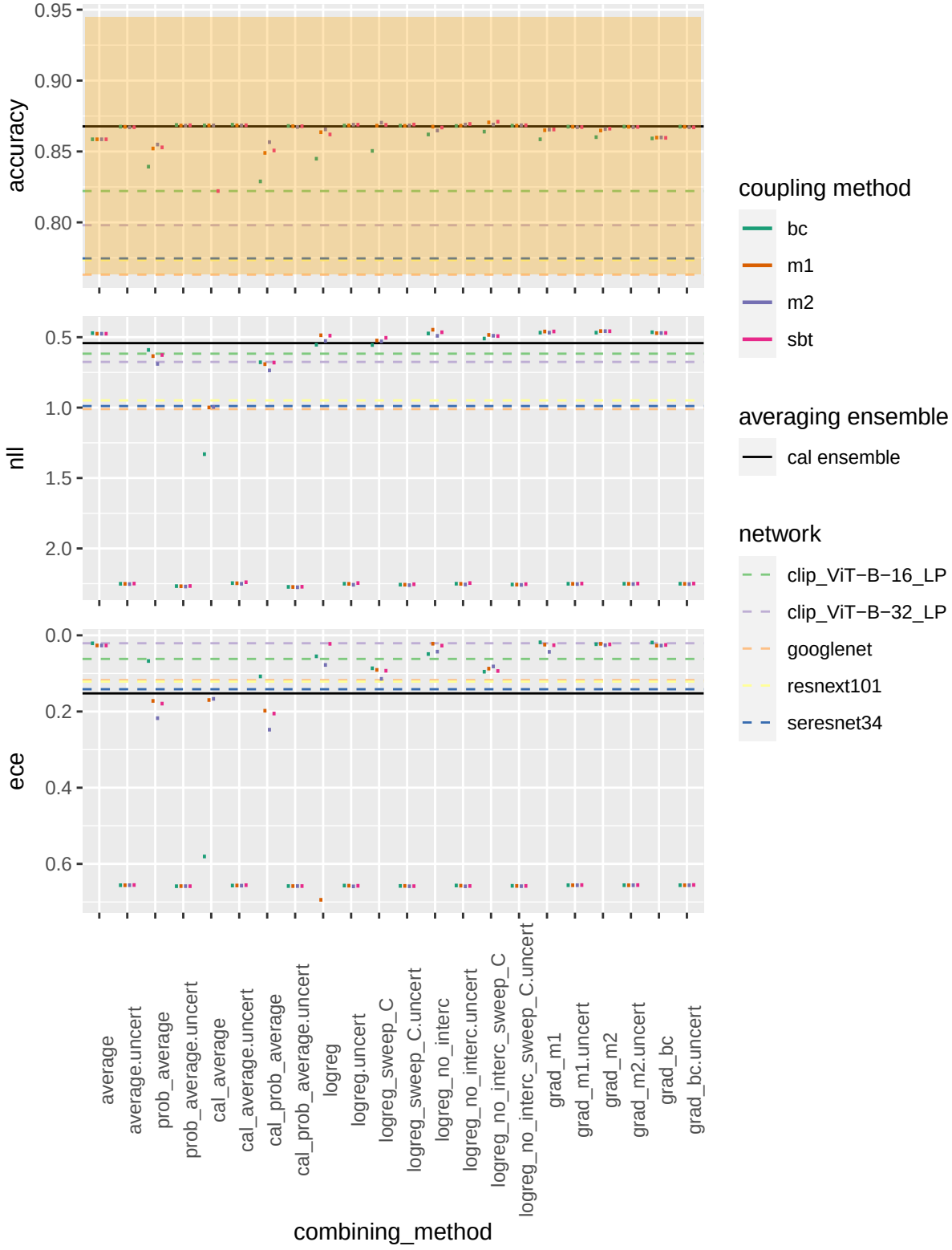
Ensemble metrics
Error inconsistency 0.373400002717972
Average pairwise accuracy variance 4.92575818498153e−05



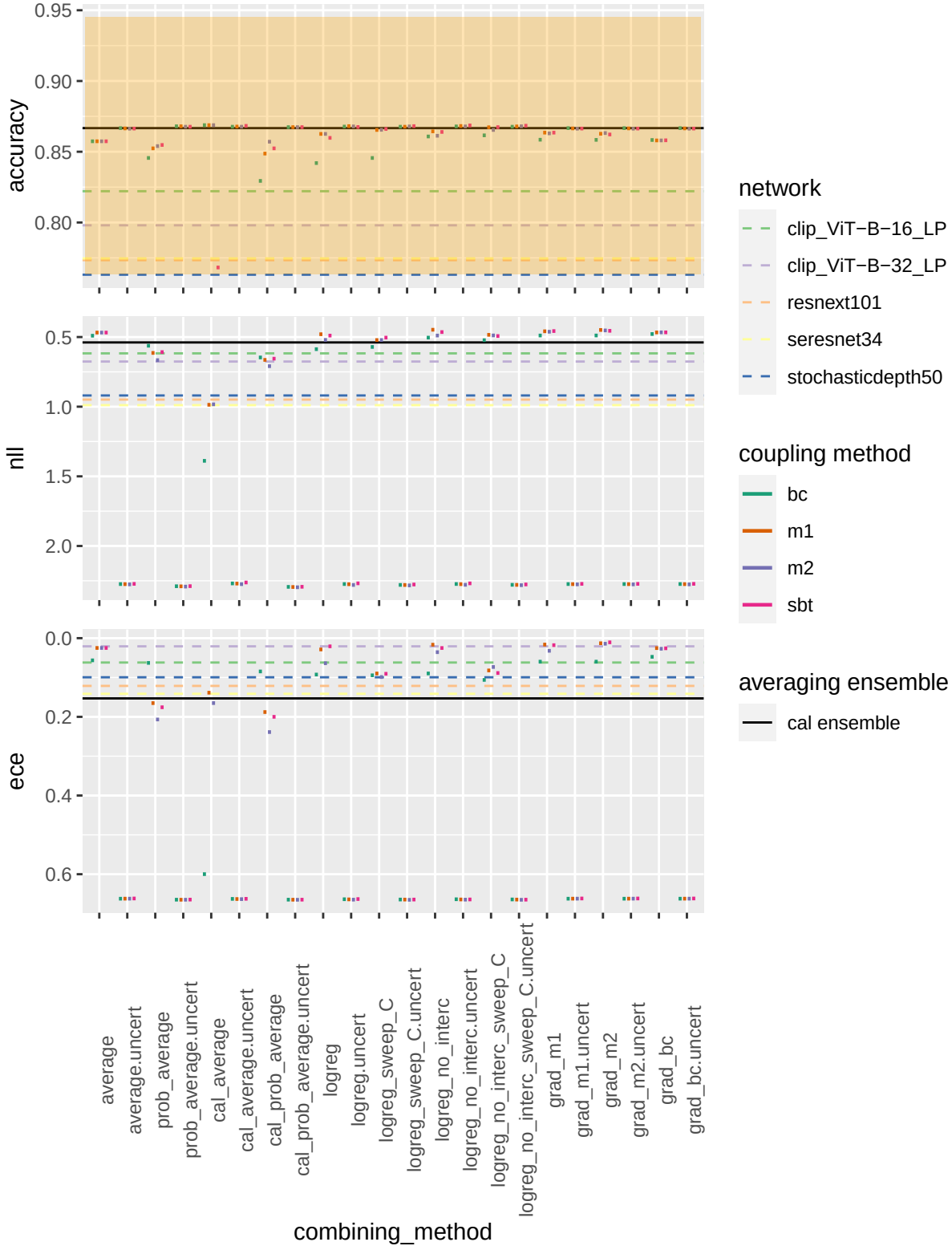
Ensemble metrics
Error inconsistency 0.351300001144409
Average pairwise accuracy variance 4.11695873481221e−05



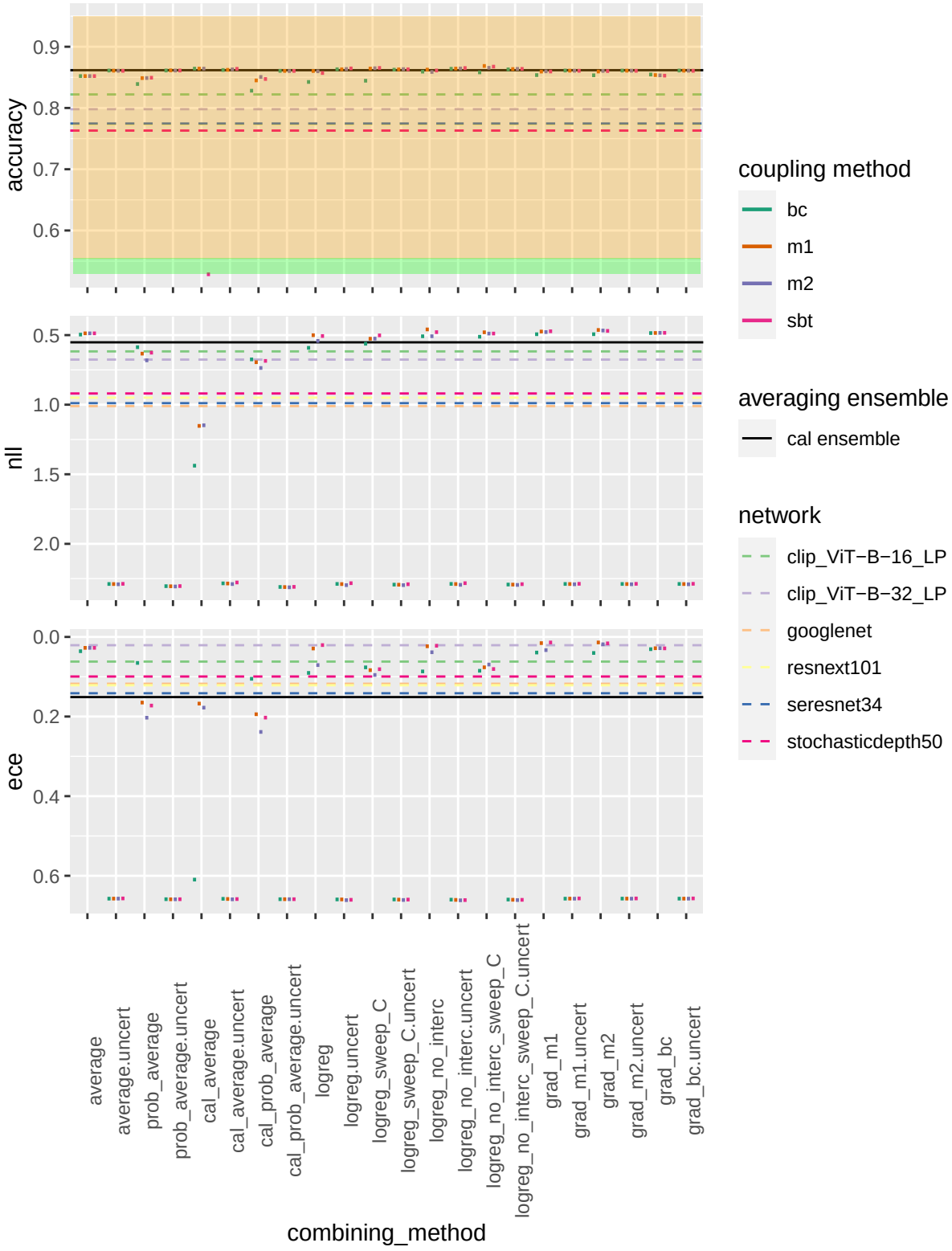
Ensemble metrics
Error inconsistency 0.369699984788895
Average pairwise accuracy variance 4.94615851494018e−05



Ensemble metrics
Error inconsistency 0.368099987506866
Average pairwise accuracy variance 4.2230385588482e−05



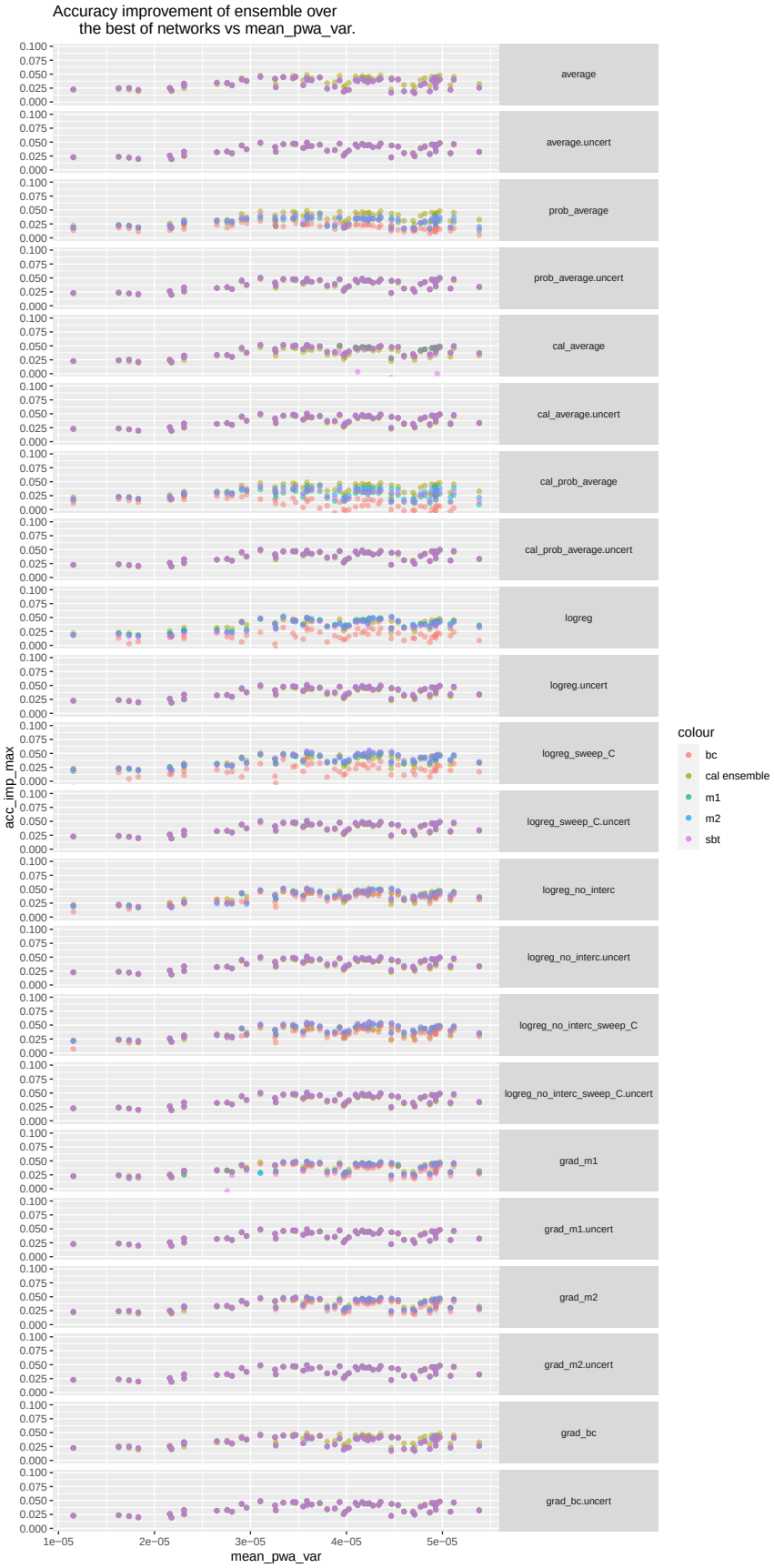
Ensemble metrics
Error inconsistency 0.394899994134903
Average pairwise accuracy variance 4.77338762721047e−05

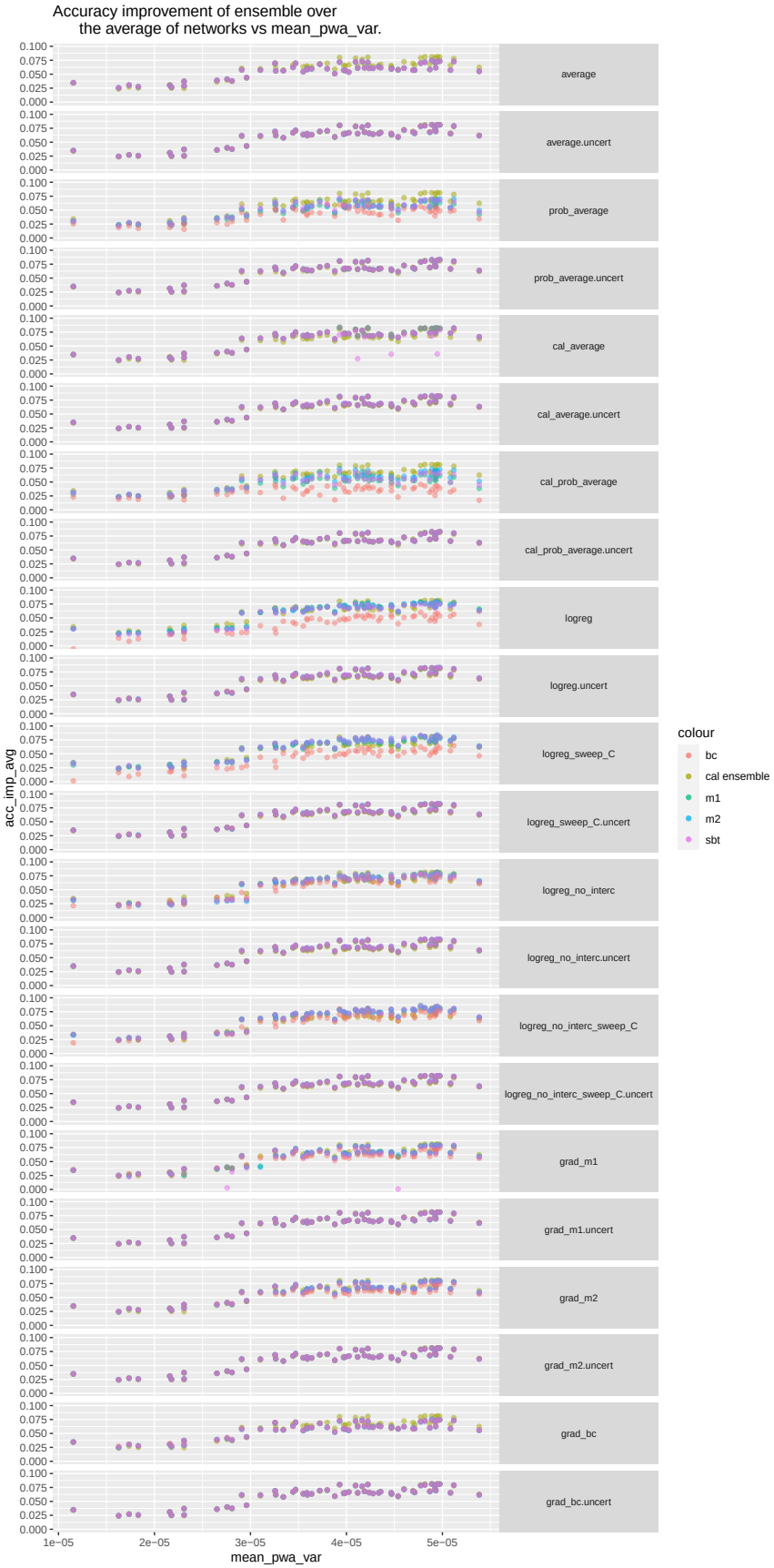


```
xax <- c(
  "err_incons", "mean_pwa_var"
)
yax <- c("acc_imp_max", "acc_imp_avg")

for (xa in xax)
{
  for (ya in yax)
  {
    cur_plot <- ggplot() +
      geom_point(
        data = ens_cal_plt_df,
        mapping = aes_string(x = xa, y = ya, color = shQuote("cal ensemble")),
        alpha = 0.5
      ) +
      geom_point(
        data = 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.",
        if (ya == "acc_imp_max") "best" else "average", xa
      )) +
      coord_cartesian(ylim=c(0, 0.1)) +
      theme(strip.text.y = element_text(size = 8, angle = 0))

    print(cur_plot)
  }
}
```

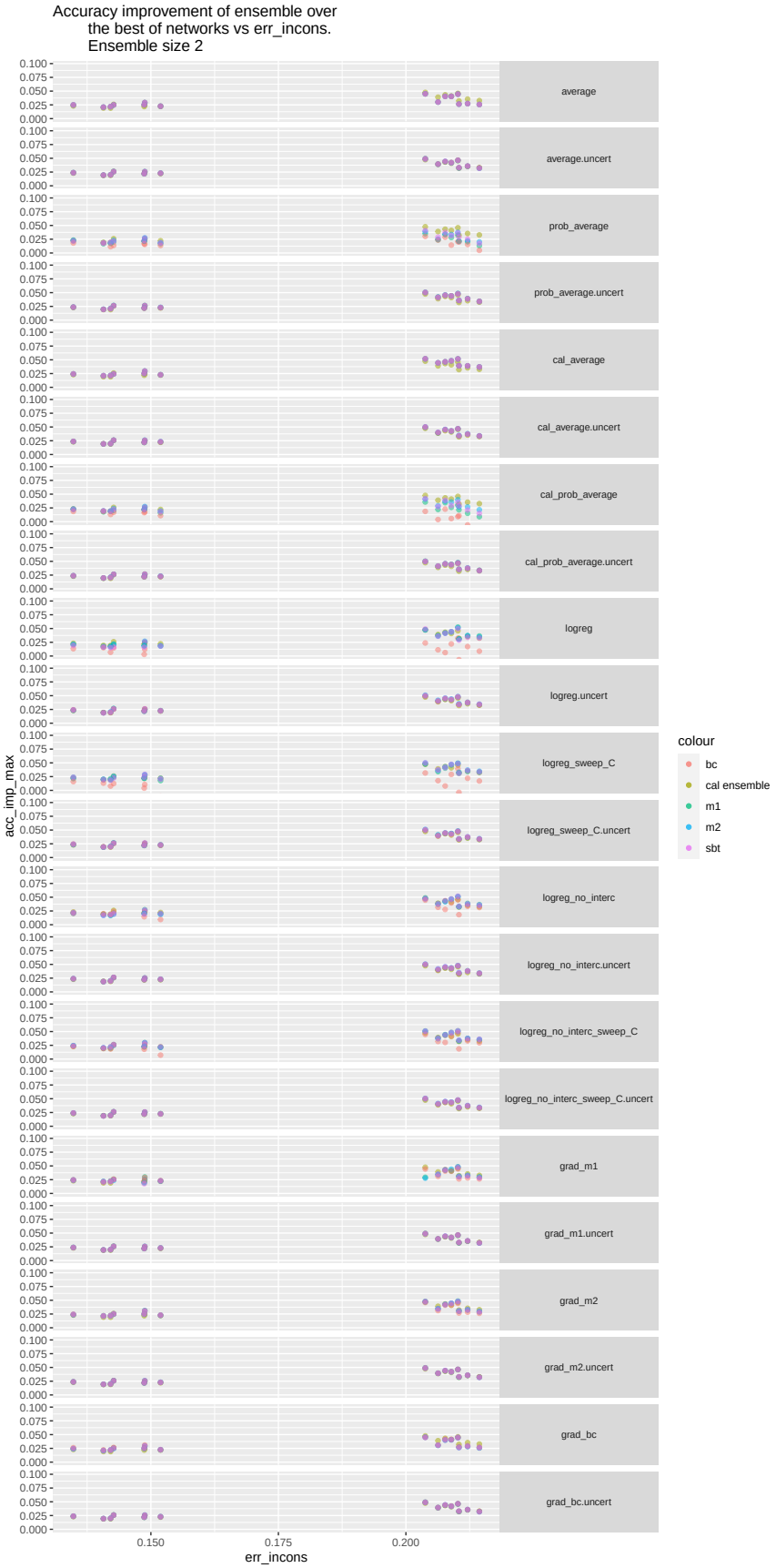

Accuracy improvement of ensemble over the best of networks vs err_incons.
acc_imp_max
err_incons
0.2
0.3
0.4
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
colour
bc
cal ensemble
m1
m2
sbt

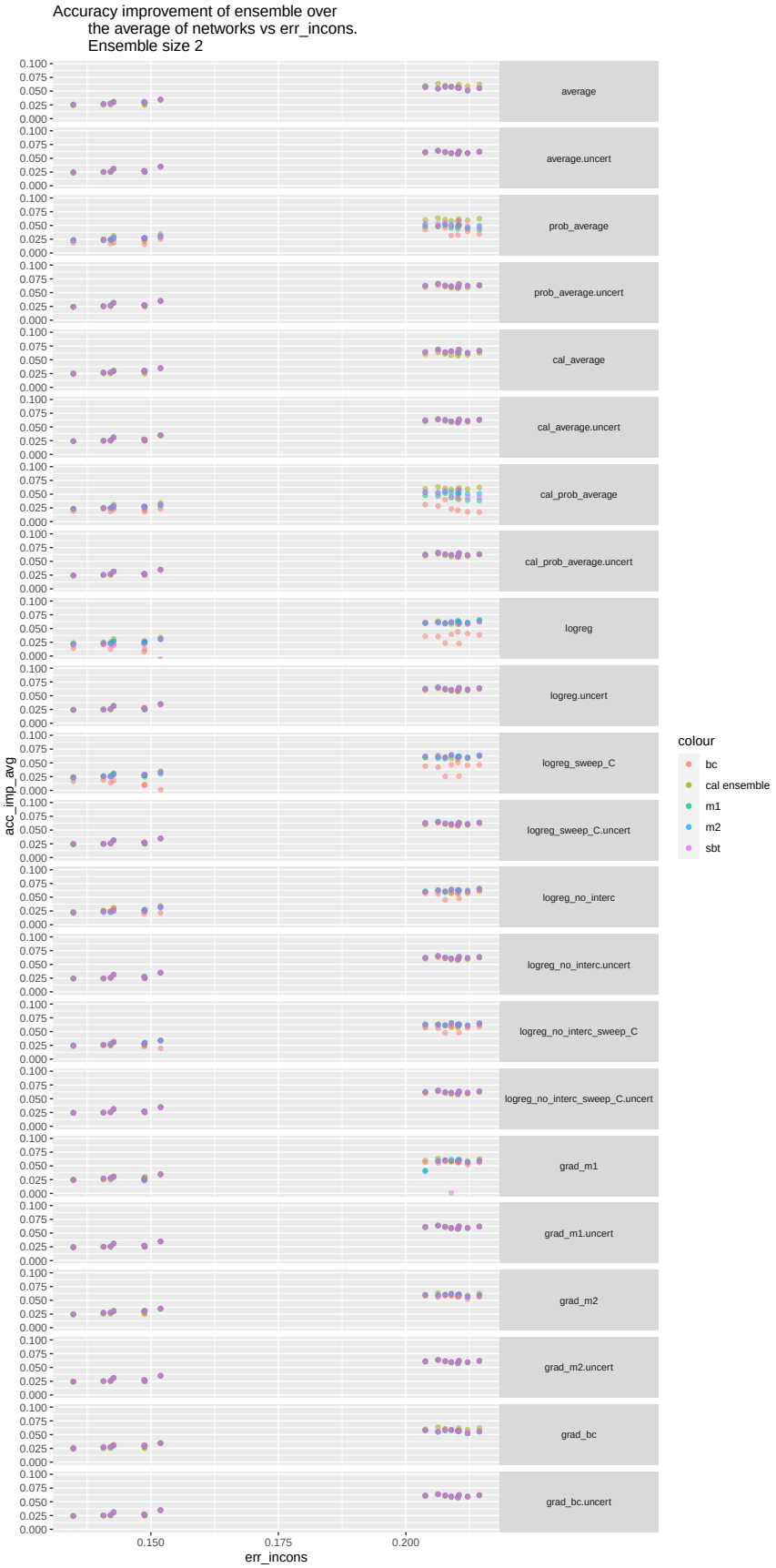

Accuracy improvement of ensemble over
the average of networks vs err_incons.
acc_imp_avg
err_incons
0.2
0.3
0.4
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
colour
bc
cal ensemble
m1
m2
sbt

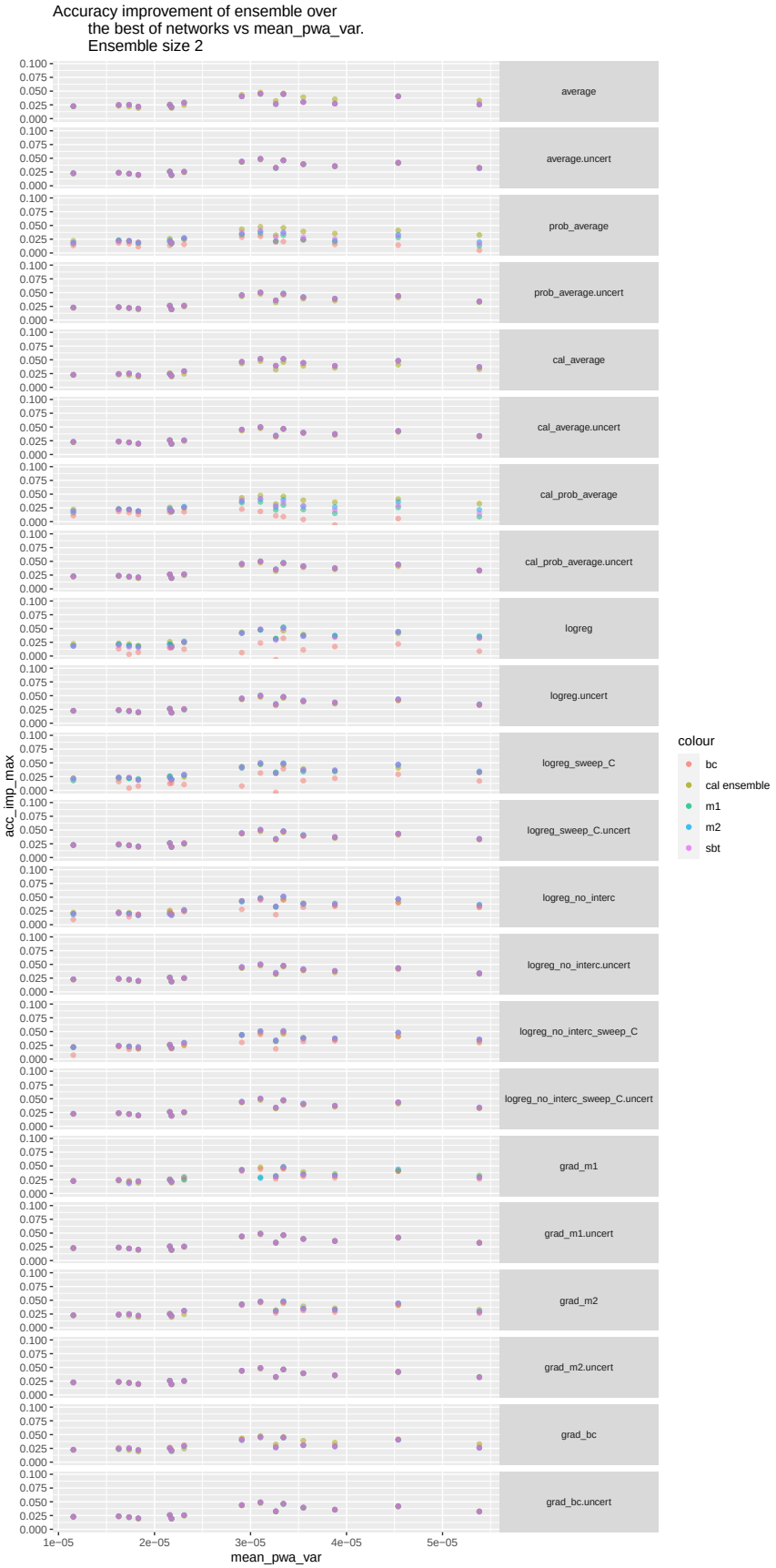

Accuracy improvement of ensemble over
the best of networks vs mean_pwa_var.
acc_imp_max
mean_pwa_var
0.100
0.075
0.050
0.025
0.000
1e−05
2e−05
3e−05
4e−05
5e−05
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
colour
bc
cal ensemble
m1
m2
sbt


Accuracy improvement of ensemble over
the average of networks vs mean_pwa_var.
acc_imp_avg
mean_pwa_var
0.100
0.075
0.050
0.025
0.000
1e−05
2e−05
3e−05
4e−05
5e−05
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
colour
bc
cal ensemble
m1
m2
sbt

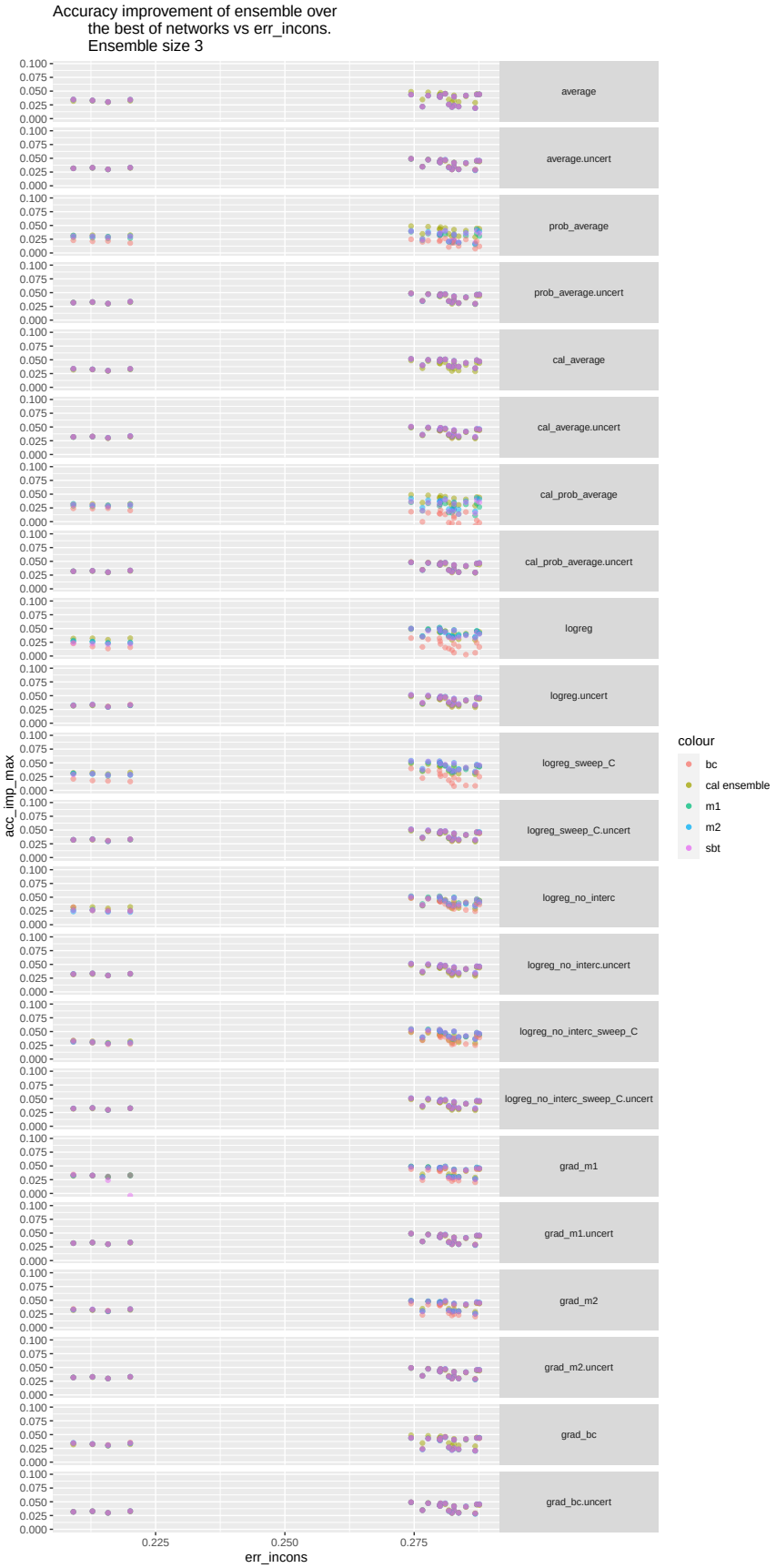
```
for (sss in unique(ens_cal_plt_df$combination_size))
{
  cur_ens_cal_plt_df <- ens_cal_plt_df %>% filter(combination_size == sss)
  cur_ens_pwc_plt_df <- ens_pwc_plt_df %>% filter(combination_size == sss)
  for (xa in xax)
  {
    for (ya in yax)
    {
      cur_plot <- ggplot() +
        geom_point(
          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
        )) +
        coord_cartesian(ylim=c(0, 0.1)) +
        theme(strip.text.y = element_text(size = 8, angle = 0))

      print(cur_plot)
    }
  }
}
```

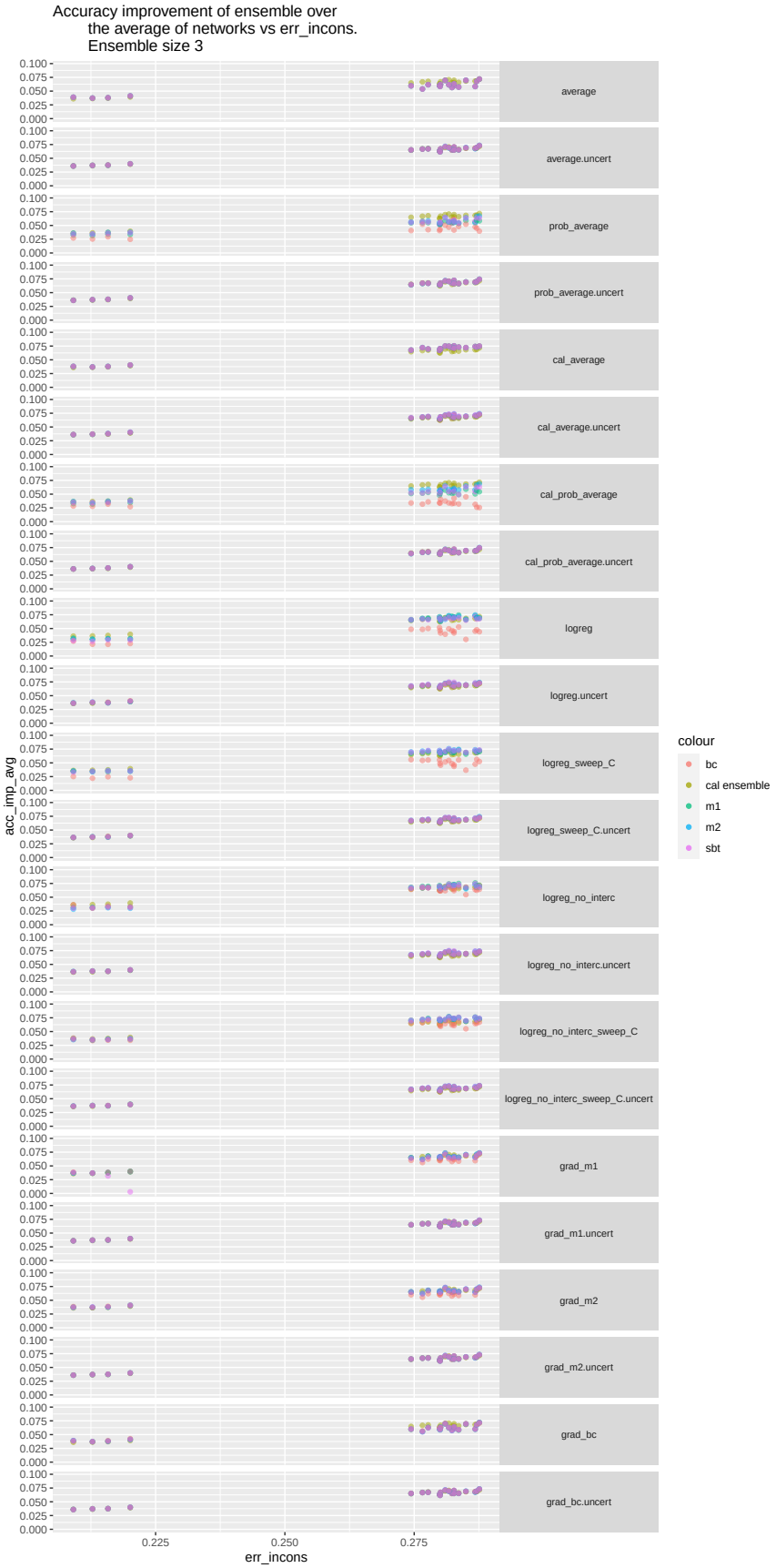

Accuracy improvement of ensemble over the best of networks vs err_incons.
Ensemble size 2
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
acc_imp_max
0.000
0.025
0.050
0.075
0.100
0.150
0.175
0.200
err_incons
colour
bc
cal ensemble
m1
m2
sbt

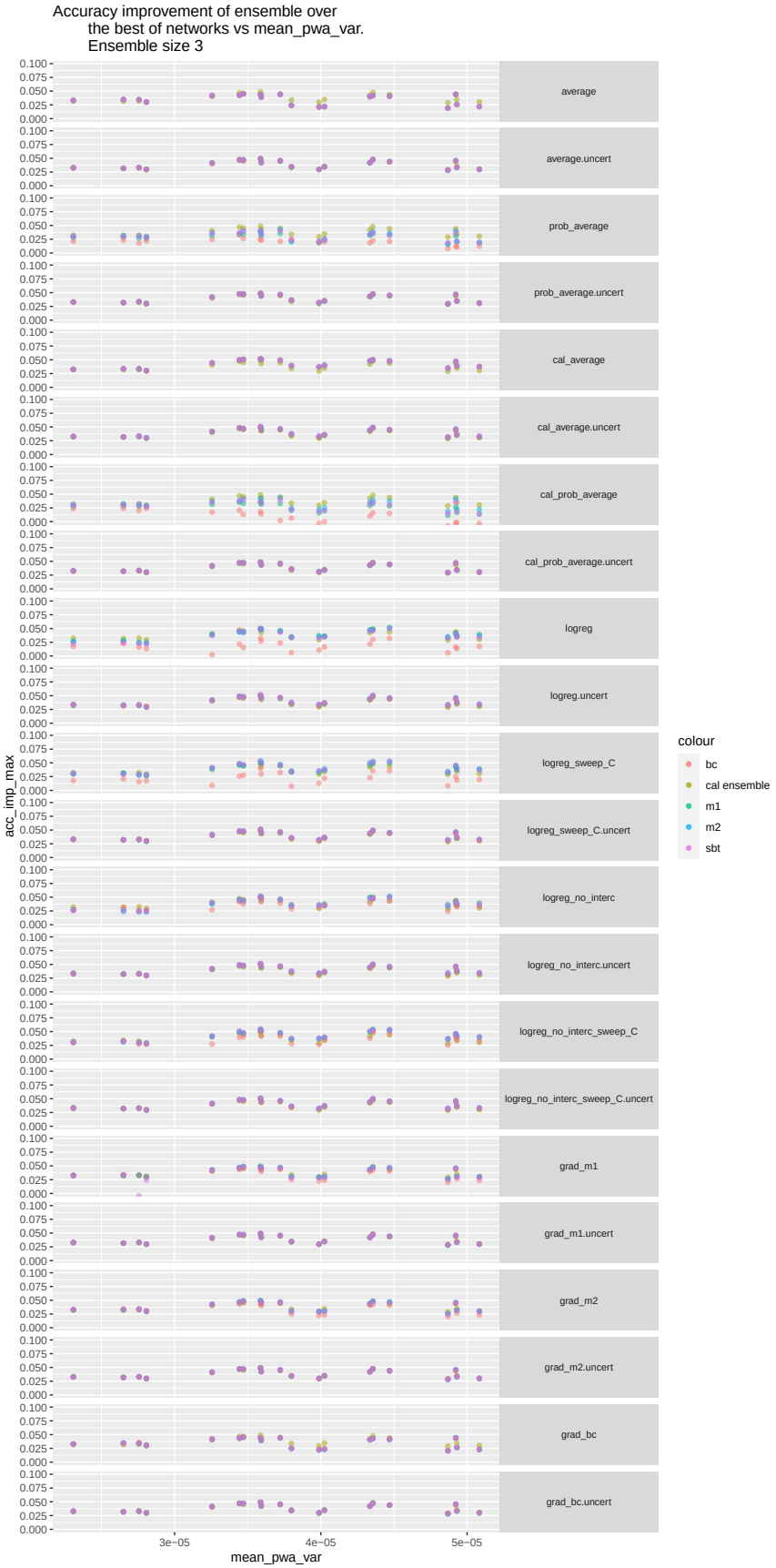

Accuracy improvement of ensemble over
the average of networks vs err_incons.
Ensemble size 2
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
acc_imp_avg
0.000
0.025
0.050
0.075
0.100
0.150
0.175
0.200
err_incons
colour
bc
cal ensemble
m1
m2
sbt

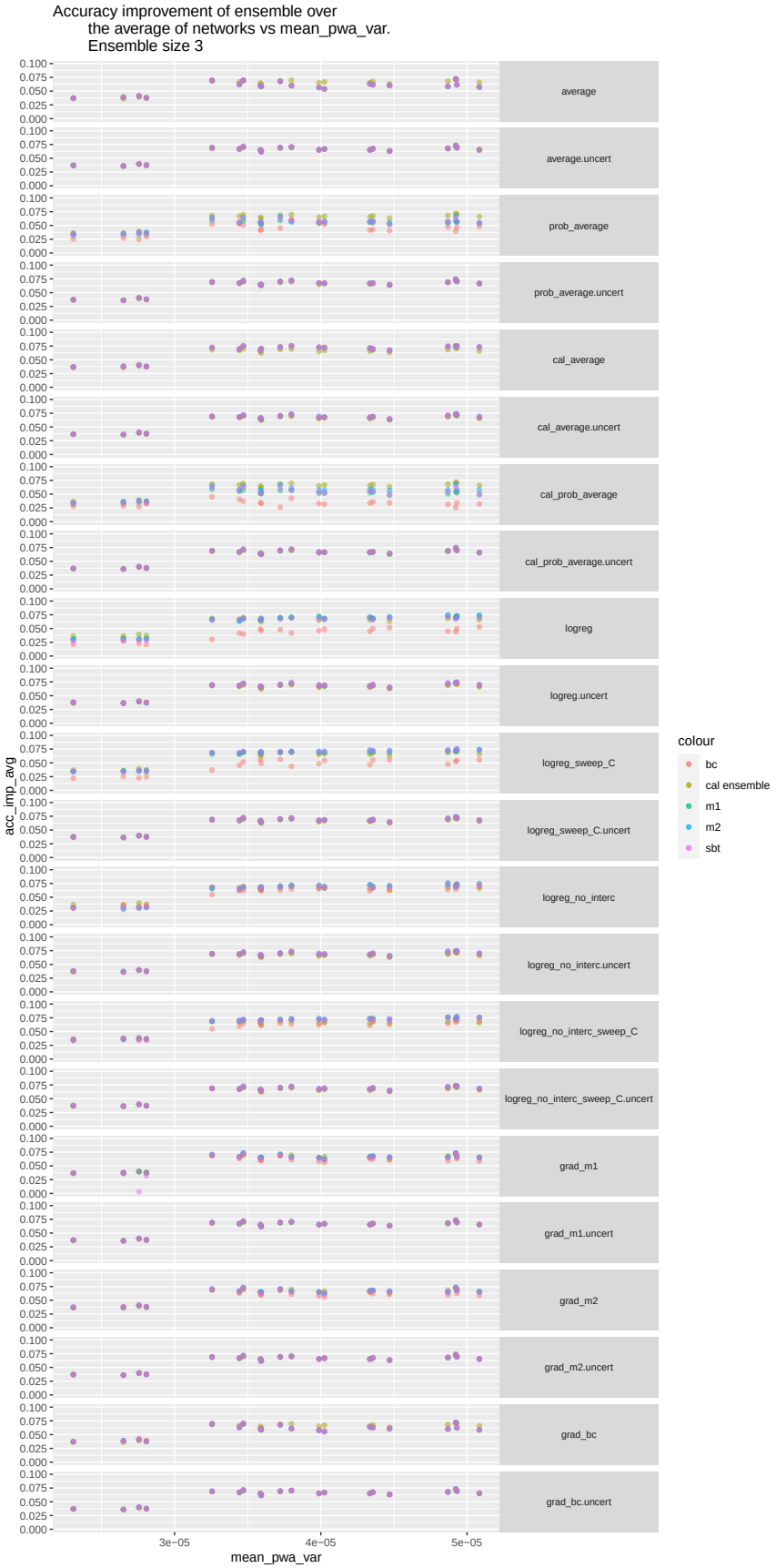

Accuracy improvement of ensemble over
the best of networks vs mean_pwa_var.
Ensemble size 2
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
0.100
0.075
0.050
0.025
0.000
acc_imp_max
1e−05
2e−05
3e−05
4e−05
5e−05
mean_pwa_var
colour
bc
cal ensemble
m1
m2
sbt

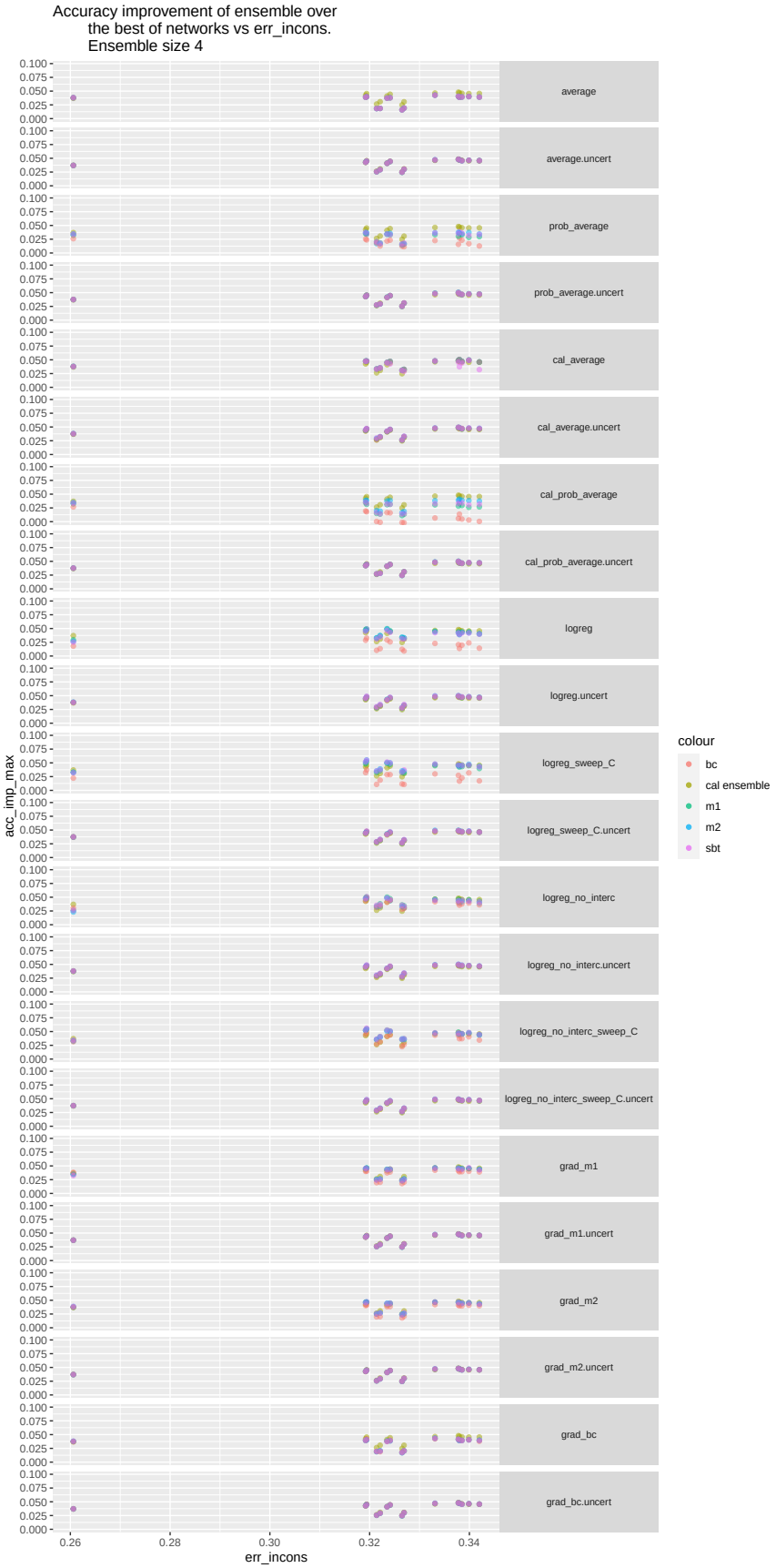

Accuracy improvement of ensemble over
the average of networks vs mean_pwa_var.
Ensemble size 2
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
0.100
0.075
0.050
0.025
0.000
acc_imp_avg
1e−05
2e−05
3e−05
4e−05
5e−05
mean_pwa_var
colour
bc
cal ensemble
m1
m2
sbt

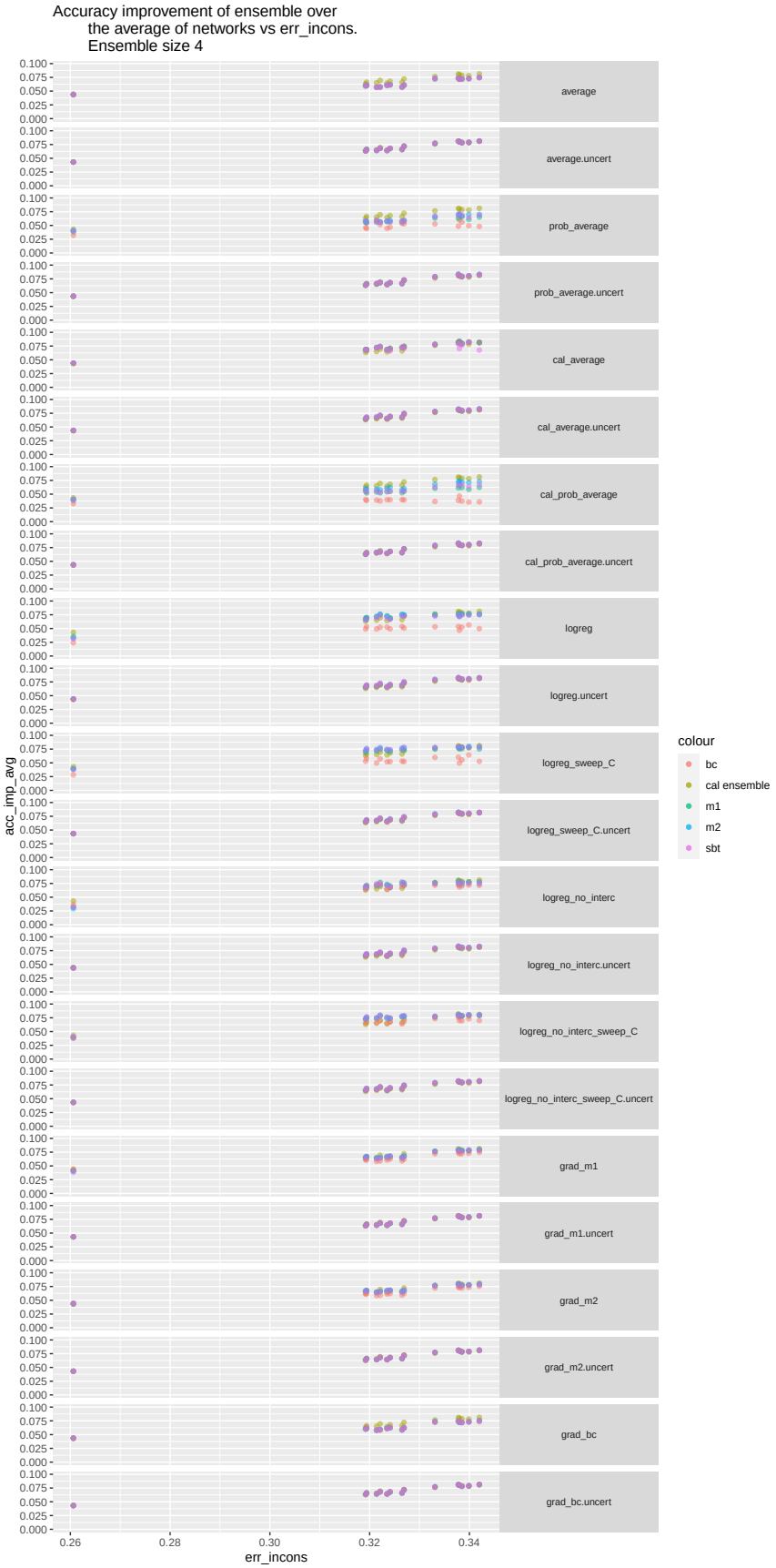

Accuracy improvement of ensemble over the best of networks vs err_incons.
Ensemble size 3
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
acc_imp_max
0.100
0.075
0.050
0.025
0.000
0.225
0.250
0.275
err_incons
colour
bc
cal ensemble
m1
m2
sbt


Accuracy improvement of ensemble over
the average of networks vs err_incons.
Ensemble size 3
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
acc_imp_avg
0.100
0.075
0.050
0.025
0.000
0.225
0.250
0.275
err_incons
colour
bc
cal ensemble
m1
m2
sbt

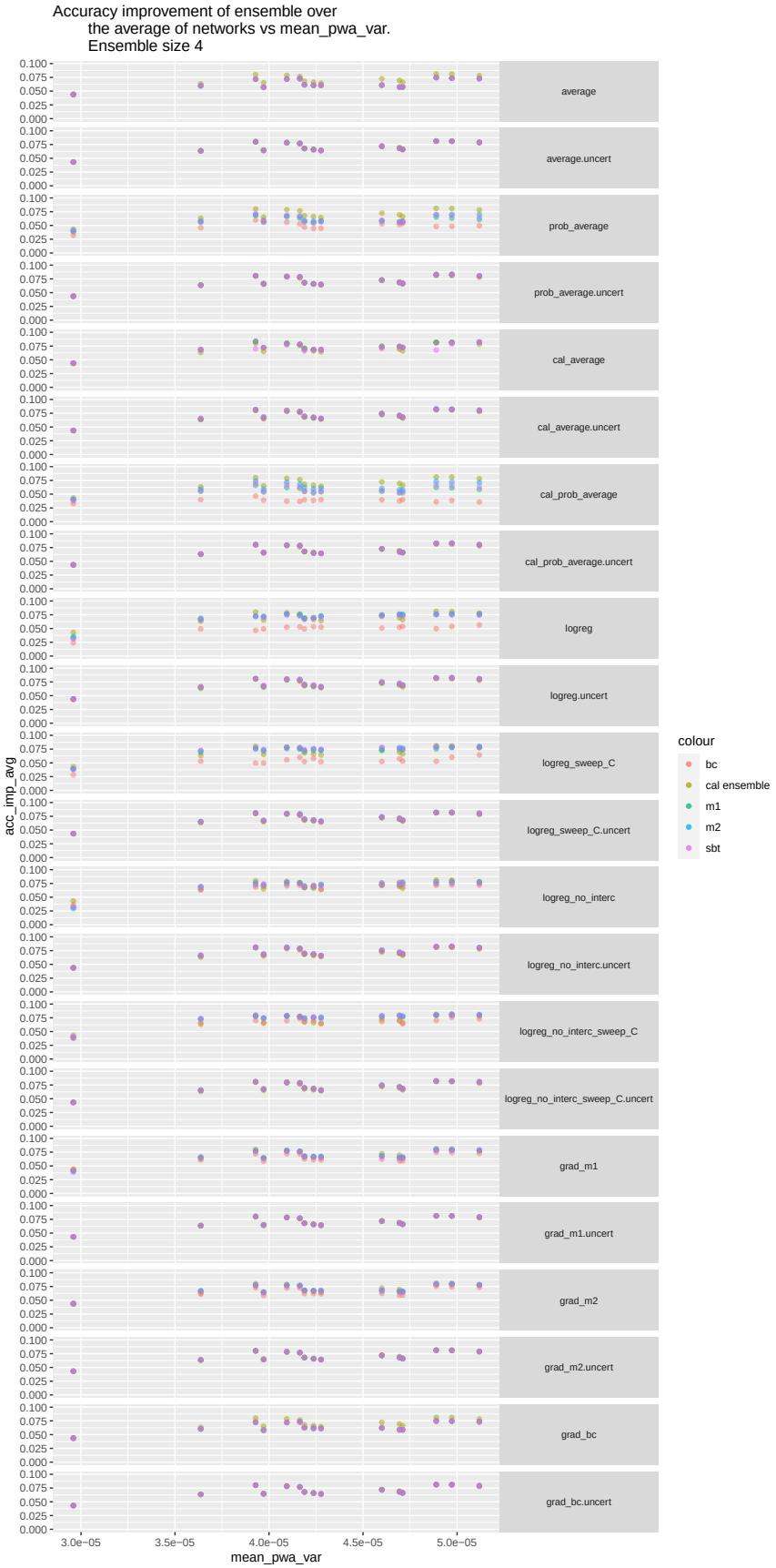

Accuracy improvement of ensemble over
the best of networks vs mean_pwa_var.
Ensemble size 3
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
0.100
0.075
0.050
0.025
0.000
acc_imp_max
3e−05
4e−05
5e−05
mean_pwa_var
colour
bc
cal ensemble
m1
m2
sbt


Accuracy improvement of ensemble over
the average of networks vs mean_pwa_var.
Ensemble size 3
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
0.100
0.075
0.050
0.025
0.000
acc_imp_avg
3e−05
4e−05
5e−05
mean_pwa_var
colour
bc
cal ensemble
m1
m2
sbt


Accuracy improvement of ensemble over the best of networks vs err_incons.
Ensemble size 4
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
acc_imp_max
err_incons
0.000
0.025
0.050
0.075
0.100
0.26
0.28
0.30
0.32
0.34
colour
bc
cal ensemble
m1
m2
sbt


Accuracy improvement of ensemble over
the average of networks vs err_incons.
Ensemble size 4
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
acc_imp_avg
0.100
0.075
0.050
0.025
0.000
0.26
0.28
0.30
0.32
0.34
err_incons
colour
bc
cal ensemble
m1
m2
sbt

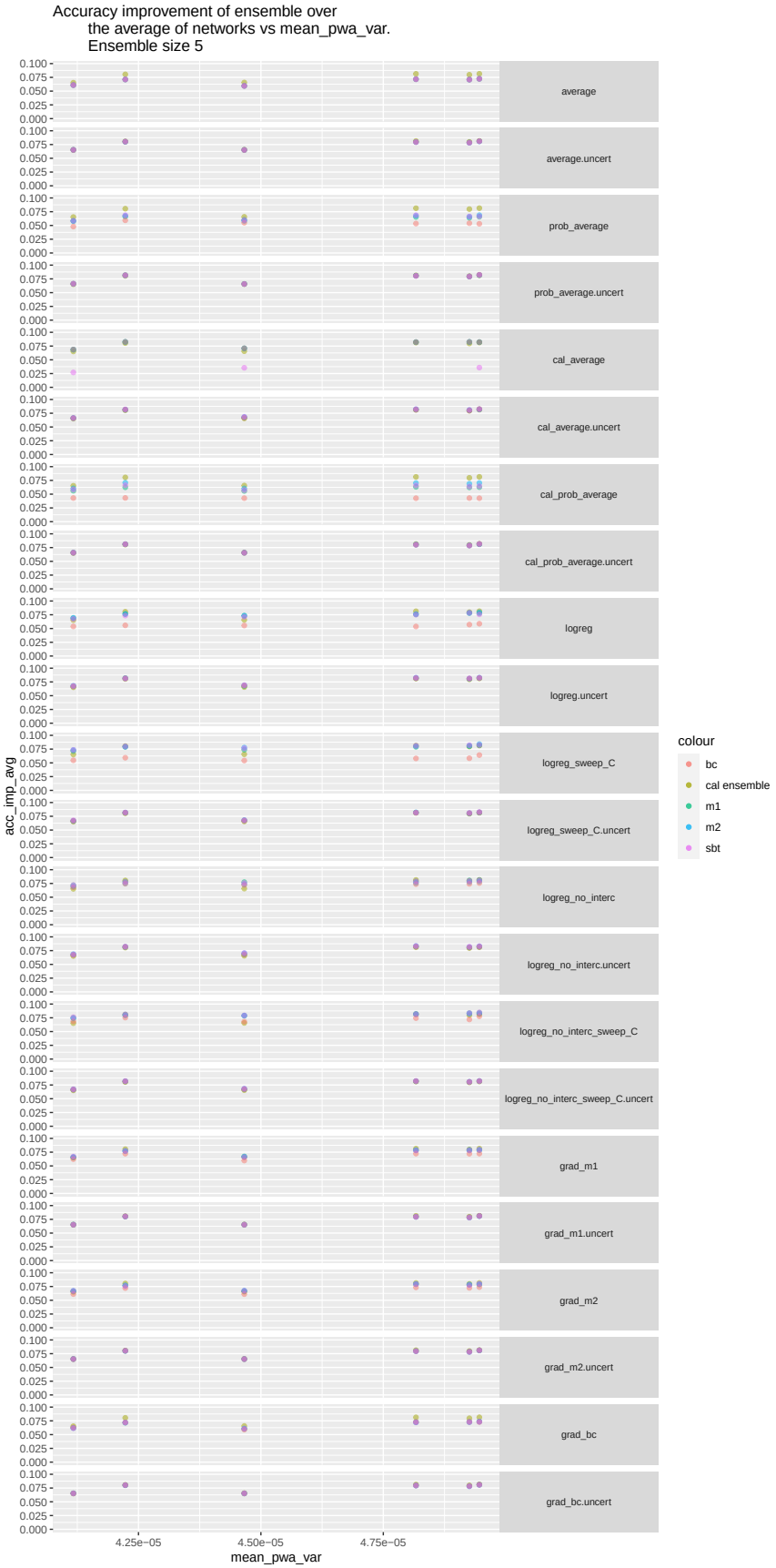

Accuracy improvement of ensemble over
the best of networks vs mean_pwa_var.
Ensemble size 4
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
acc_imp_max
0.100
0.075
0.050
0.025
0.000
3.0e−05
3.5e−05
4.0e−05
4.5e−05
5.0e−05
mean_pwa_var
colour
bc
cal ensemble
m1
m2
sbt

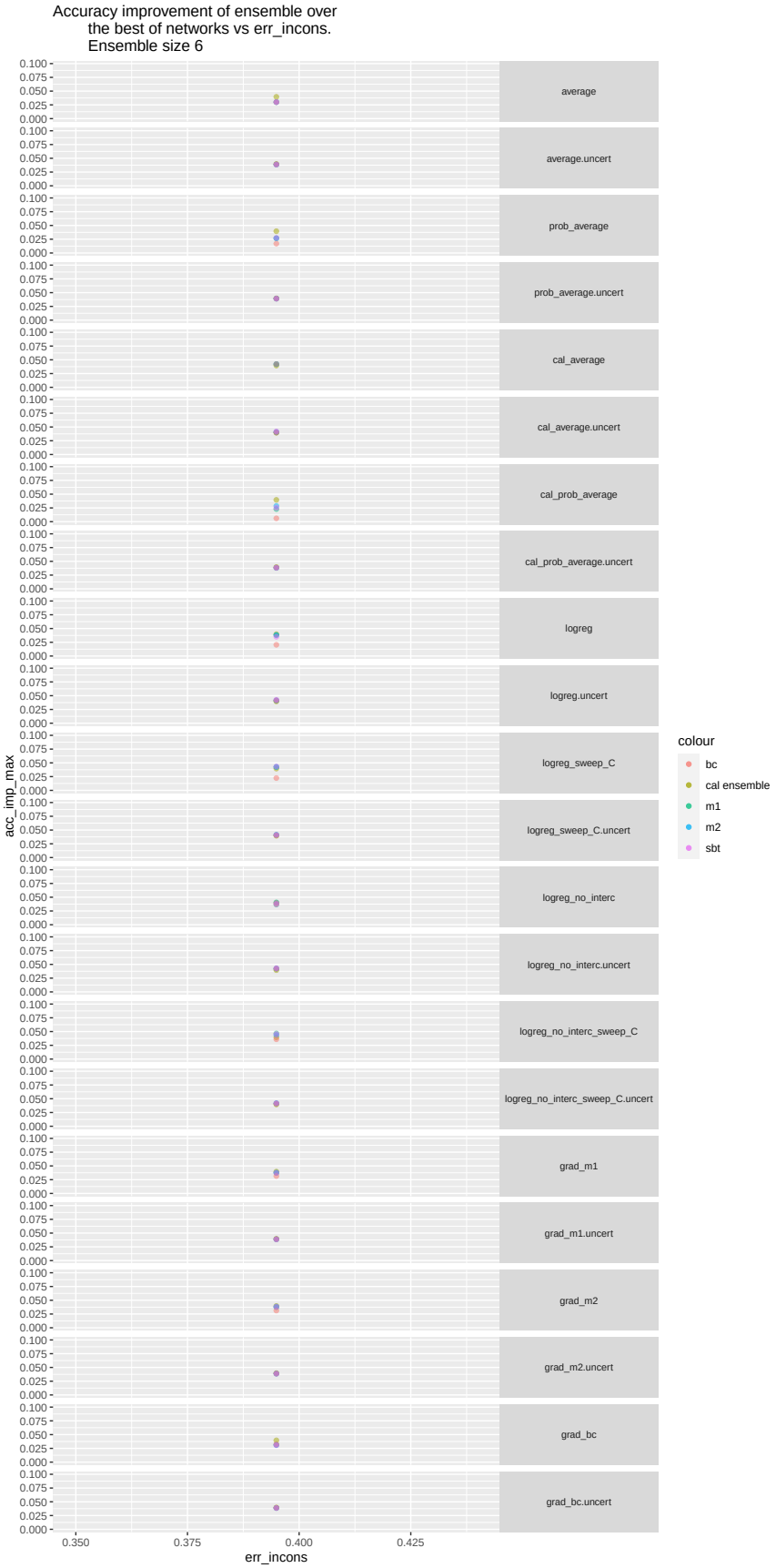

Accuracy improvement of ensemble over
the average of networks vs mean_pwa_var.
Ensemble size 4
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
0.100
0.075
0.050
0.025
0.000
acc_imp_avg
3.0e−05
3.5e−05
4.0e−05
4.5e−05
5.0e−05
mean_pwa_var
colour
bc
cal ensemble
m1
m2
sbt

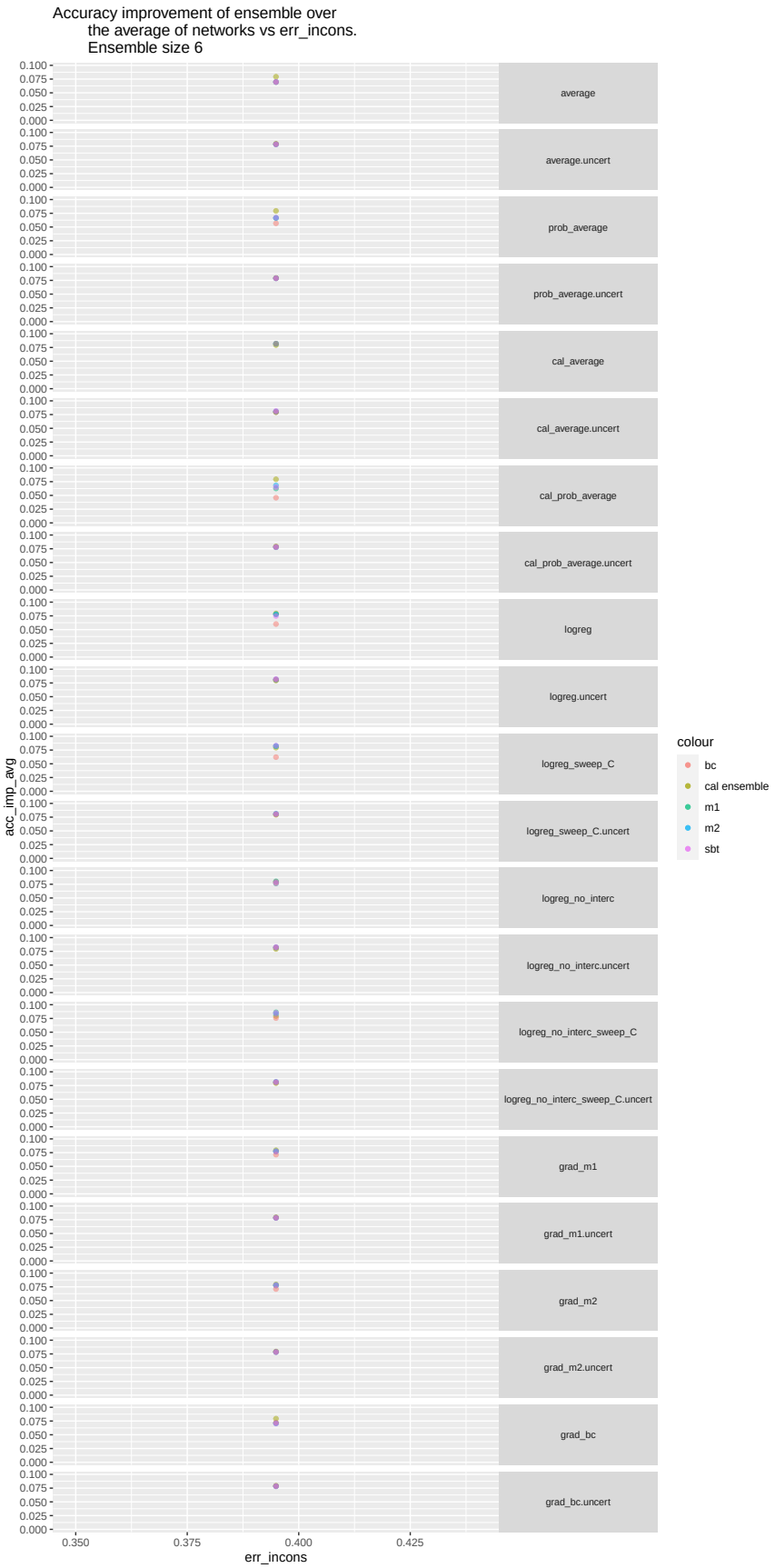

Accuracy improvement of ensemble over the best of networks vs err_incons. Ensemble size 5
acc_imp_max
err_incons
0.355
0.360
0.365
0.370
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
colour
bc
cal ensemble
m1
m2
sbt

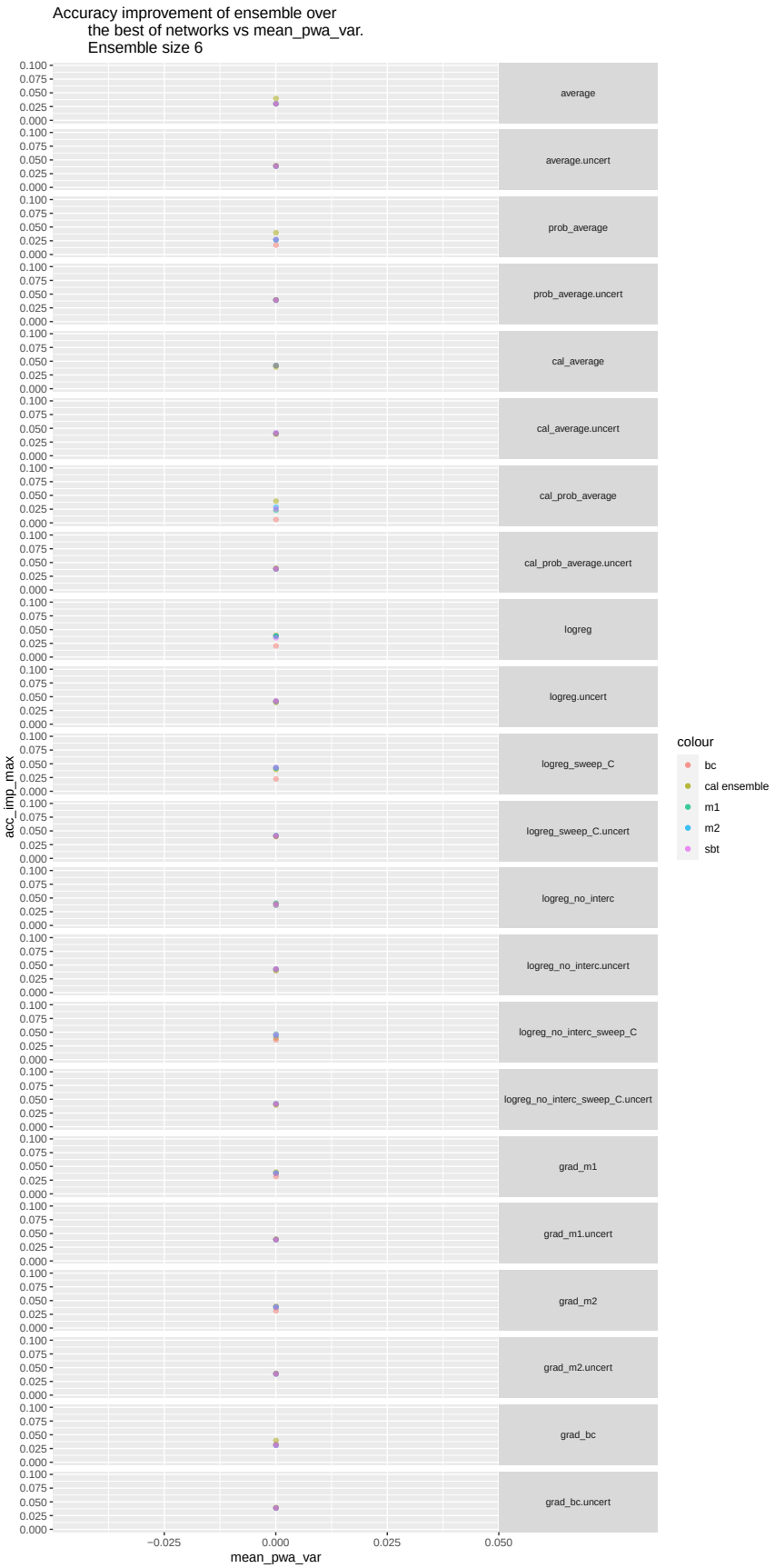

Accuracy improvement of ensemble over
the average of networks vs err_incons.
Ensemble size 5
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
acc_imp_avg
0.100
0.075
0.050
0.025
0.000
0.355
0.360
0.365
0.370
err_incons
colour
bc
cal ensemble
m1
m2
sbt

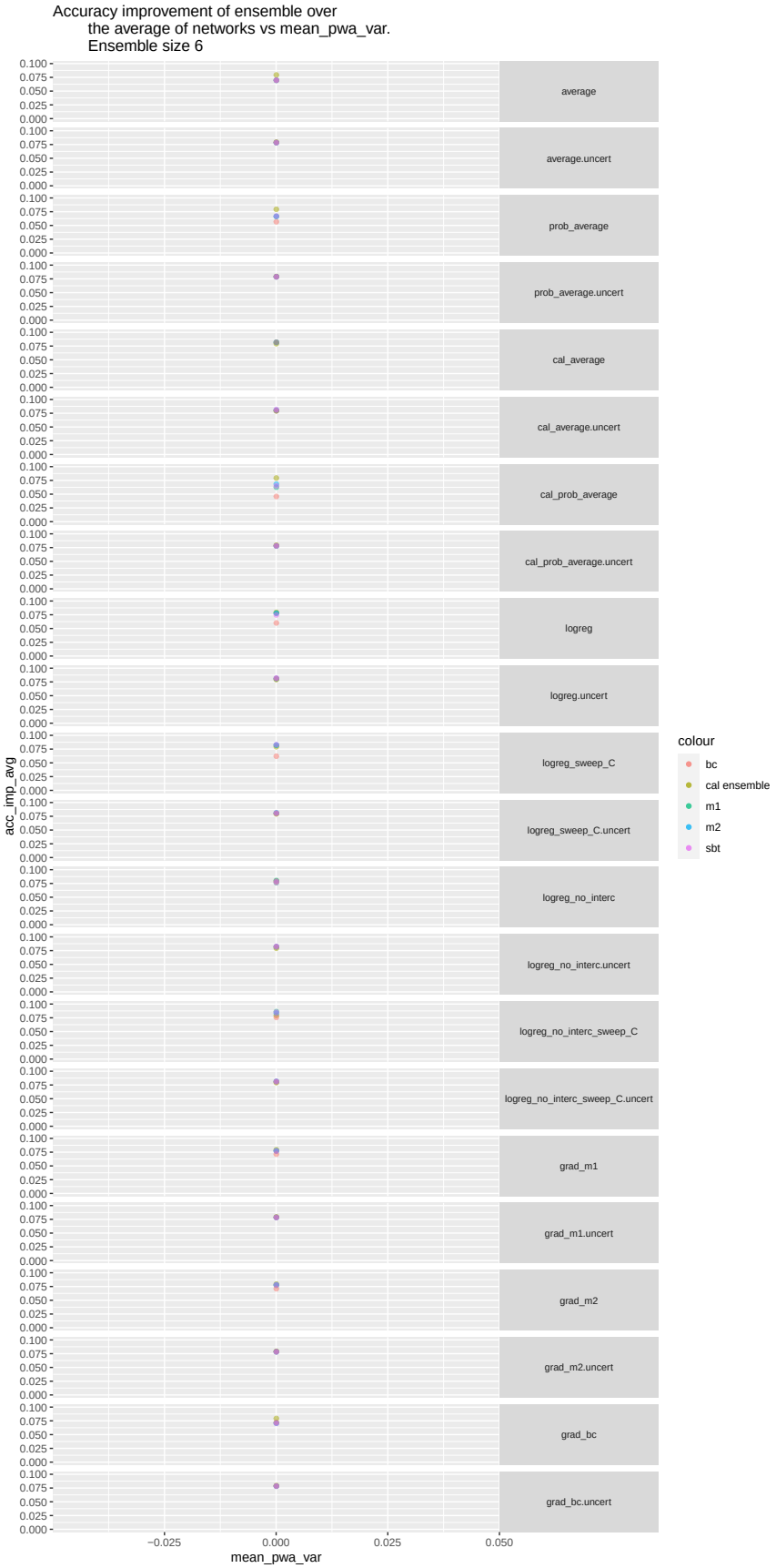

Accuracy improvement of ensemble over
the best of networks vs mean_pwa_var.
Ensemble size 5
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
0.100
0.075
0.050
0.025
0.000
acc_imp_max
4.25e−05
4.50e−05
4.75e−05
mean_pwa_var
colour
bc
cal ensemble
m1
m2
sbt


Accuracy improvement of ensemble over
the average of networks vs mean_pwa_var.
Ensemble size 5
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
acc_imp_avg
0.100
0.075
0.050
0.025
0.000
4.25e−05
4.50e−05
4.75e−05
mean_pwa_var
colour
bc
cal ensemble
m1
m2
sbt


Accuracy improvement of ensemble over the best of networks vs err_incons. Ensemble size 6
acc_imp_max
err_incons
0.350
0.375
0.400
0.425
0.000
0.025
0.050
0.075
0.100
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
colour
bc
cal ensemble
m1
m2
sbt

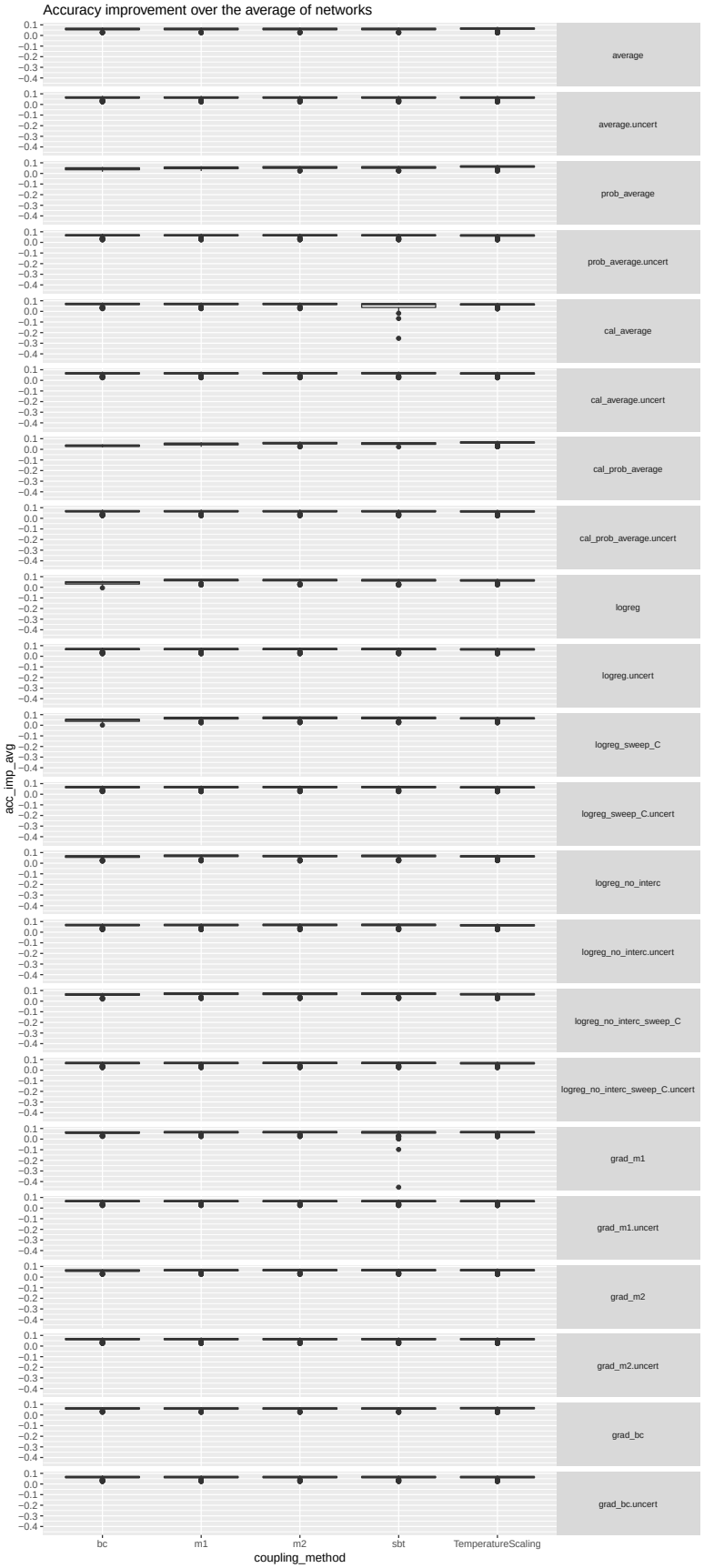

Accuracy improvement of ensemble over
the average of networks vs err_incons.
Ensemble size 6
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
acc_imp_avg
0.100
0.075
0.050
0.025
0.000
0.350
0.375
0.400
0.425
err_incons
colour
bc
cal ensemble
m1
m2
sbt


Accuracy improvement of ensemble over
the best of networks vs mean_pwa_var.
Ensemble size 6
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
acc_imp_max
mean_pwa_var
colour
bc
cal ensemble
m1
m2
sbt

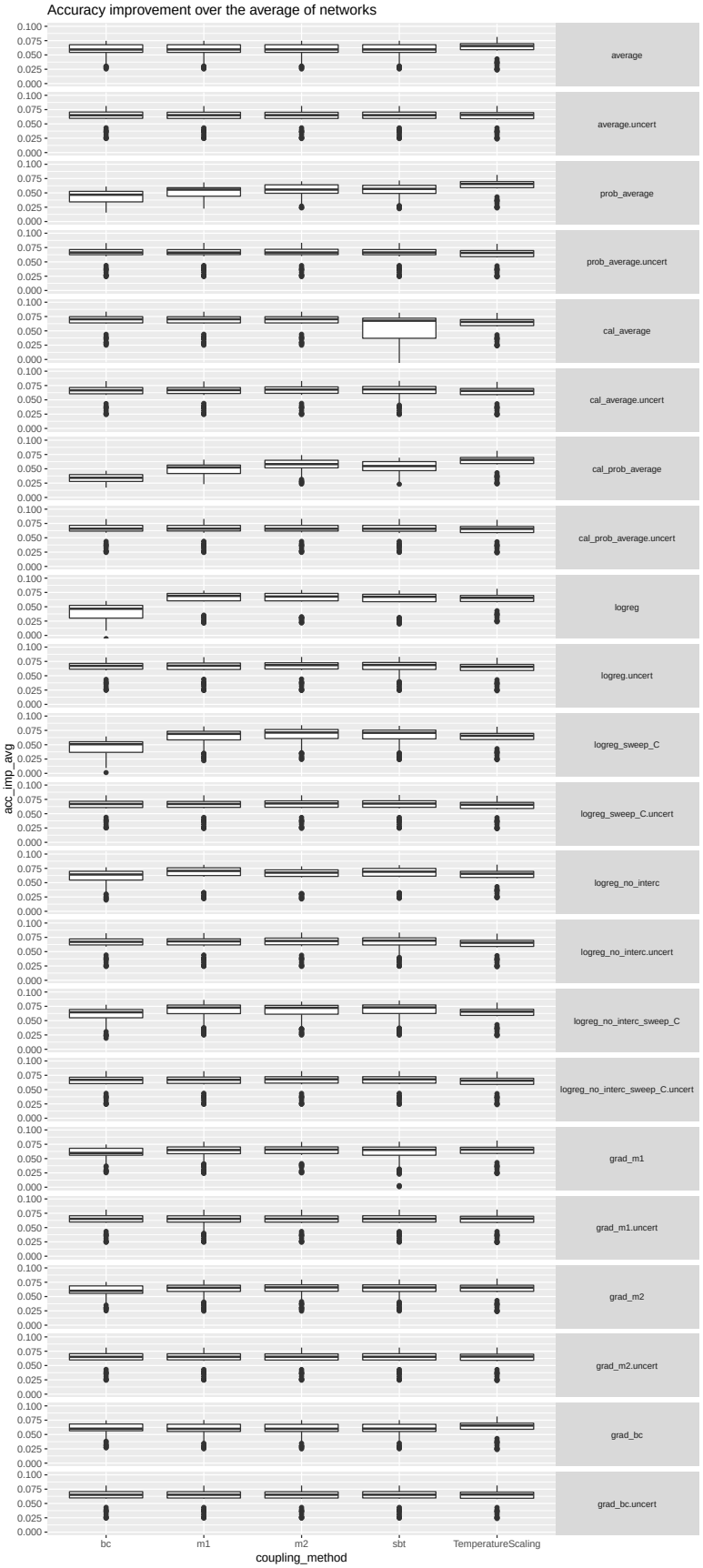

Accuracy improvement of ensemble over
the average of networks vs mean_pwa_var.
Ensemble size 6
acc_imp_avg
mean_pwa_var
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
colour
bc
cal ensemble
m1
m2
sbt

```r
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))
)
```

```r
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)
```


Accuracy improvement over the average of networks
acc_imp_avg
0.1
0.0
−0.1
−0.2
−0.3
−0.4
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
bc
m1
m2
sbt
TemperatureScaling
coupling_method

```
print(imp_avg_plot + coord_cartesian(ylim=c(0, 0.1)))
```


Accuracy improvement over the average of networks
acc_imp_avg
0.100
0.075
0.050
0.025
0.000
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
bc
m1
m2
sbt
TemperatureScaling
coupling_method

Sorted by average improvement over the average of networks.

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

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Wed Feb 09 10:34:06 2022

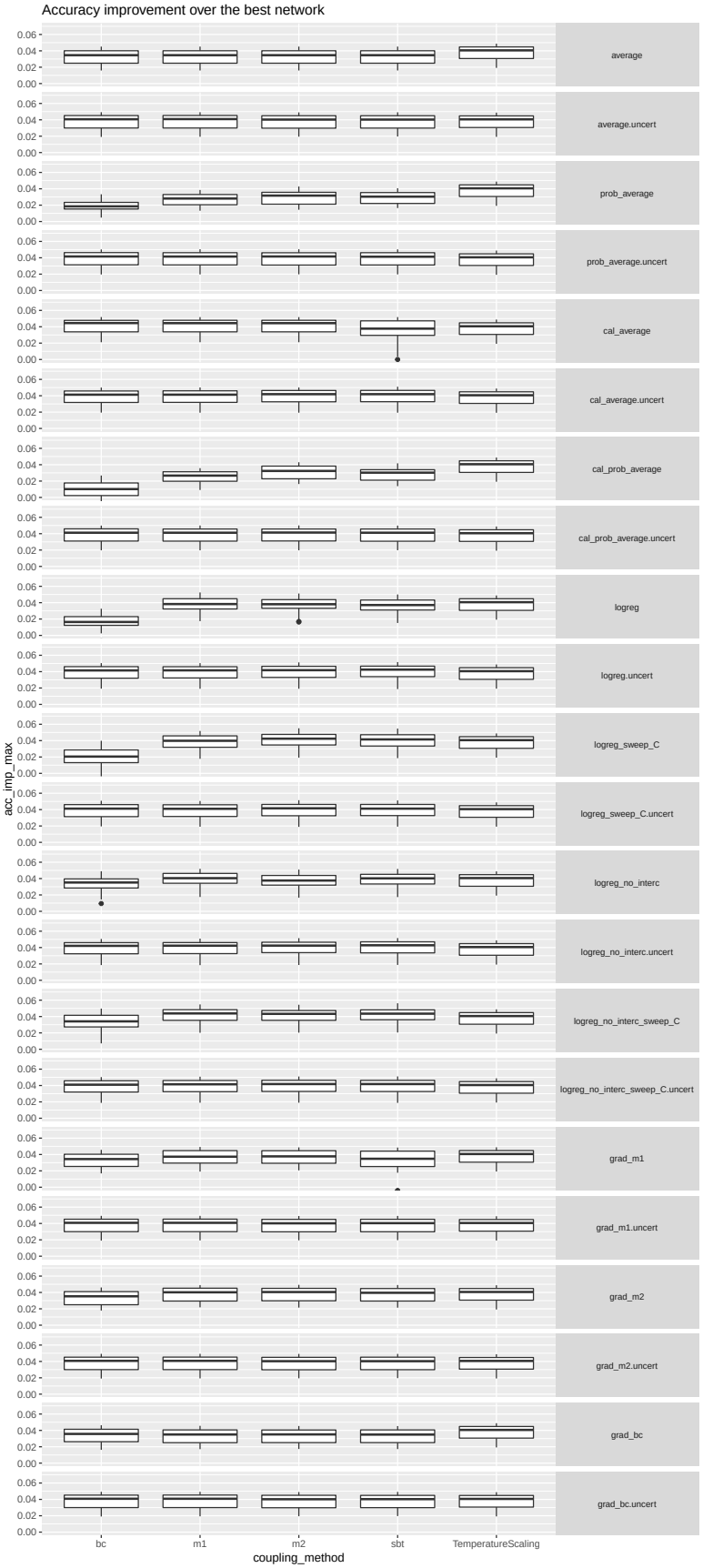
	method	imp_o_avg	imp_o_max
1	logreg_no_interc_sweep_C sbt	0.0653	0.0413
2	logreg_no_interc_sweep_C m1	0.0651	0.0411
3	cal_average m1	0.0646	0.0407
4	cal_average m2	0.0646	0.0407
5	cal_average bc	0.0646	0.0407
6	logreg_no_interc_sweep_C m2	0.0644	0.0404
7	logreg_sweep_C m2	0.0640	0.0400
8	logreg_sweep_C sbt	0.0635	0.0396
9	logreg_no_interc.uncert sbt	0.0635	0.0396
10	logreg_no_interc.uncert m2	0.0633	0.0394
11	logreg.uncert sbt	0.0633	0.0393
12	logreg.uncert m2	0.0632	0.0392
13	cal_average.uncert sbt	0.0629	0.0390
14	logreg_no_interc_sweep_C.uncert sbt	0.0629	0.0389
15	logreg_no_interc_sweep_C.uncert m2	0.0628	0.0389
16	logreg_no_interc m1	0.0627	0.0388
17	logreg_sweep_C.uncert m2	0.0627	0.0388
18	logreg_no_interc.uncert m1	0.0627	0.0388
19	logreg_sweep_C.uncert sbt	0.0627	0.0388
20	cal_average.uncert m2	0.0626	0.0387
21	logreg_no_interc.uncert bc	0.0625	0.0386
22	logreg.uncert m1	0.0625	0.0385
23	logreg_no_interc_sweep_C.uncert m1	0.0625	0.0385
24	logreg.uncert bc	0.0624	0.0385
25	logreg_sweep_C.uncert m1	0.0623	0.0384
26	logreg_no_interc_sweep_C.uncert bc	0.0623	0.0383
27	prob_average.uncert m1	0.0622	0.0383
28	prob_average.uncert bc	0.0622	0.0383
29	prob_average.uncert m2	0.0622	0.0383
30	logreg_sweep_C.uncert bc	0.0622	0.0383
31	cal_average.uncert m1	0.0622	0.0383
32	prob_average.uncert sbt	0.0621	0.0382
33	logreg_sweep_C m1	0.0621	0.0381
34	logreg_no_interc sbt	0.0620	0.0381
35	cal_average.uncert bc	0.0620	0.0381
36	cal_prob_average.uncert m2	0.0620	0.0380
37	cal_prob_average.uncert bc	0.0619	0.0380
38	cal_prob_average.uncert m1	0.0619	0.0379
39	cal_prob_average.uncert sbt	0.0618	0.0379
40	logreg m1	0.0614	0.0375
41	grad_bc.uncert bc	0.0613	0.0373
42	grad_m1.uncert bc	0.0613	0.0373
43	average.uncert bc	0.0613	0.0373
44	grad_m2.uncert bc	0.0613	0.0373
45	grad_m2.uncert m1	0.0613	0.0373
46	grad_m1.uncert sbt	0.0612	0.0373

47	grad_m2.uncert sbt	0.0612	0.0373
48	grad_m1.uncert m1	0.0612	0.0373
49	grad_bc.uncert sbt	0.0612	0.0373
50	grad_bc.uncert m1	0.0612	0.0373
51	average.uncert sbt	0.0612	0.0373
52	average.uncert m1	0.0612	0.0373
53	grad_m2.uncert m2	0.0612	0.0373
54	average.uncert m2	0.0612	0.0373
55	grad_m1.uncert m2	0.0612	0.0373
56	grad_bc.uncert m2	0.0612	0.0372
57	average of TemperatureScaling	0.0611	0.0372
58	grad_m2 m2	0.0611	0.0372
59	grad_m2 sbt	0.0610	0.0371
60	logreg m2	0.0610	0.0370
61	grad_m2 m1	0.0608	0.0369
62	logreg_no_interc m2	0.0605	0.0366
63	grad_m1 m2	0.0603	0.0364
64	grad_m1 m1	0.0603	0.0363
65	logreg sbt	0.0596	0.0357
66	logreg_no_interc_sweep_C bc	0.0580	0.0340
67	logreg_no_interc bc	0.0577	0.0338
68	grad_bc bc	0.0576	0.0337
69	grad_m2 bc	0.0574	0.0335
70	grad_bc m2	0.0573	0.0333
71	grad_bc m1	0.0573	0.0333
72	grad_bc sbt	0.0572	0.0333
73	grad_m1 bc	0.0572	0.0332
74	average m1	0.0567	0.0327
75	average m2	0.0567	0.0327
76	average sbt	0.0567	0.0327
77	average bc	0.0567	0.0327
78	cal_prob_average m2	0.0551	0.0311
79	prob_average sbt	0.0532	0.0292
80	prob_average m2	0.0530	0.0291
81	cal_prob_average sbt	0.0517	0.0277
82	prob_average m1	0.0510	0.0271
83	cal_average sbt	0.0498	0.0258
84	cal_prob_average m1	0.0490	0.0250
85	grad_m1 sbt	0.0474	0.0234
86	logreg_sweep_C bc	0.0443	0.0204
87	prob_average bc	0.0431	0.0191
88	logreg bc	0.0408	0.0169
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)
```


Accuracy improvement over the best network
acc_imp_max
0.0
−0.1
−0.2
−0.3
−0.4
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
bc
m1
m2
sbt
TemperatureScaling
coupling_method

```
print(imp_max_plot + coord_cartesian(ylim=c(0, 0.07)))
```


Accuracy improvement over the best network
acc_imp_max
0.06
0.04
0.02
0.00
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
bc
m1
m2
sbt
TemperatureScaling
coupling_method

Sorted by average improvement over the best of the networks.

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

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Wed Feb 09 10:34:10 2022

	method	imp_o_avg	imp_o_max
1	logreg_no_interc_sweep_C sbt	0.0653	0.0413
2	logreg_no_interc_sweep_C m1	0.0651	0.0411
3	cal_average m1	0.0646	0.0407
4	cal_average m2	0.0646	0.0407
5	cal_average bc	0.0646	0.0407
6	logreg_no_interc_sweep_C m2	0.0644	0.0404
7	logreg_sweep_C m2	0.0640	0.0400
8	logreg_sweep_C sbt	0.0635	0.0396
9	logreg_no_interc.uncert sbt	0.0635	0.0396
10	logreg_no_interc.uncert m2	0.0633	0.0394
11	logreg.uncert sbt	0.0633	0.0393
12	logreg.uncert m2	0.0632	0.0392
13	cal_average.uncert sbt	0.0629	0.0390
14	logreg_no_interc_sweep_C.uncert sbt	0.0629	0.0389
15	logreg_no_interc_sweep_C.uncert m2	0.0628	0.0389
16	logreg_no_interc m1	0.0627	0.0388
17	logreg_sweep_C.uncert m2	0.0627	0.0388
18	logreg_no_interc.uncert m1	0.0627	0.0388
19	logreg_sweep_C.uncert sbt	0.0627	0.0388
20	cal_average.uncert m2	0.0626	0.0387
21	logreg_no_interc.uncert bc	0.0625	0.0386
22	logreg.uncert m1	0.0625	0.0385
23	logreg_no_interc_sweep_C.uncert m1	0.0625	0.0385
24	logreg.uncert bc	0.0624	0.0385
25	logreg_sweep_C.uncert m1	0.0623	0.0384
26	logreg_no_interc_sweep_C.uncert bc	0.0623	0.0383
27	prob_average.uncert m1	0.0622	0.0383
28	prob_average.uncert bc	0.0622	0.0383
29	prob_average.uncert m2	0.0622	0.0383
30	logreg_sweep_C.uncert bc	0.0622	0.0383
31	cal_average.uncert m1	0.0622	0.0383
32	prob_average.uncert sbt	0.0621	0.0382
33	logreg_sweep_C m1	0.0621	0.0381
34	logreg_no_interc sbt	0.0620	0.0381
35	cal_average.uncert bc	0.0620	0.0381
36	cal_prob_average.uncert m2	0.0620	0.0380
37	cal_prob_average.uncert bc	0.0619	0.0380
38	cal_prob_average.uncert m1	0.0619	0.0379
39	cal_prob_average.uncert sbt	0.0618	0.0379
40	logreg m1	0.0614	0.0375
41	grad_bc.uncert bc	0.0613	0.0373
42	grad_m1.uncert bc	0.0613	0.0373
43	average.uncert bc	0.0613	0.0373
44	grad_m2.uncert bc	0.0613	0.0373
45	grad_m2.uncert m1	0.0613	0.0373
46	grad_m1.uncert sbt	0.0612	0.0373

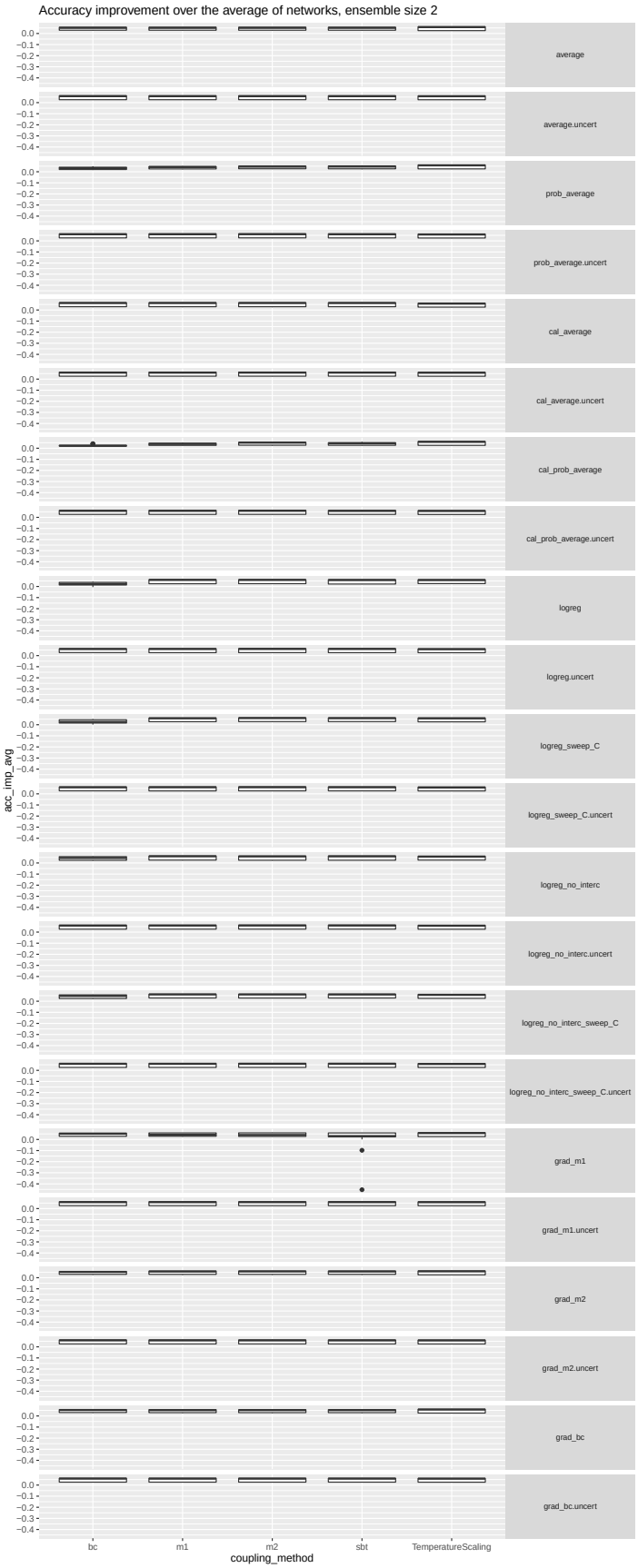
47	grad_m2.uncert sbt	0.0612	0.0373
48	grad_m1.uncert m1	0.0612	0.0373
49	grad_bc.uncert sbt	0.0612	0.0373
50	grad_bc.uncert m1	0.0612	0.0373
51	average.uncert sbt	0.0612	0.0373
52	average.uncert m1	0.0612	0.0373
53	grad_m2.uncert m2	0.0612	0.0373
54	average.uncert m2	0.0612	0.0373
55	grad_m1.uncert m2	0.0612	0.0373
56	grad_bc.uncert m2	0.0612	0.0372
57	average of TemperatureScaling	0.0611	0.0372
58	grad_m2 m2	0.0611	0.0372
59	grad_m2 sbt	0.0610	0.0371
60	logreg m2	0.0610	0.0370
61	grad_m2 m1	0.0608	0.0369
62	logreg_no_interc m2	0.0605	0.0366
63	grad_m1 m2	0.0603	0.0364
64	grad_m1 m1	0.0603	0.0363
65	logreg sbt	0.0596	0.0357
66	logreg_no_interc_sweep_C bc	0.0580	0.0340
67	logreg_no_interc bc	0.0577	0.0338
68	grad_bc bc	0.0576	0.0337
69	grad_m2 bc	0.0574	0.0335
70	grad_bc m2	0.0573	0.0333
71	grad_bc m1	0.0573	0.0333
72	grad_bc sbt	0.0572	0.0333
73	grad_m1 bc	0.0572	0.0332
74	average m1	0.0567	0.0327
75	average m2	0.0567	0.0327
76	average sbt	0.0567	0.0327
77	average bc	0.0567	0.0327
78	cal_prob_average m2	0.0551	0.0311
79	prob_average sbt	0.0532	0.0292
80	prob_average m2	0.0530	0.0291
81	cal_prob_average sbt	0.0517	0.0277
82	prob_average m1	0.0510	0.0271
83	cal_average sbt	0.0498	0.0258
84	cal_prob_average m1	0.0490	0.0250
85	grad_m1 sbt	0.0474	0.0234
86	logreg_sweep_C bc	0.0443	0.0204
87	prob_average bc	0.0431	0.0191
88	logreg bc	0.0408	0.0169
89	cal_prob_average bc	0.0336	0.0096

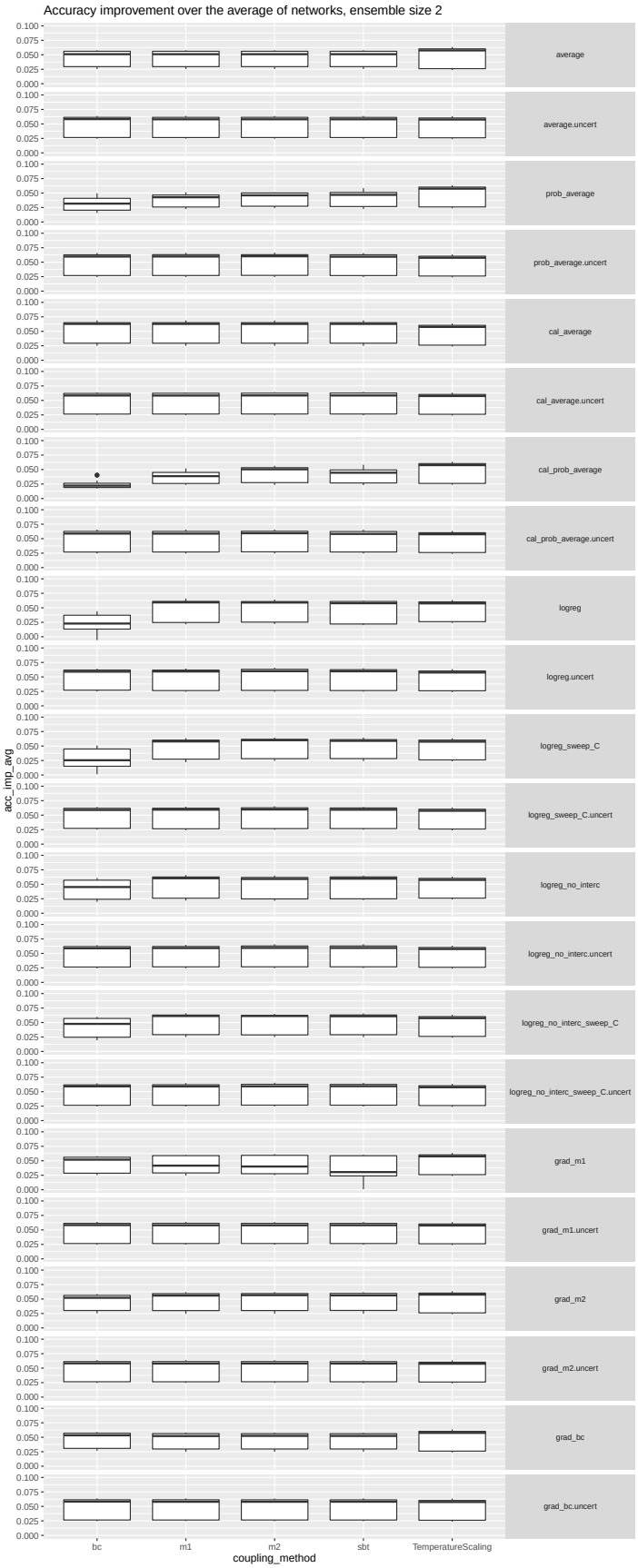
```
avg_imp_table_cs <- rbind(
  ens_pwc_plt_df %>%
    mutate(method = paste(combining_method, coupling_method, sep=" ")) %>%
    group_by(method, combination_size) %>%
    summarise(imp_o_avg = mean(acc_imp_avg), imp_o_max = mean(acc_imp_max)),
  ens_cal_plt_df %>%
    mutate(method = paste0("average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

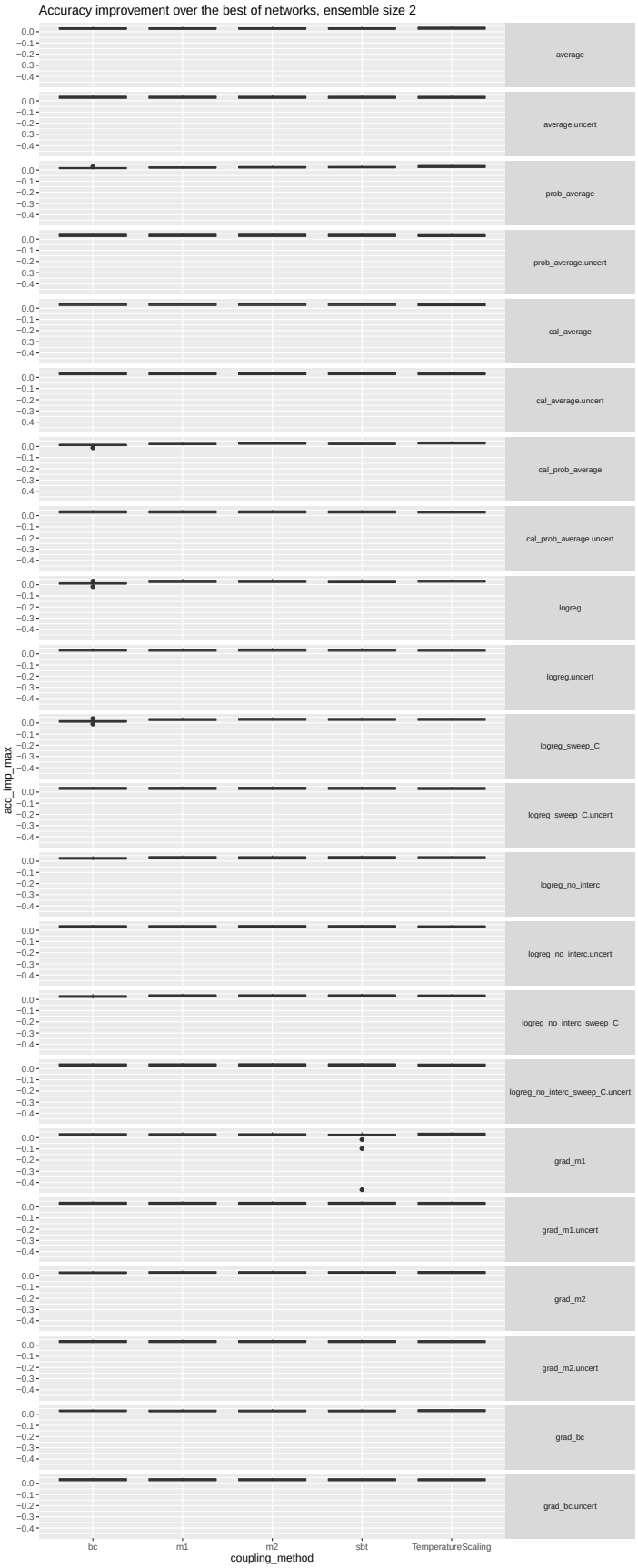
```
)
```

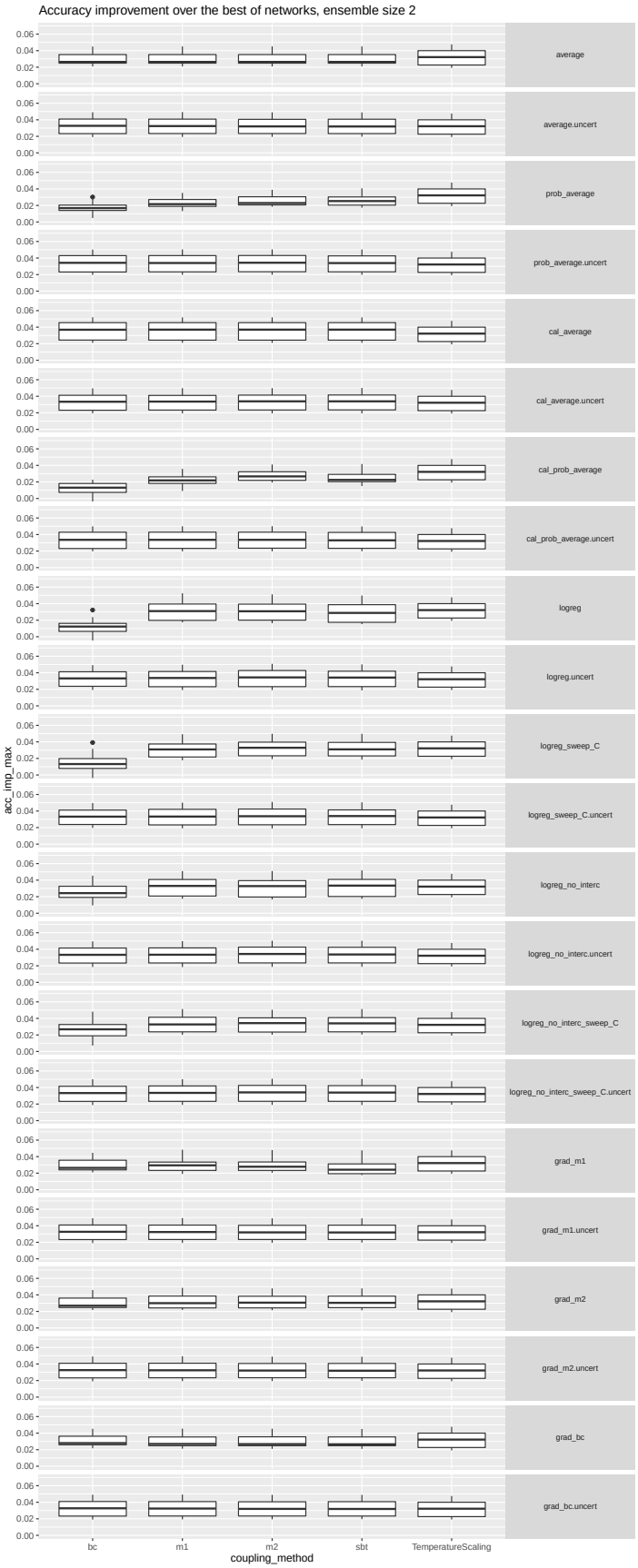
```
## `summarise()` has grouped output by 'method'. You can override using the `.groups` argument.
## `summarise()` has grouped output by 'method'. You can override using the `.groups` argument.
```

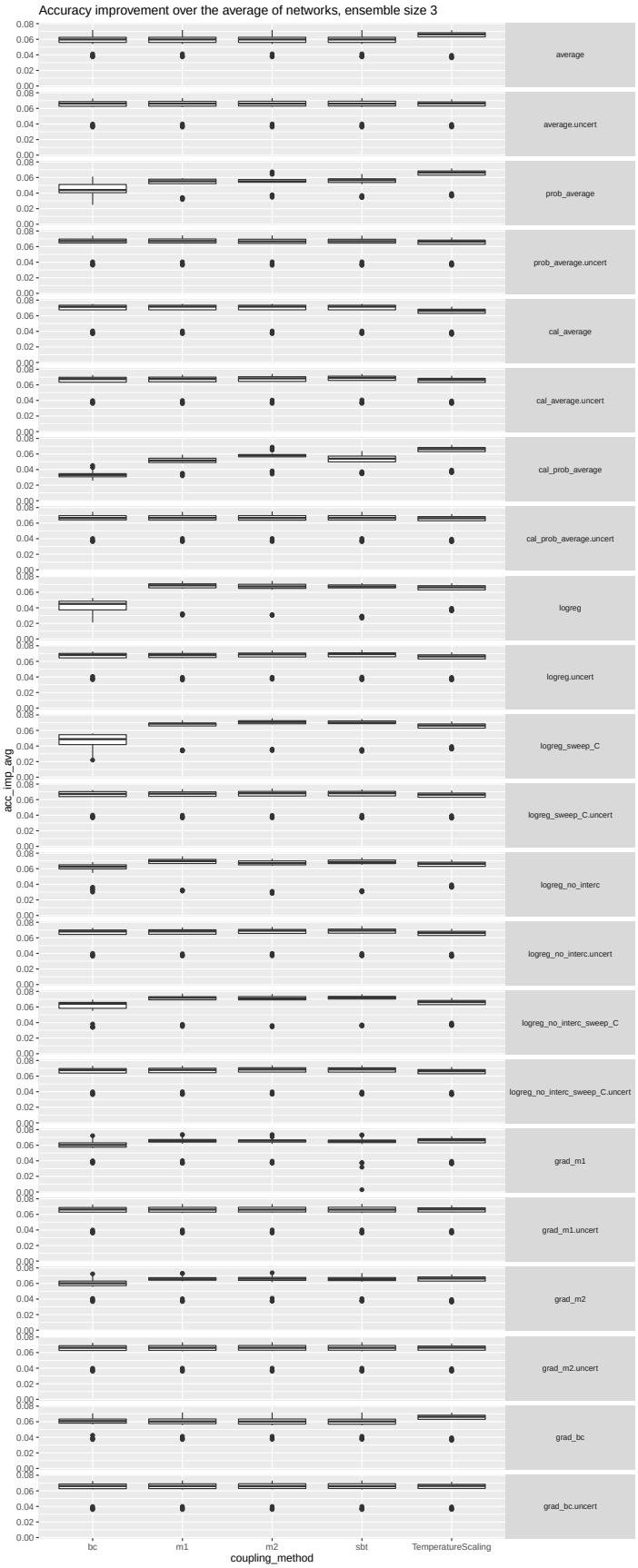
```
for (sss in unique(ens_cal_plt_df$combination_size))
{
  imp_avg_plot <- ggplot() +
    geom_boxplot(data = ens_pwc_plt_df %>% filter(combination_size == sss), mapping = aes(x = coupling_m
    geom_boxplot(data = ens_cal_plt_df %>% filter(combination_size == sss), mapping = aes(x = calibrati
    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)
  print(imp_avg_plot + coord_cartesian(ylim=c(0, 0.1)))

  imp_max_plot <- ggplot() +
    geom_boxplot(data = ens_pwc_plt_df %>% filter(combination_size == sss), mapping = aes(x = coupling_m
    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)
  print(imp_max_plot + coord_cartesian(ylim=c(0, 0.07)))

}
```

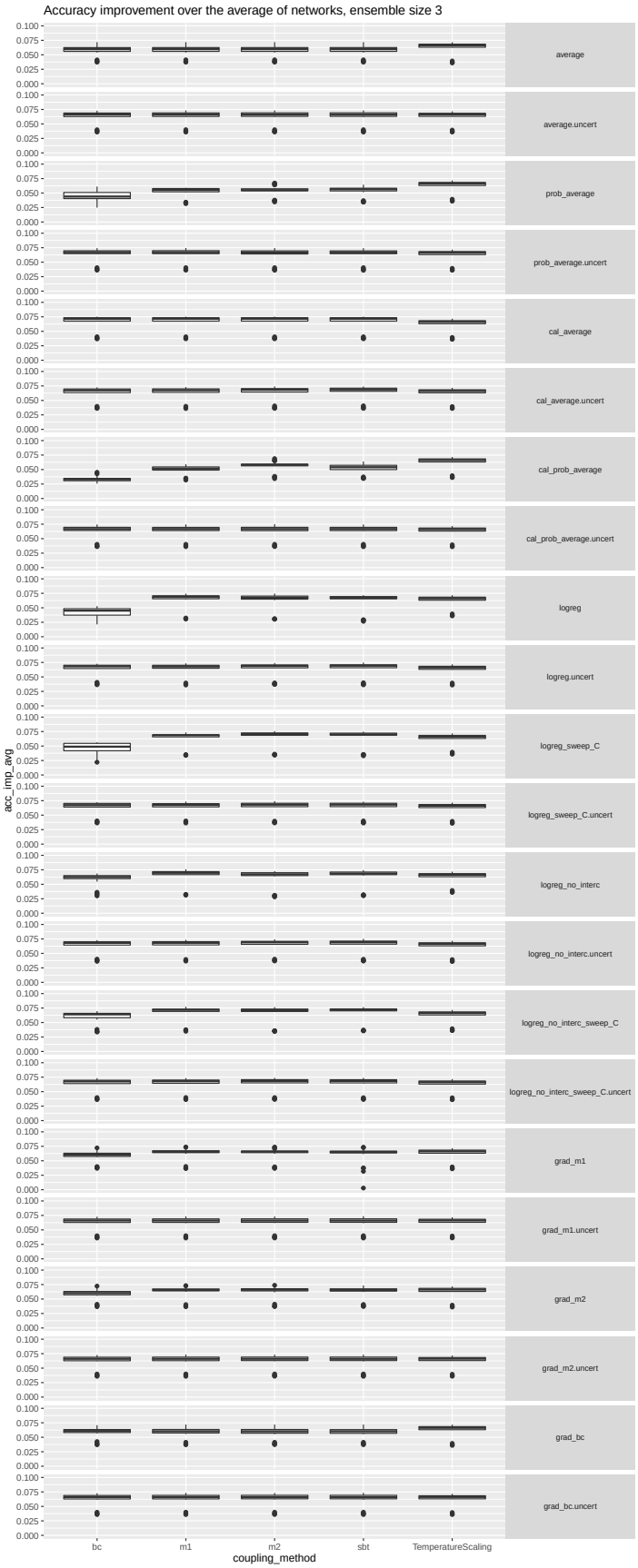

Accuracy improvement over the average of networks, ensemble size 2
acc_imp_avg
0.0
−0.1
−0.2
−0.3
−0.4
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
bc
m1
m2
sbt
TemperatureScaling
coupling_method

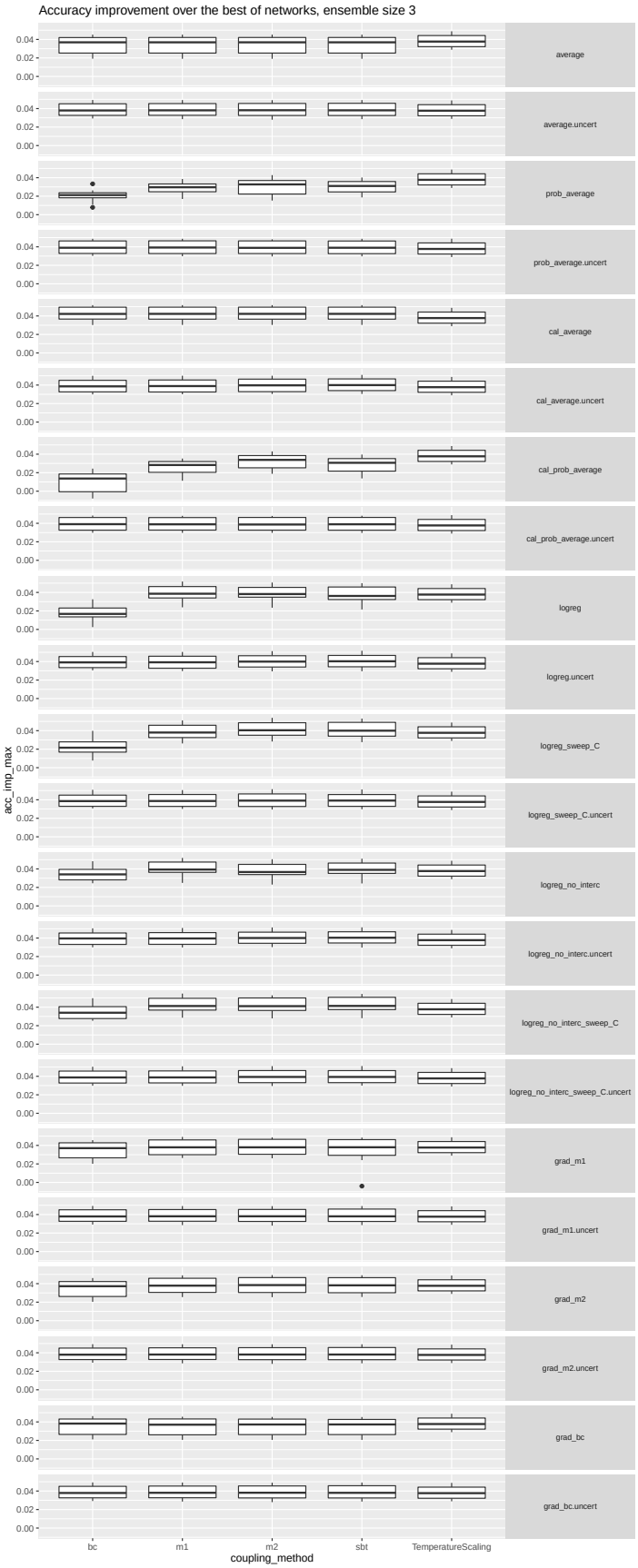

Accuracy improvement over the average of networks, ensemble size 2
acc_imp_avg
0.100
0.075
0.050
0.025
0.000
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
bc
m1
m2
sbt
TemperatureScaling
coupling_method

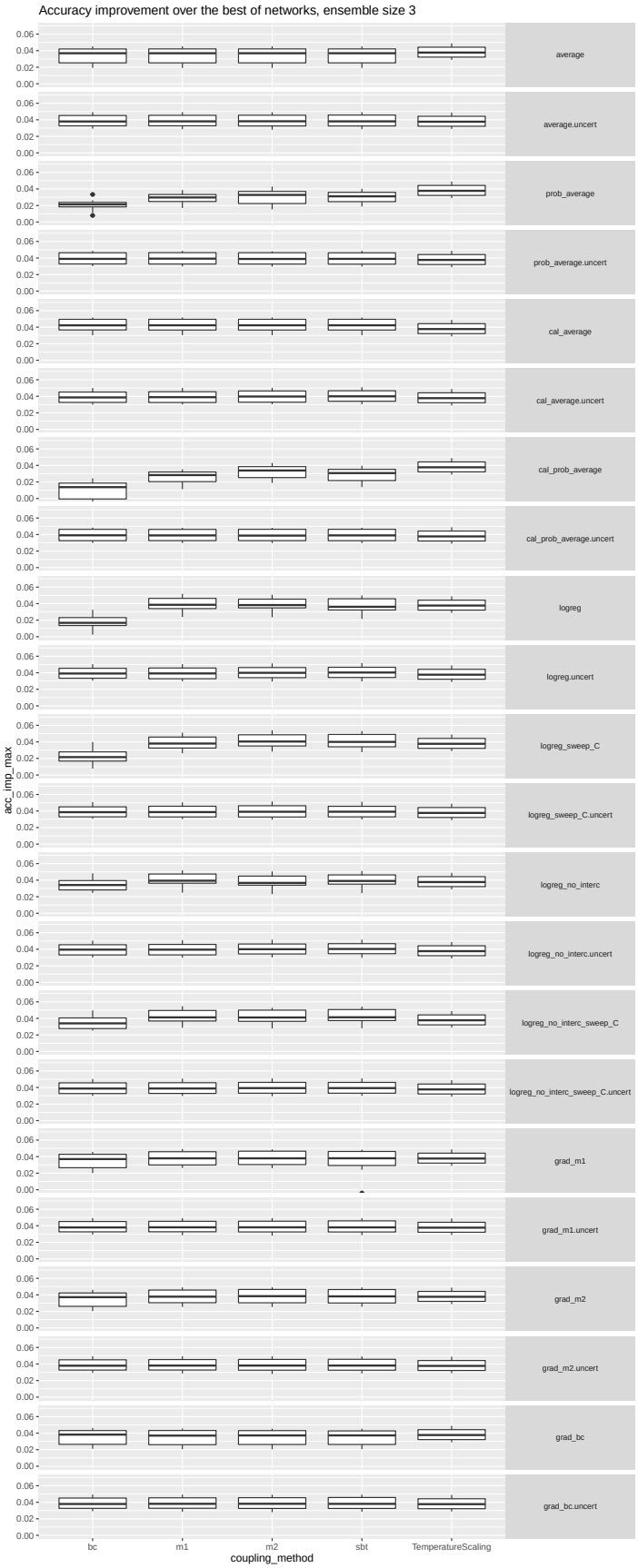

Accuracy improvement over the best of networks, ensemble size 2
acc_imp_max
0.0
−0.1
−0.2
−0.3
−0.4
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
bc
m1
m2
sbt
TemperatureScaling
coupling_method

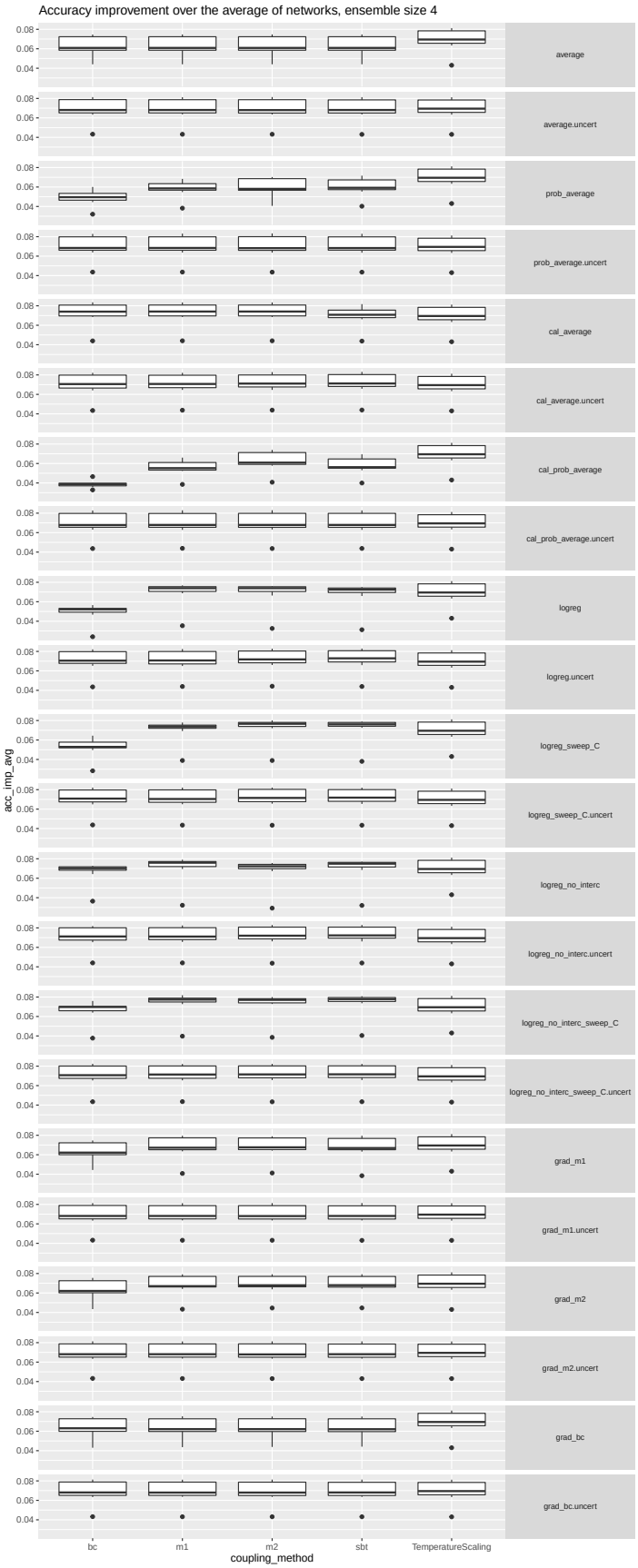

Accuracy improvement over the best of networks, ensemble size 2
acc_imp_max
0.06
0.04
0.02
0.00
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
bc
m1
m2
sbt
TemperatureScaling
coupling_method

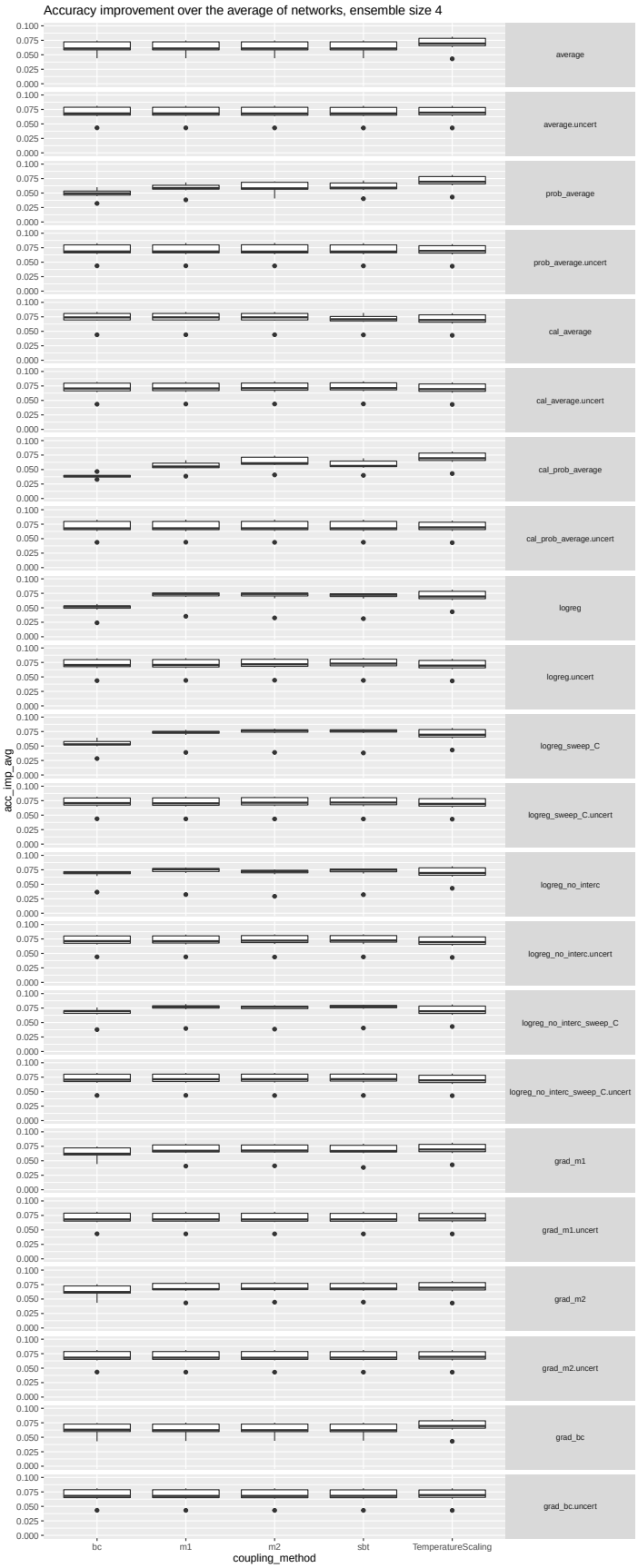

Accuracy improvement over the average of networks, ensemble size 3
acc_imp_avg
coupling_method
bc
m1
m2
sbt
TemperatureScaling
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert

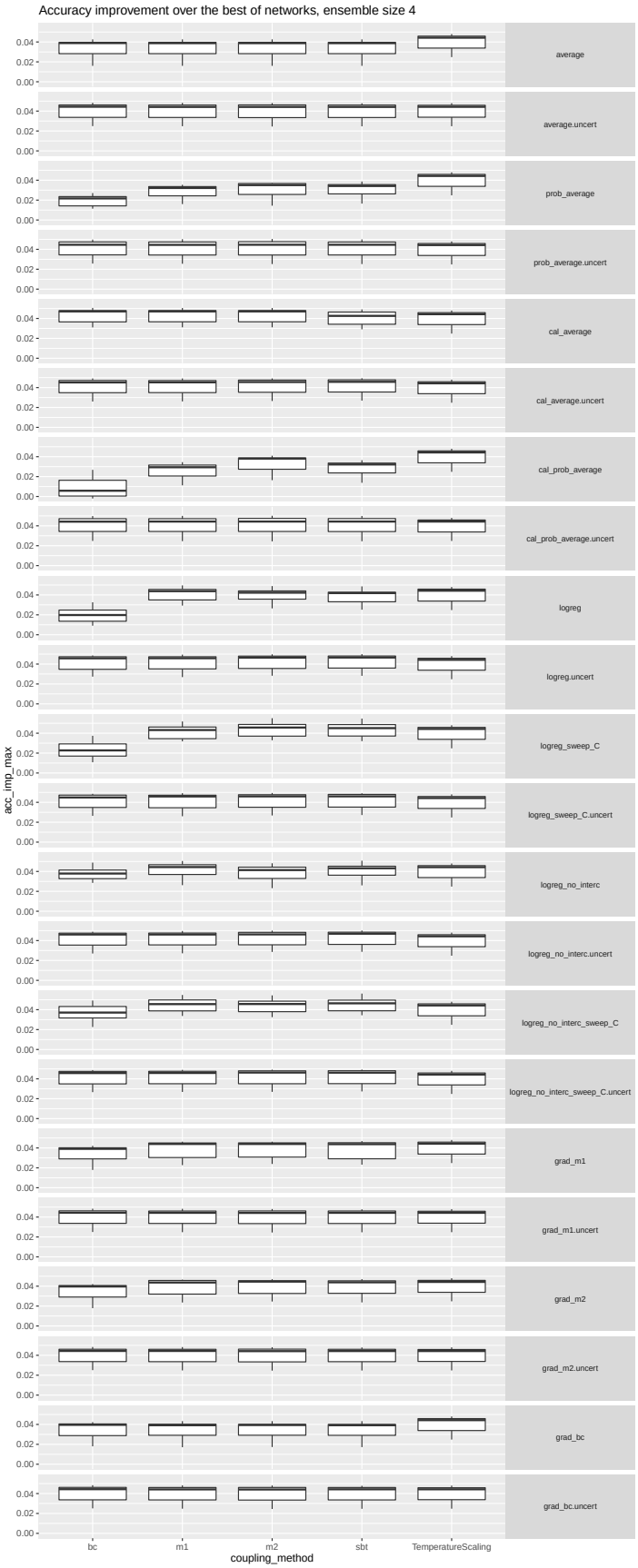

Accuracy improvement over the average of networks, ensemble size 3
acc_imp_avg
0.100
0.075
0.050
0.025
0.000
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
bc
m1
m2
sbt
TemperatureScaling
coupling_method

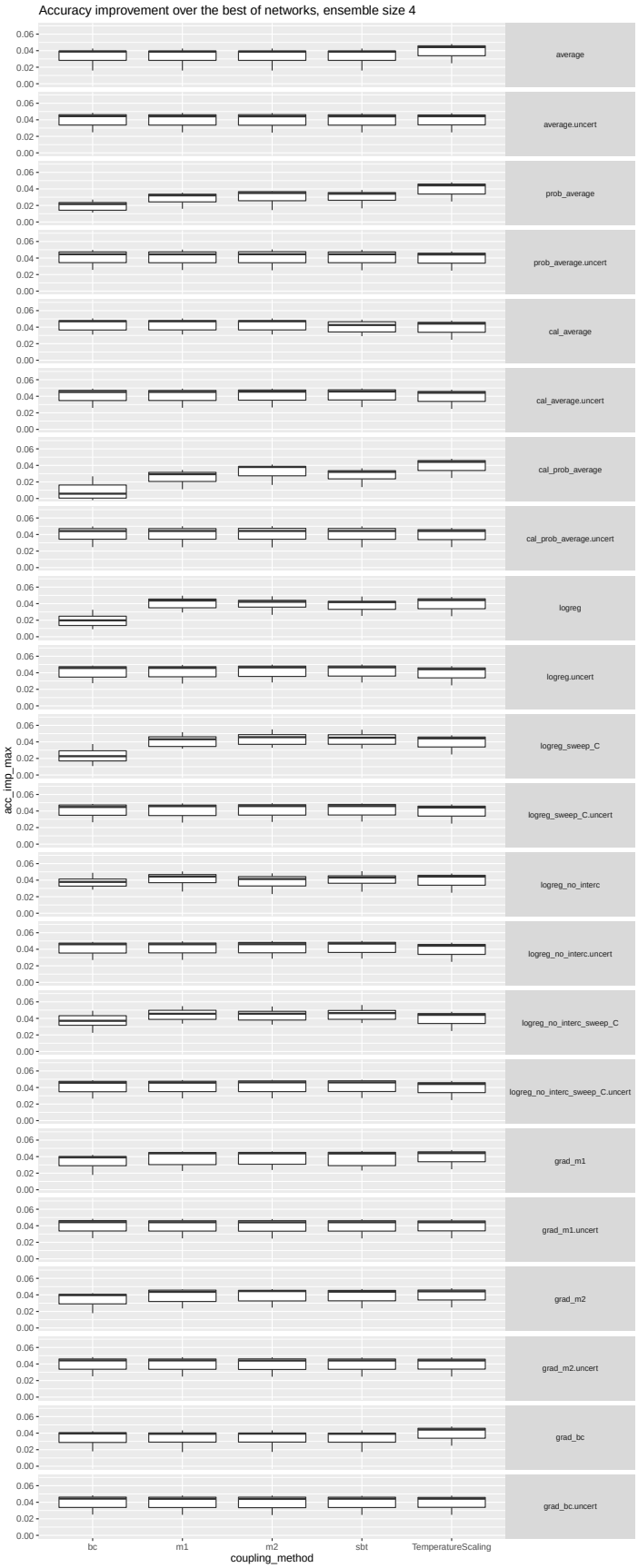

Accuracy improvement over the best of networks, ensemble size 3
acc_imp_max
0.04
0.02
0.00
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
bc
m1
m2
sbt
TemperatureScaling
coupling_method

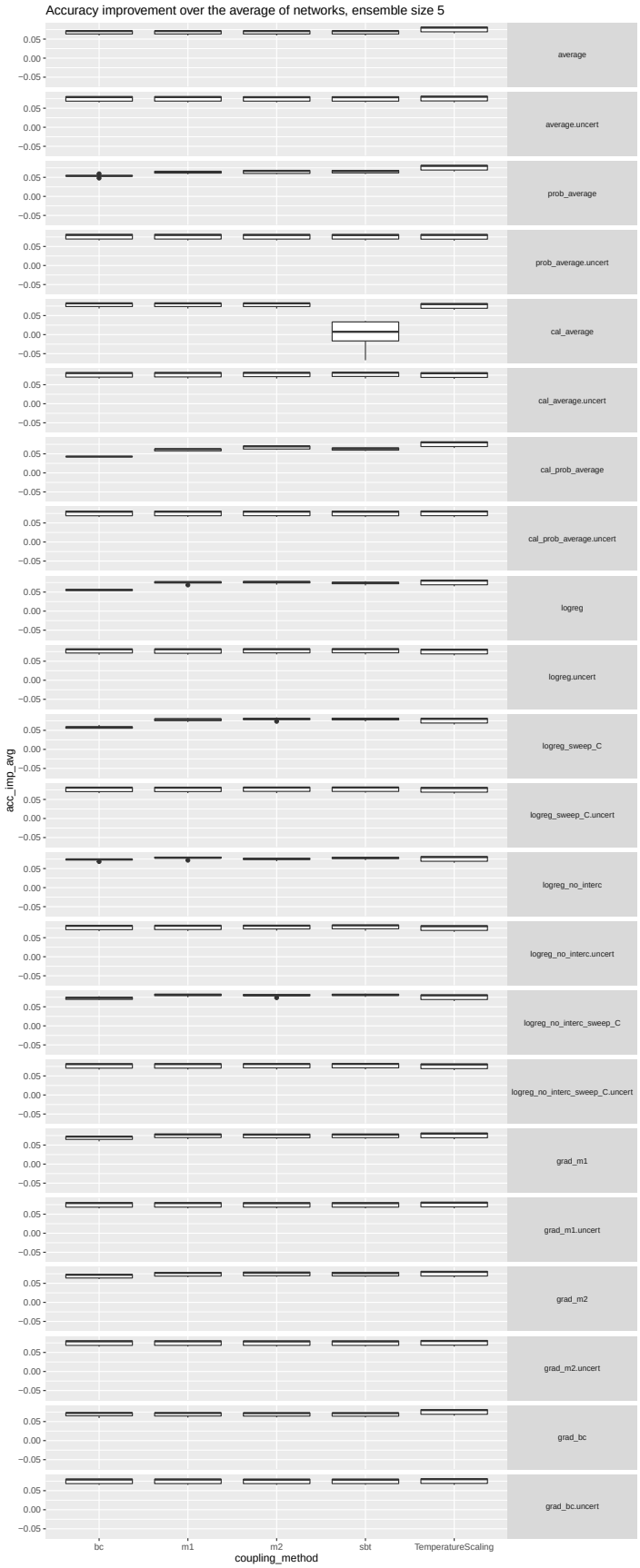

Accuracy improvement over the best of networks, ensemble size 3
acc_imp_max
0.06
0.04
0.02
0.00
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
bc
m1
m2
sbt
TemperatureScaling
coupling_method

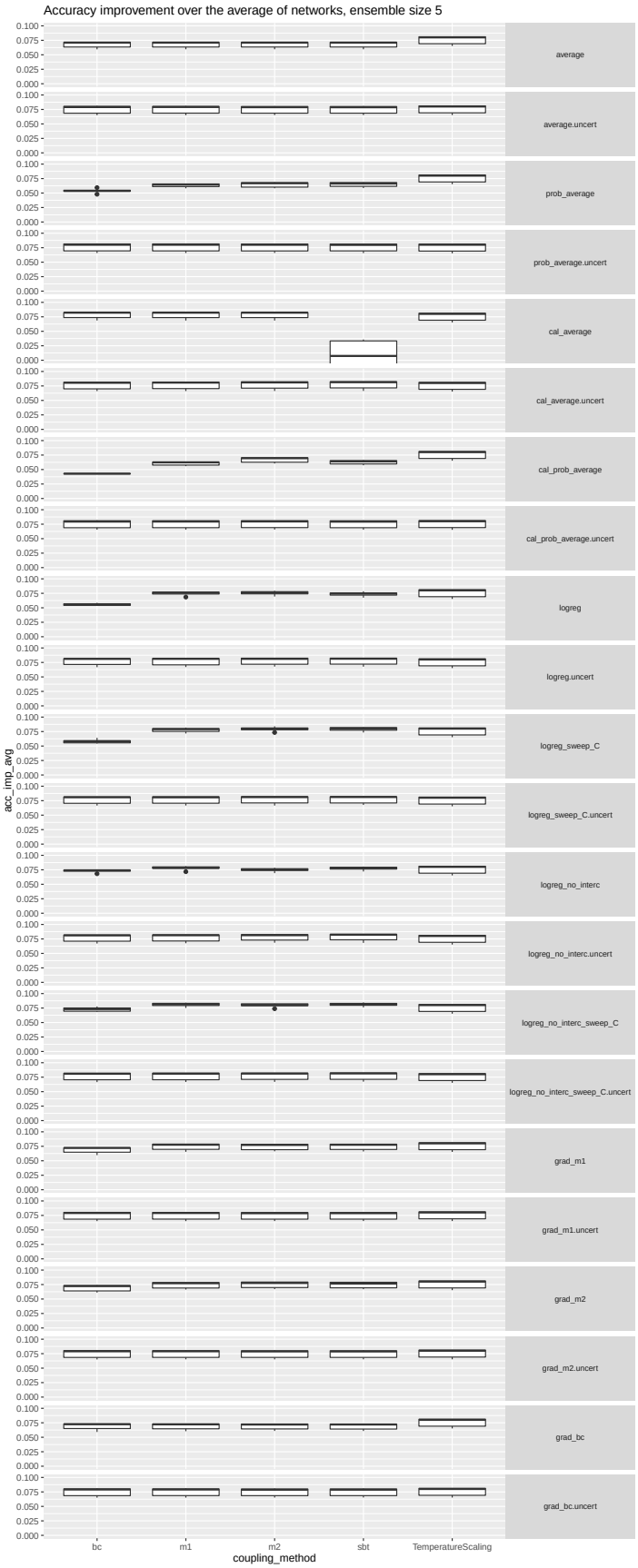

Accuracy improvement over the average of networks, ensemble size 4
acc_imp_avg
0.08
0.06
0.04
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
bc
m1
m2
sbt
TemperatureScaling
coupling_method

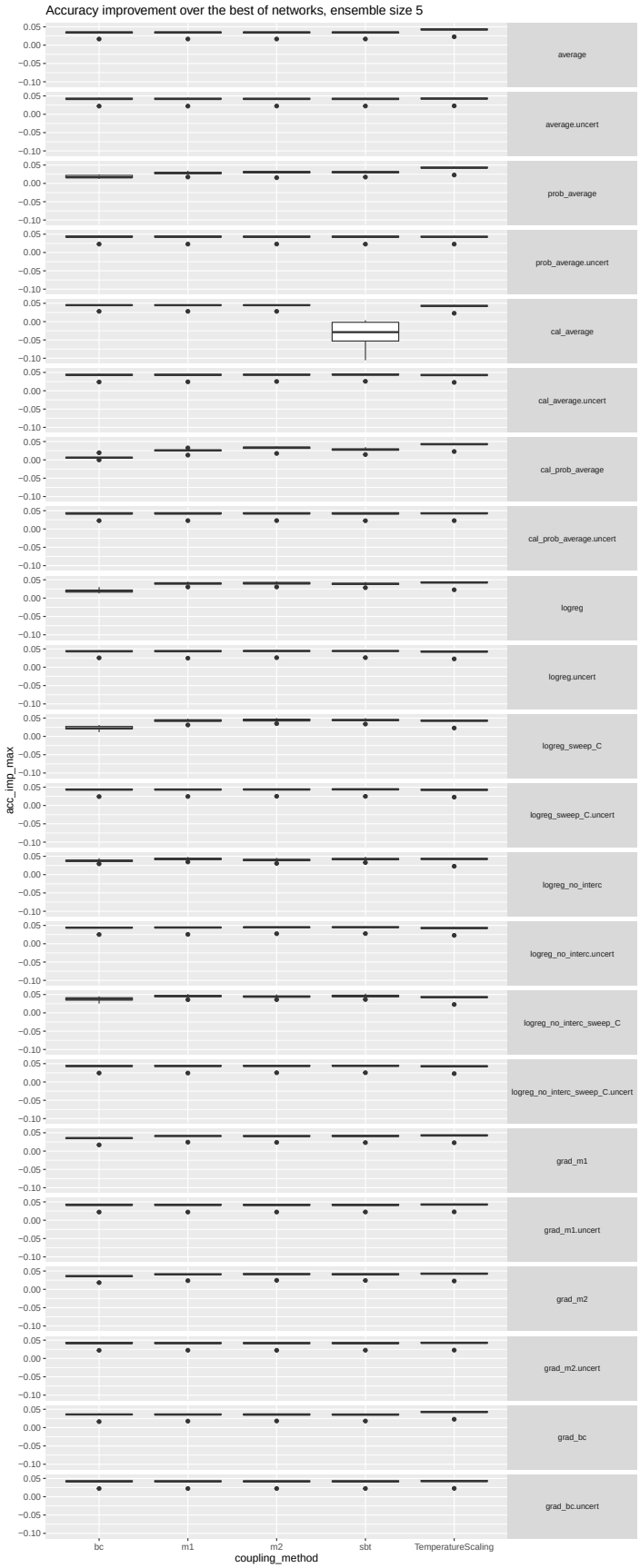

Accuracy improvement over the average of networks, ensemble size 4
acc_imp_avg
0.100
0.075
0.050
0.025
0.000
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
bc
m1
m2
sbt
TemperatureScaling
coupling_method

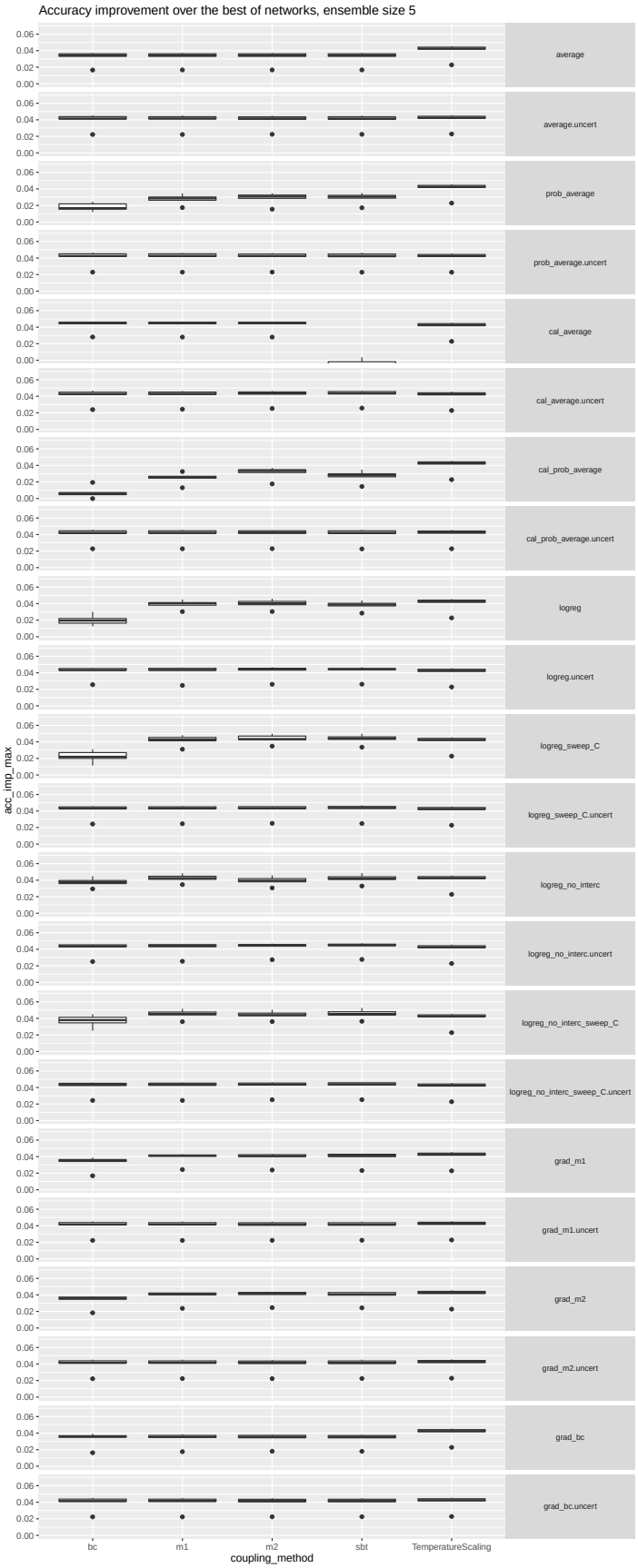

Accuracy improvement over the best of networks, ensemble size 4
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
0.04
0.02
0.00
acc_imp_max
bc
m1
m2
sbt
TemperatureScaling
coupling_method

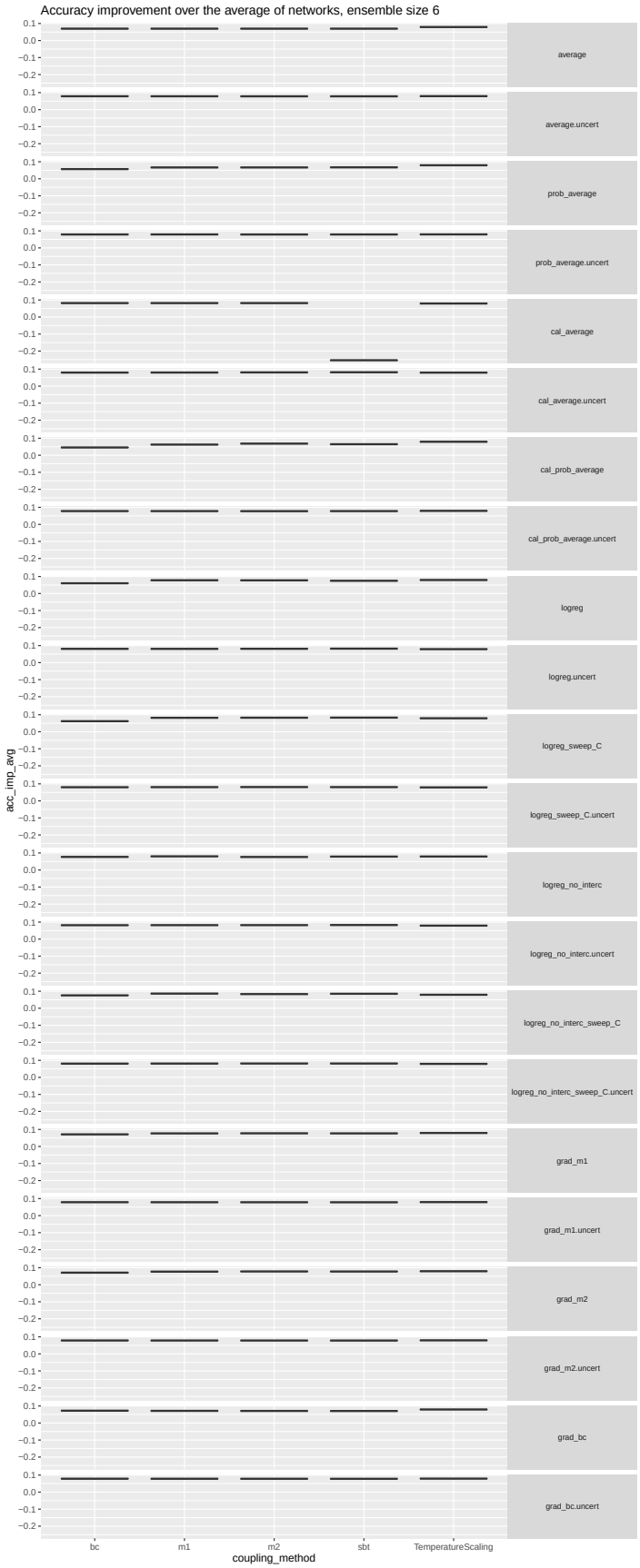

Accuracy improvement over the best of networks, ensemble size 4
acc_imp_max
0.06
0.04
0.02
0.00
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
bc
m1
m2
sbt
TemperatureScaling
coupling_method

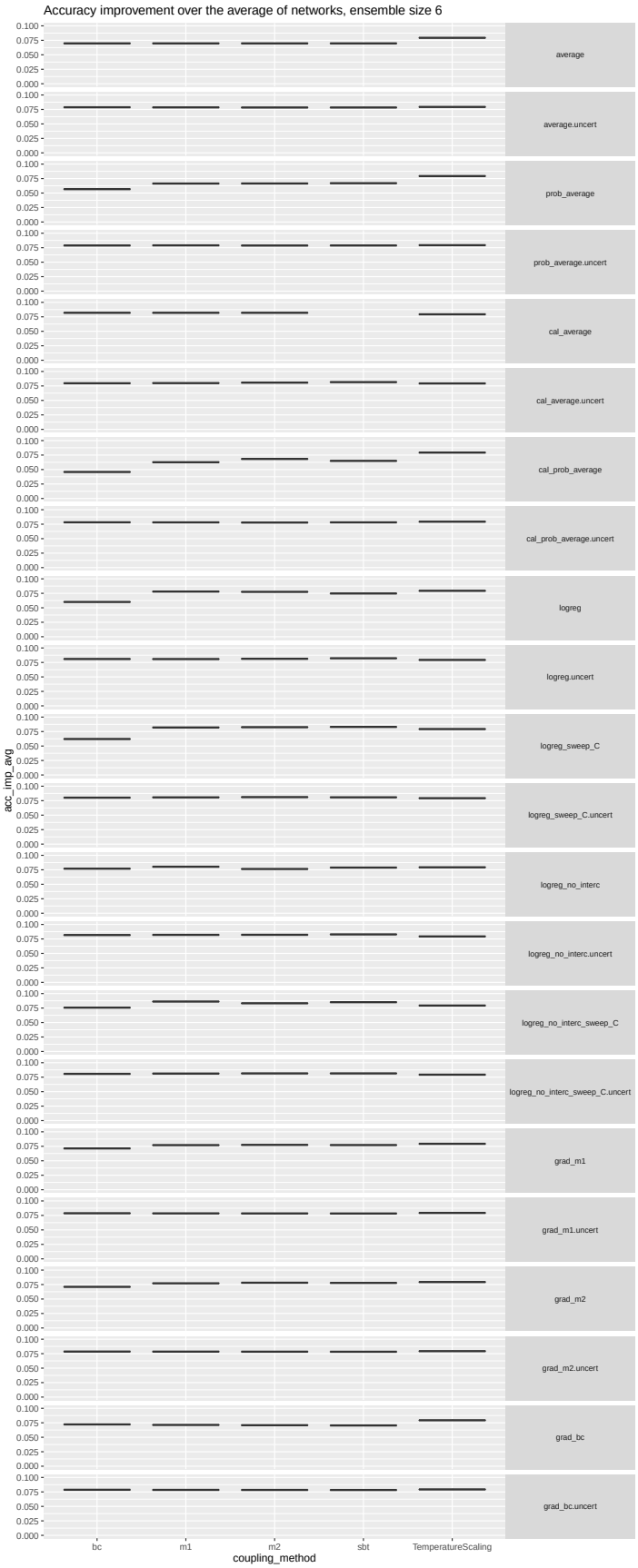

Accuracy improvement over the average of networks, ensemble size 5
acc_imp_avg
0.05
0.00
−0.05
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
bc
m1
m2
sbt
TemperatureScaling
coupling_method

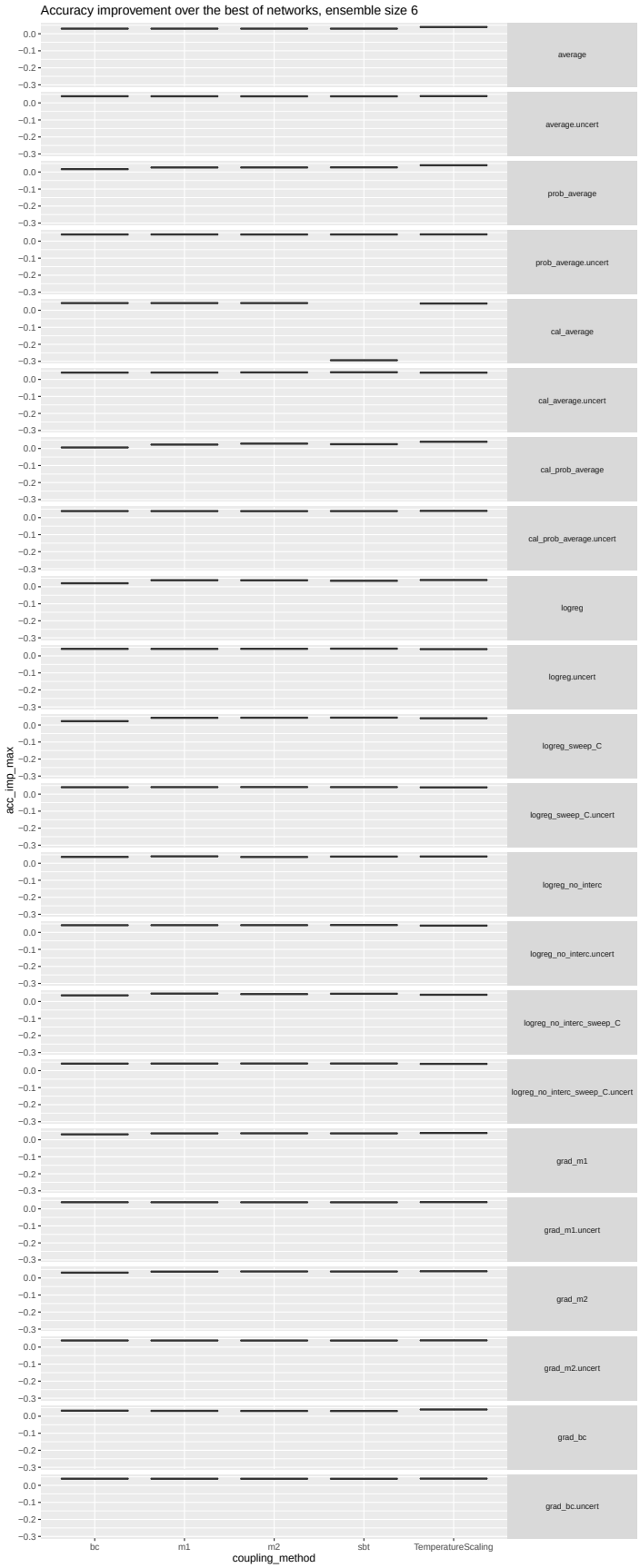

Accuracy improvement over the average of networks, ensemble size 5
acc_imp_avg
coupling_method
bc
m1
m2
sbt
TemperatureScaling
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert

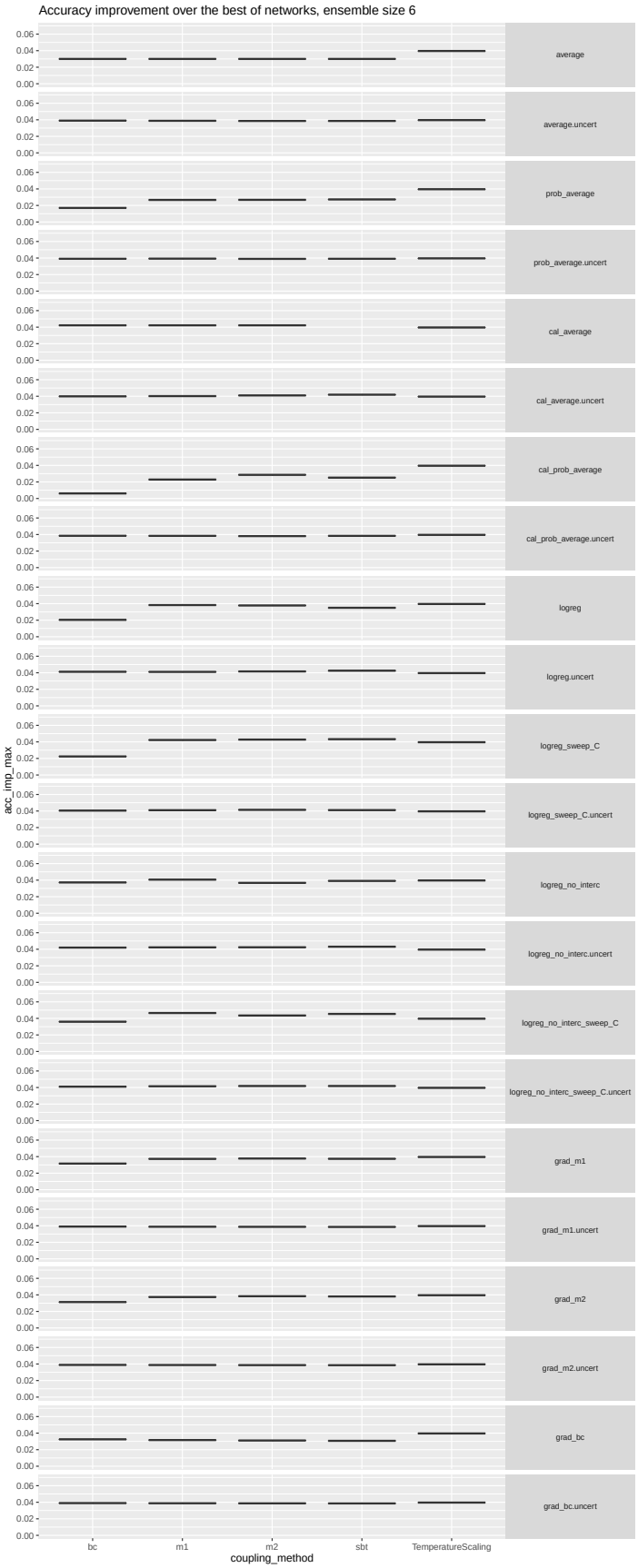

Accuracy improvement over the best of networks, ensemble size 5
acc_imp_max
coupling_method
bc
m1
m2
sbt
TemperatureScaling
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert


Accuracy improvement over the best of networks, ensemble size 5
acc_imp_max
coupling_method
bc
m1
m2
sbt
TemperatureScaling
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert


Accuracy improvement over the average of networks, ensemble size 6
acc_imp_avg
coupling_method
bc
m1
m2
sbt
TemperatureScaling
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert


Accuracy improvement over the average of networks, ensemble size 6
acc_imp_avg
0.100
0.075
0.050
0.025
0.000
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
bc
m1
m2
sbt
TemperatureScaling
coupling_method


Accuracy improvement over the best of networks, ensemble size 6
acc_imp_max
coupling_method
bc
m1
m2
sbt
TemperatureScaling
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert


Accuracy improvement over the best of networks, ensemble size 6
acc_imp_max
0.06
0.04
0.02
0.00
average
average.uncert
prob_average
prob_average.uncert
cal_average
cal_average.uncert
cal_prob_average
cal_prob_average.uncert
logreg
logreg.uncert
logreg_sweep_C
logreg_sweep_C.uncert
logreg_no_interc
logreg_no_interc.uncert
logreg_no_interc_sweep_C
logreg_no_interc_sweep_C.uncert
grad_m1
grad_m1.uncert
grad_m2
grad_m2.uncert
grad_bc
grad_bc.uncert
bc
m1
m2
sbt
TemperatureScaling
coupling_method

```
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
  print(xtable(avg_imp_table_cs %>% filter(combination_size == sss) %>% arrange(desc(imp_o_max)), digits
}
```

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Wed Feb 09 10:34:54 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	logreg_no_interc_sweep_C sbt	2	0.0472	0.0338
6	logreg_no_interc_sweep_C m1	2	0.0471	0.0338
7	prob_average.uncert m2	2	0.0469	0.0335
8	prob_average.uncert m1	2	0.0468	0.0334
9	cal_prob_average.uncert m2	2	0.0467	0.0333
10	prob_average.uncert bc	2	0.0467	0.0333
11	logreg_no_interc_sweep_C m2	2	0.0467	0.0333
12	prob_average.uncert sbt	2	0.0466	0.0333
13	logreg.uncert m2	2	0.0465	0.0332
14	cal_prob_average.uncert m1	2	0.0465	0.0331
15	logreg_no_interc.uncert m2	2	0.0464	0.0331
16	cal_prob_average.uncert bc	2	0.0464	0.0331
17	cal_prob_average.uncert sbt	2	0.0464	0.0330
18	logreg_sweep_C.uncert m2	2	0.0464	0.0330
19	logreg_no_interc.uncert sbt	2	0.0463	0.0330
20	logreg_no_interc_sweep_C.uncert m2	2	0.0463	0.0329
21	logreg_no_interc_sweep_C.uncert sbt	2	0.0462	0.0328
22	logreg.uncert sbt	2	0.0461	0.0328
23	logreg_sweep_C.uncert sbt	2	0.0461	0.0328
24	cal_average.uncert sbt	2	0.0461	0.0327
25	logreg_sweep_C m2	2	0.0461	0.0327
26	cal_average.uncert m2	2	0.0461	0.0327
27	logreg.uncert bc	2	0.0460	0.0326
28	logreg_sweep_C.uncert m1	2	0.0460	0.0326
29	logreg_no_interc.uncert m1	2	0.0460	0.0326
30	logreg.uncert m1	2	0.0460	0.0326
31	logreg_no_interc_sweep_C.uncert m1	2	0.0459	0.0326
32	logreg_sweep_C.uncert bc	2	0.0459	0.0325
33	cal_average.uncert m1	2	0.0459	0.0325
34	logreg_no_interc.uncert bc	2	0.0459	0.0325
35	logreg_no_interc_sweep_C.uncert bc	2	0.0458	0.0324
36	cal_average.uncert bc	2	0.0458	0.0324
37	logreg_sweep_C sbt	2	0.0455	0.0322
38	logreg_no_interc m1	2	0.0455	0.0322
39	grad_bc.uncert sbt	2	0.0455	0.0321
40	grad_m1.uncert sbt	2	0.0455	0.0321
41	average.uncert sbt	2	0.0455	0.0321
42	grad_bc.uncert bc	2	0.0455	0.0321
43	grad_m2.uncert sbt	2	0.0455	0.0321
44	grad_m2.uncert m1	2	0.0455	0.0321

45	average.uncert bc	2	0.0455	0.0321
46	grad_m1.uncert bc	2	0.0455	0.0321
47	grad_m2.uncert bc	2	0.0455	0.0321
48	grad_bc.uncert m1	2	0.0454	0.0321
49	grad_m1.uncert m1	2	0.0454	0.0321
50	grad_m2.uncert m2	2	0.0454	0.0321
51	average.uncert m1	2	0.0454	0.0321
52	average.uncert m2	2	0.0454	0.0321
53	grad_m2 m2	2	0.0454	0.0321
54	grad_bc.uncert m2	2	0.0454	0.0321
55	grad_m1.uncert m2	2	0.0454	0.0321
56	grad_m2 sbt	2	0.0453	0.0320
57	grad_m2 m1	2	0.0453	0.0319
58	logreg_no_interc sbt	2	0.0453	0.0319
59	average of TemperatureScaling	2	0.0449	0.0316
60	logreg_sweep_C m1	2	0.0447	0.0313
61	logreg m1	2	0.0445	0.0312
62	logreg_no_interc m2	2	0.0445	0.0311
63	grad_bc bc	2	0.0443	0.0309
64	logreg m2	2	0.0442	0.0309
65	grad_m2 bc	2	0.0438	0.0305
66	grad_m1 m1	2	0.0436	0.0302
67	grad_bc m2	2	0.0435	0.0302
68	grad_bc m1	2	0.0435	0.0302
69	grad_bc sbt	2	0.0435	0.0302
70	grad_m1 m2	2	0.0434	0.0301
71	grad_m1 bc	2	0.0434	0.0300
72	average m1	2	0.0433	0.0300
73	average m2	2	0.0433	0.0300
74	average sbt	2	0.0433	0.0300
75	average bc	2	0.0433	0.0299
76	logreg sbt	2	0.0429	0.0296
77	logreg_no_interc_sweep_C bc	2	0.0411	0.0278
78	cal_prob_average m2	2	0.0409	0.0275
79	logreg_no_interc bc	2	0.0403	0.0269
80	prob_average sbt	2	0.0397	0.0263
81	cal_prob_average sbt	2	0.0392	0.0259
82	prob_average m2	2	0.0389	0.0256
83	prob_average m1	2	0.0367	0.0233
84	cal_prob_average m1	2	0.0359	0.0226
85	prob_average bc	2	0.0310	0.0177
86	logreg_sweep_C bc	2	0.0276	0.0143
87	cal_prob_average bc	2	0.0244	0.0110
88	logreg bc	2	0.0242	0.0108
89	grad_m1 sbt	2	-0.0020	-0.0154

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Wed Feb 09 10:34:54 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	logreg__no_interc_sweep_C sbt	2	0.0472	0.0338
6	logreg__no_interc_sweep_C m1	2	0.0471	0.0338
7	prob_average.uncert m2	2	0.0469	0.0335
8	prob_average.uncert m1	2	0.0468	0.0334
9	cal_prob_average.uncert m2	2	0.0467	0.0333
10	prob_average.uncert bc	2	0.0467	0.0333
11	logreg__no_interc_sweep_C m2	2	0.0467	0.0333
12	prob_average.uncert sbt	2	0.0466	0.0333
13	logreg.uncert m2	2	0.0465	0.0332
14	cal_prob_average.uncert m1	2	0.0465	0.0331
15	logreg__no_interc.uncert m2	2	0.0464	0.0331
16	cal_prob_average.uncert bc	2	0.0464	0.0331
17	cal_prob_average.uncert sbt	2	0.0464	0.0330
18	logreg__sweep_C.uncert m2	2	0.0464	0.0330
19	logreg__no_interc.uncert sbt	2	0.0463	0.0330
20	logreg__no_interc_sweep_C.uncert m2	2	0.0463	0.0329
21	logreg__no_interc_sweep_C.uncert sbt	2	0.0462	0.0328
22	logreg.uncert sbt	2	0.0461	0.0328
23	logreg__sweep_C.uncert sbt	2	0.0461	0.0328
24	cal_average.uncert sbt	2	0.0461	0.0327
25	logreg__sweep_C m2	2	0.0461	0.0327
26	cal_average.uncert m2	2	0.0461	0.0327
27	logreg.uncert bc	2	0.0460	0.0326
28	logreg__sweep_C.uncert m1	2	0.0460	0.0326
29	logreg__no_interc.uncert m1	2	0.0460	0.0326
30	logreg.uncert m1	2	0.0460	0.0326
31	logreg__no_interc_sweep_C.uncert m1	2	0.0459	0.0326
32	logreg__sweep_C.uncert bc	2	0.0459	0.0325
33	cal_average.uncert m1	2	0.0459	0.0325
34	logreg__no_interc.uncert bc	2	0.0459	0.0325
35	logreg__no_interc_sweep_C.uncert bc	2	0.0458	0.0324
36	cal_average.uncert bc	2	0.0458	0.0324
37	logreg__sweep_C sbt	2	0.0455	0.0322
38	logreg__no_interc m1	2	0.0455	0.0322
39	grad_bc.uncert sbt	2	0.0455	0.0321
40	grad_m1.uncert sbt	2	0.0455	0.0321
41	average.uncert sbt	2	0.0455	0.0321
42	grad_bc.uncert bc	2	0.0455	0.0321
43	grad_m2.uncert sbt	2	0.0455	0.0321
44	grad_m2.uncert m1	2	0.0455	0.0321
45	average.uncert bc	2	0.0455	0.0321
46	grad_m1.uncert bc	2	0.0455	0.0321
47	grad_m2.uncert bc	2	0.0455	0.0321
48	grad_bc.uncert m1	2	0.0454	0.0321
49	grad_m1.uncert m1	2	0.0454	0.0321
50	grad_m2.uncert m2	2	0.0454	0.0321
51	average.uncert m1	2	0.0454	0.0321
52	average.uncert m2	2	0.0454	0.0321
53	grad_m2 m2	2	0.0454	0.0321
54	grad_bc.uncert m2	2	0.0454	0.0321
55	grad_m1.uncert m2	2	0.0454	0.0321
56	grad_m2 sbt	2	0.0453	0.0320
57	grad_m2 m1	2	0.0453	0.0319
58	logreg__no_interc sbt	2	0.0453	0.0319

59	average of TemperatureScaling	2	0.0449	0.0316
60	logreg_sweep_C m1	2	0.0447	0.0313
61	logreg m1	2	0.0445	0.0312
62	logreg_no_interc m2	2	0.0445	0.0311
63	grad_bc bc	2	0.0443	0.0309
64	logreg m2	2	0.0442	0.0309
65	grad_m2 bc	2	0.0438	0.0305
66	grad_m1 m1	2	0.0436	0.0302
67	grad_bc m2	2	0.0435	0.0302
68	grad_bc m1	2	0.0435	0.0302
69	grad_bc sbt	2	0.0435	0.0302
70	grad_m1 m2	2	0.0434	0.0301
71	grad_m1 bc	2	0.0434	0.0300
72	average m1	2	0.0433	0.0300
73	average m2	2	0.0433	0.0300
74	average sbt	2	0.0433	0.0300
75	average bc	2	0.0433	0.0299
76	logreg sbt	2	0.0429	0.0296
77	logreg_no_interc_sweep_C bc	2	0.0411	0.0278
78	cal_prob_average m2	2	0.0409	0.0275
79	logreg_no_interc bc	2	0.0403	0.0269
80	prob_average sbt	2	0.0397	0.0263
81	cal_prob_average sbt	2	0.0392	0.0259
82	prob_average m2	2	0.0389	0.0256
83	prob_average m1	2	0.0367	0.0233
84	cal_prob_average m1	2	0.0359	0.0226
85	prob_average bc	2	0.0310	0.0177
86	logreg_sweep_C bc	2	0.0276	0.0143
87	cal_prob_average bc	2	0.0244	0.0110
88	logreg bc	2	0.0242	0.0108
89	grad_m1 sbt	2	-0.0020	-0.0154

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Wed Feb 09 10:34:54 2022

	method	combination_size	imp_o_avg	imp_o_max
1	logreg_no_interc_sweep_C sbt	3	0.0656	0.0426
2	logreg_no_interc_sweep_C m1	3	0.0654	0.0425
3	cal_average m1	3	0.0653	0.0424
4	cal_average m2	3	0.0653	0.0424
5	cal_average sbt	3	0.0653	0.0424
6	cal_average bc	3	0.0653	0.0424
7	logreg_no_interc_sweep_C m2	3	0.0648	0.0419
8	logreg_sweep_C m2	3	0.0643	0.0414
9	logreg_no_interc.uncert sbt	3	0.0640	0.0411
10	logreg.uncert sbt	3	0.0638	0.0409
11	logreg_no_interc.uncert m2	3	0.0637	0.0408
12	logreg_sweep_C sbt	3	0.0637	0.0408
13	logreg.uncert m2	3	0.0635	0.0406
14	cal_average.uncert sbt	3	0.0635	0.0405
15	logreg_no_interc_sweep_C.uncert sbt	3	0.0632	0.0403
16	logreg_no_interc_sweep_C.uncert m2	3	0.0632	0.0402
17	logreg_no_interc m1	3	0.0630	0.0401
18	logreg_no_interc.uncert m1	3	0.0630	0.0401

19	logreg_sweep_C.uncert sbt	3	0.0630	0.0400
20	logreg_sweep_C.uncert m2	3	0.0630	0.0400
21	cal_average.uncert m2	3	0.0629	0.0400
22	logreg_no_interc.uncert bc	3	0.0628	0.0398
23	logreg.uncert m1	3	0.0627	0.0398
24	logreg.uncert bc	3	0.0627	0.0397
25	logreg_no_interc_sweep_C.uncert m1	3	0.0627	0.0397
26	logreg_sweep_C.uncert m1	3	0.0624	0.0395
27	cal_average.uncert m1	3	0.0624	0.0395
28	logreg_no_interc_sweep_C.uncert bc	3	0.0624	0.0395
29	logreg_sweep_C.uncert bc	3	0.0624	0.0394
30	prob_average.uncert bc	3	0.0623	0.0394
31	prob_average.uncert m1	3	0.0623	0.0394
32	logreg_sweep_C m1	3	0.0622	0.0393
33	prob_average.uncert sbt	3	0.0622	0.0393
34	cal_average.uncert bc	3	0.0621	0.0392
35	prob_average.uncert m2	3	0.0621	0.0392
36	logreg_no_interc sbt	3	0.0621	0.0392
37	cal_prob_average.uncert bc	3	0.0620	0.0391
38	logreg m1	3	0.0620	0.0391
39	cal_prob_average.uncert m2	3	0.0620	0.0390
40	cal_prob_average.uncert m1	3	0.0619	0.0390
41	cal_prob_average.uncert sbt	3	0.0619	0.0390
42	grad_m2 m2	3	0.0615	0.0386
43	grad_m1.uncert sbt	3	0.0615	0.0386
44	grad_m2.uncert sbt	3	0.0615	0.0385
45	average.uncert sbt	3	0.0615	0.0385
46	grad_bc.uncert sbt	3	0.0615	0.0385
47	average.uncert bc	3	0.0614	0.0385
48	grad_m1.uncert bc	3	0.0614	0.0385
49	grad_m2.uncert bc	3	0.0614	0.0385
50	grad_bc.uncert bc	3	0.0614	0.0385
51	grad_m2.uncert m2	3	0.0614	0.0385
52	grad_bc.uncert m1	3	0.0614	0.0385
53	grad_m1.uncert m1	3	0.0614	0.0385
54	grad_m2.uncert m1	3	0.0614	0.0385
55	average.uncert m2	3	0.0614	0.0385
56	grad_m2 sbt	3	0.0614	0.0385
57	average.uncert m1	3	0.0614	0.0385
58	grad_bc.uncert m2	3	0.0614	0.0385
59	grad_m1.uncert m2	3	0.0614	0.0385
60	grad_m1 m2	3	0.0613	0.0384
61	logreg m2	3	0.0613	0.0384
62	grad_m1 m1	3	0.0613	0.0384
63	average of TemperatureScaling	3	0.0612	0.0383
64	grad_m2 m1	3	0.0612	0.0383
65	logreg_no_interc m2	3	0.0608	0.0378
66	logreg sbt	3	0.0600	0.0371
67	grad_m1 sbt	3	0.0589	0.0359
68	logreg_no_interc_sweep_C bc	3	0.0585	0.0355
69	grad_bc bc	3	0.0584	0.0354
70	grad_m2 bc	3	0.0580	0.0350
71	grad_bc m2	3	0.0580	0.0350
72	grad_m1 bc	3	0.0579	0.0350

73	grad_bc m1	3	0.0579	0.0350
74	grad_bc sbt	3	0.0579	0.0350
75	logreg_no_interc bc	3	0.0577	0.0348
76	average m1	3	0.0574	0.0344
77	average m2	3	0.0574	0.0344
78	average sbt	3	0.0574	0.0344
79	average bc	3	0.0574	0.0344
80	cal_prob_average m2	3	0.0554	0.0325
81	prob_average m2	3	0.0537	0.0307
82	prob_average sbt	3	0.0535	0.0306
83	cal_prob_average sbt	3	0.0520	0.0290
84	prob_average m1	3	0.0515	0.0285
85	cal_prob_average m1	3	0.0493	0.0264
86	logreg_sweep_C bc	3	0.0452	0.0222
87	prob_average bc	3	0.0433	0.0203
88	logreg bc	3	0.0411	0.0182
89	cal_prob_average bc	3	0.0334	0.0104

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Wed Feb 09 10:34:54 2022

	method	combination_size	imp_o_avg	imp_o_max
1	logreg_no_interc_sweep_C sbt	3	0.0656	0.0426
2	logreg_no_interc_sweep_C m1	3	0.0654	0.0425
3	cal_average m1	3	0.0653	0.0424
4	cal_average m2	3	0.0653	0.0424
5	cal_average sbt	3	0.0653	0.0424
6	cal_average bc	3	0.0653	0.0424
7	logreg_no_interc_sweep_C m2	3	0.0648	0.0419
8	logreg_sweep_C m2	3	0.0643	0.0414
9	logreg_no_interc.uncert sbt	3	0.0640	0.0411
10	logreg.uncert sbt	3	0.0638	0.0409
11	logreg_no_interc.uncert m2	3	0.0637	0.0408
12	logreg_sweep_C sbt	3	0.0637	0.0408
13	logreg.uncert m2	3	0.0635	0.0406
14	cal_average.uncert sbt	3	0.0635	0.0405
15	logreg_no_interc_sweep_C.uncert sbt	3	0.0632	0.0403
16	logreg_no_interc_sweep_C.uncert m2	3	0.0632	0.0402
17	logreg_no_interc m1	3	0.0630	0.0401
18	logreg_no_interc.uncert m1	3	0.0630	0.0401
19	logreg_sweep_C.uncert sbt	3	0.0630	0.0400
20	logreg_sweep_C.uncert m2	3	0.0630	0.0400
21	cal_average.uncert m2	3	0.0629	0.0400
22	logreg_no_interc.uncert bc	3	0.0628	0.0398
23	logreg.uncert m1	3	0.0627	0.0398
24	logreg.uncert bc	3	0.0627	0.0397
25	logreg_no_interc_sweep_C.uncert m1	3	0.0627	0.0397
26	logreg_sweep_C.uncert m1	3	0.0624	0.0395
27	cal_average.uncert m1	3	0.0624	0.0395
28	logreg_no_interc_sweep_C.uncert bc	3	0.0624	0.0395
29	logreg_sweep_C.uncert bc	3	0.0624	0.0394
30	prob_average.uncert bc	3	0.0623	0.0394
31	prob_average.uncert m1	3	0.0623	0.0394
32	logreg_sweep_C m1	3	0.0622	0.0393

33	prob_average.uncert sbt	3	0.0622	0.0393
34	cal_average.uncert bc	3	0.0621	0.0392
35	prob_average.uncert m2	3	0.0621	0.0392
36	logreg_no_interc sbt	3	0.0621	0.0392
37	cal_prob_average.uncert bc	3	0.0620	0.0391
38	logreg m1	3	0.0620	0.0391
39	cal_prob_average.uncert m2	3	0.0620	0.0390
40	cal_prob_average.uncert m1	3	0.0619	0.0390
41	cal_prob_average.uncert sbt	3	0.0619	0.0390
42	grad_m2 m2	3	0.0615	0.0386
43	grad_m1.uncert sbt	3	0.0615	0.0386
44	grad_m2.uncert sbt	3	0.0615	0.0385
45	average.uncert sbt	3	0.0615	0.0385
46	grad_bc.uncert sbt	3	0.0615	0.0385
47	average.uncert bc	3	0.0614	0.0385
48	grad_m1.uncert bc	3	0.0614	0.0385
49	grad_m2.uncert bc	3	0.0614	0.0385
50	grad_bc.uncert bc	3	0.0614	0.0385
51	grad_m2.uncert m2	3	0.0614	0.0385
52	grad_bc.uncert m1	3	0.0614	0.0385
53	grad_m1.uncert m1	3	0.0614	0.0385
54	grad_m2.uncert m1	3	0.0614	0.0385
55	average.uncert m2	3	0.0614	0.0385
56	grad_m2 sbt	3	0.0614	0.0385
57	average.uncert m1	3	0.0614	0.0385
58	grad_bc.uncert m2	3	0.0614	0.0385
59	grad_m1.uncert m2	3	0.0614	0.0385
60	grad_m1 m2	3	0.0613	0.0384
61	logreg m2	3	0.0613	0.0384
62	grad_m1 m1	3	0.0613	0.0384
63	average of TemperatureScaling	3	0.0612	0.0383
64	grad_m2 m1	3	0.0612	0.0383
65	logreg_no_interc m2	3	0.0608	0.0378
66	logreg sbt	3	0.0600	0.0371
67	grad_m1 sbt	3	0.0589	0.0359
68	logreg_no_interc_sweep_C bc	3	0.0585	0.0355
69	grad_bc bc	3	0.0584	0.0354
70	grad_m2 bc	3	0.0580	0.0350
71	grad_bc m2	3	0.0580	0.0350
72	grad_m1 bc	3	0.0579	0.0350
73	grad_bc m1	3	0.0579	0.0350
74	grad_bc sbt	3	0.0579	0.0350
75	logreg_no_interc bc	3	0.0577	0.0348
76	average m1	3	0.0574	0.0344
77	average m2	3	0.0574	0.0344
78	average sbt	3	0.0574	0.0344
79	average bc	3	0.0574	0.0344
80	cal_prob_average m2	3	0.0554	0.0325
81	prob_average m2	3	0.0537	0.0307
82	prob_average sbt	3	0.0535	0.0306
83	cal_prob_average sbt	3	0.0520	0.0290
84	prob_average m1	3	0.0515	0.0285
85	cal_prob_average m1	3	0.0493	0.0264
86	logreg_sweep_C bc	3	0.0452	0.0222

87	prob_average bc	3	0.0433	0.0203
88	logreg bc	3	0.0411	0.0182
89	cal_prob_average bc	3	0.0334	0.0104

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Wed Feb 09 10:34:54 2022

	method	combination_size	imp_o_avg	imp_o_max
1	logreg_no_interc_sweep_C sbt	4	0.0753	0.0452
2	logreg_no_interc_sweep_C m1	4	0.0749	0.0447
3	logreg_no_interc_sweep_C m2	4	0.0742	0.0440
4	logreg_sweep_C m2	4	0.0740	0.0439
5	logreg_sweep_C sbt	4	0.0737	0.0436
6	cal_average m1	4	0.0734	0.0432
7	cal_average m2	4	0.0734	0.0432
8	cal_average bc	4	0.0733	0.0432
9	logreg_no_interc.uncert sbt	4	0.0728	0.0427
10	logreg.uncert sbt	4	0.0727	0.0426
11	logreg_no_interc.uncert m2	4	0.0726	0.0425
12	logreg.uncert m2	4	0.0724	0.0423
13	logreg_no_interc m1	4	0.0723	0.0422
14	cal_average.uncert sbt	4	0.0722	0.0420
15	logreg_no_interc_sweep_C.uncert sbt	4	0.0721	0.0420
16	logreg_no_interc_sweep_C.uncert m2	4	0.0721	0.0419
17	logreg_no_interc.uncert m1	4	0.0720	0.0419
18	logreg_sweep_C.uncert sbt	4	0.0719	0.0418
19	logreg_sweep_C.uncert m2	4	0.0719	0.0418
20	logreg_no_interc.uncert bc	4	0.0718	0.0417
21	logreg_no_interc_sweep_C.uncert m1	4	0.0718	0.0417
22	cal_average.uncert m2	4	0.0718	0.0417
23	logreg.uncert m1	4	0.0717	0.0416
24	logreg_sweep_C m1	4	0.0717	0.0416
25	logreg_no_interc_sweep_C.uncert bc	4	0.0717	0.0415
26	logreg.uncert bc	4	0.0716	0.0415
27	logreg_sweep_C.uncert m1	4	0.0715	0.0414
28	logreg_no_interc sbt	4	0.0714	0.0413
29	logreg_sweep_C.uncert bc	4	0.0714	0.0413
30	cal_average.uncert m1	4	0.0714	0.0412
31	cal_average.uncert bc	4	0.0712	0.0411
32	prob_average.uncert m2	4	0.0711	0.0409
33	prob_average.uncert bc	4	0.0710	0.0409
34	prob_average.uncert m1	4	0.0710	0.0409
35	logreg m1	4	0.0710	0.0409
36	prob_average.uncert sbt	4	0.0709	0.0408
37	cal_prob_average.uncert m2	4	0.0707	0.0406
38	cal_prob_average.uncert m1	4	0.0707	0.0405
39	cal_prob_average.uncert bc	4	0.0707	0.0405
40	cal_prob_average.uncert sbt	4	0.0706	0.0405
41	cal_average sbt	4	0.0704	0.0403
42	logreg m2	4	0.0703	0.0402
43	average of TemperatureScaling	4	0.0702	0.0401
44	grad_bc.uncert bc	4	0.0702	0.0401
45	grad_m1.uncert bc	4	0.0702	0.0400
46	grad_m2.uncert bc	4	0.0702	0.0400

47	grad_m2.uncert m1	4	0.0702	0.0400
48	average.uncert bc	4	0.0702	0.0400
49	grad_m1.uncert m1	4	0.0701	0.0400
50	average.uncert m1	4	0.0701	0.0400
51	grad_bc.uncert m1	4	0.0701	0.0400
52	grad_m2.uncert m2	4	0.0701	0.0400
53	grad_m2.uncert sbt	4	0.0701	0.0400
54	grad_bc.uncert sbt	4	0.0701	0.0400
55	average.uncert m2	4	0.0701	0.0400
56	grad_m1.uncert m2	4	0.0701	0.0400
57	average.uncert sbt	4	0.0701	0.0400
58	grad_m1.uncert sbt	4	0.0701	0.0400
59	grad_bc.uncert m2	4	0.0701	0.0400
60	grad_m2 m2	4	0.0697	0.0396
61	grad_m2 sbt	4	0.0696	0.0395
62	logreg_no_interc m2	4	0.0694	0.0392
63	grad_m2 m1	4	0.0693	0.0392
64	grad_m1 m2	4	0.0691	0.0390
65	logreg sbt	4	0.0691	0.0389
66	grad_m1 m1	4	0.0689	0.0388
67	grad_m1 sbt	4	0.0686	0.0385
68	logreg_no_interc bc	4	0.0676	0.0375
69	logreg_no_interc_sweep_C bc	4	0.0671	0.0370
70	grad_m2 bc	4	0.0647	0.0346
71	grad_bc m2	4	0.0645	0.0344
72	grad_bc bc	4	0.0645	0.0344
73	grad_bc m1	4	0.0645	0.0344
74	grad_bc sbt	4	0.0645	0.0344
75	grad_m1 bc	4	0.0645	0.0343
76	average bc	4	0.0638	0.0336
77	average m1	4	0.0638	0.0336
78	average m2	4	0.0638	0.0336
79	average sbt	4	0.0638	0.0336
80	cal_prob_average m2	4	0.0632	0.0330
81	prob_average m2	4	0.0608	0.0307
82	prob_average sbt	4	0.0606	0.0305
83	prob_average m1	4	0.0588	0.0287
84	cal_prob_average sbt	4	0.0584	0.0283
85	cal_prob_average m1	4	0.0561	0.0260
86	logreg_sweep_C bc	4	0.0532	0.0231
87	logreg bc	4	0.0498	0.0196
88	prob_average bc	4	0.0497	0.0195
89	cal_prob_average bc	4	0.0386	0.0085

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Wed Feb 09 10:34:54 2022

	method	combination_size	imp_o_avg	imp_o_max
1	logreg_no_interc_sweep_C sbt	4	0.0753	0.0452
2	logreg_no_interc_sweep_C m1	4	0.0749	0.0447
3	logreg_no_interc_sweep_C m2	4	0.0742	0.0440
4	logreg_sweep_C m2	4	0.0740	0.0439
5	logreg_sweep_C sbt	4	0.0737	0.0436
6	cal_average m1	4	0.0734	0.0432

7	cal_average m2	4	0.0734	0.0432
8	cal_average bc	4	0.0733	0.0432
9	logreg_no_interc.uncert sbt	4	0.0728	0.0427
10	logreg.uncert sbt	4	0.0727	0.0426
11	logreg_no_interc.uncert m2	4	0.0726	0.0425
12	logreg.uncert m2	4	0.0724	0.0423
13	logreg_no_interc m1	4	0.0723	0.0422
14	cal_average.uncert sbt	4	0.0722	0.0420
15	logreg_no_interc_sweep_C.uncert sbt	4	0.0721	0.0420
16	logreg_no_interc_sweep_C.uncert m2	4	0.0721	0.0419
17	logreg_no_interc.uncert m1	4	0.0720	0.0419
18	logreg_sweep_C.uncert sbt	4	0.0719	0.0418
19	logreg_sweep_C.uncert m2	4	0.0719	0.0418
20	logreg_no_interc_sweep_C.uncert m1	4	0.0718	0.0417
21	logreg_no_interc.uncert bc	4	0.0718	0.0417
22	cal_average.uncert m2	4	0.0718	0.0417
23	logreg.uncert m1	4	0.0717	0.0416
24	logreg_sweep_C m1	4	0.0717	0.0416
25	logreg_no_interc_sweep_C.uncert bc	4	0.0717	0.0415
26	logreg.uncert bc	4	0.0716	0.0415
27	logreg_sweep_C.uncert m1	4	0.0715	0.0414
28	logreg_no_interc sbt	4	0.0714	0.0413
29	logreg_sweep_C.uncert bc	4	0.0714	0.0413
30	cal_average.uncert m1	4	0.0714	0.0412
31	cal_average.uncert bc	4	0.0712	0.0411
32	prob_average.uncert m2	4	0.0711	0.0409
33	prob_average.uncert bc	4	0.0710	0.0409
34	prob_average.uncert m1	4	0.0710	0.0409
35	logreg m1	4	0.0710	0.0409
36	prob_average.uncert sbt	4	0.0709	0.0408
37	cal_prob_average.uncert m2	4	0.0707	0.0406
38	cal_prob_average.uncert m1	4	0.0707	0.0405
39	cal_prob_average.uncert bc	4	0.0707	0.0405
40	cal_prob_average.uncert sbt	4	0.0706	0.0405
41	cal_average sbt	4	0.0704	0.0403
42	logreg m2	4	0.0703	0.0402
43	average of TemperatureScaling	4	0.0702	0.0401
44	grad_bc.uncert bc	4	0.0702	0.0401
45	grad_m1.uncert bc	4	0.0702	0.0400
46	grad_m2.uncert bc	4	0.0702	0.0400
47	grad_m2.uncert m1	4	0.0702	0.0400
48	average.uncert bc	4	0.0702	0.0400
49	grad_m1.uncert m1	4	0.0701	0.0400
50	average.uncert m1	4	0.0701	0.0400
51	grad_bc.uncert m1	4	0.0701	0.0400
52	grad_m2.uncert m2	4	0.0701	0.0400
53	grad_m2.uncert sbt	4	0.0701	0.0400
54	grad_bc.uncert sbt	4	0.0701	0.0400
55	average.uncert m2	4	0.0701	0.0400
56	grad_m1.uncert m2	4	0.0701	0.0400
57	average.uncert sbt	4	0.0701	0.0400
58	grad_m1.uncert sbt	4	0.0701	0.0400
59	grad_bc.uncert m2	4	0.0701	0.0400
60	grad_m2 m2	4	0.0697	0.0396

61	grad_m2 sbt	4	0.0696	0.0395
62	logreg_no_interc m2	4	0.0694	0.0392
63	grad_m2 m1	4	0.0693	0.0392
64	grad_m1 m2	4	0.0691	0.0390
65	logreg sbt	4	0.0691	0.0389
66	grad_m1 m1	4	0.0689	0.0388
67	grad_m1 sbt	4	0.0686	0.0385
68	logreg_no_interc bc	4	0.0676	0.0375
69	logreg_no_interc_sweep_C bc	4	0.0671	0.0370
70	grad_m2 bc	4	0.0647	0.0346
71	grad_bc m2	4	0.0645	0.0344
72	grad_bc bc	4	0.0645	0.0344
73	grad_bc m1	4	0.0645	0.0344
74	grad_bc sbt	4	0.0645	0.0344
75	grad_m1 bc	4	0.0645	0.0343
76	average bc	4	0.0638	0.0336
77	average m1	4	0.0638	0.0336
78	average m2	4	0.0638	0.0336
79	average sbt	4	0.0638	0.0336
80	cal_prob_average m2	4	0.0632	0.0330
81	prob_average m2	4	0.0608	0.0307
82	prob_average sbt	4	0.0606	0.0305
83	prob_average m1	4	0.0588	0.0287
84	cal_prob_average sbt	4	0.0584	0.0283
85	cal_prob_average m1	4	0.0561	0.0260
86	logreg_sweep_C bc	4	0.0532	0.0231
87	logreg bc	4	0.0498	0.0196
88	prob_average bc	4	0.0497	0.0195
89	cal_prob_average bc	4	0.0386	0.0085

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Wed Feb 09 10:34:54 2022

	method	combination_size	imp_o_avg	imp_o_max
1	logreg_no_interc_sweep_C sbt	5	0.0811	0.0454
2	logreg_no_interc_sweep_C m1	5	0.0808	0.0451
3	logreg_no_interc_sweep_C m2	5	0.0798	0.0441
4	logreg_sweep_C m2	5	0.0794	0.0437
5	logreg_sweep_C sbt	5	0.0792	0.0435
6	logreg_no_interc.uncert sbt	5	0.0783	0.0426
7	cal_average bc	5	0.0782	0.0425
8	cal_average m1	5	0.0781	0.0424
9	cal_average m2	5	0.0781	0.0424
10	logreg_no_interc m1	5	0.0780	0.0423
11	logreg_no_interc.uncert m2	5	0.0780	0.0423
12	logreg_sweep_C m1	5	0.0777	0.0420
13	logreg.uncert sbt	5	0.0775	0.0418
14	logreg_no_interc sbt	5	0.0774	0.0417
15	logreg.uncert m2	5	0.0774	0.0417
16	logreg_no_interc_sweep_C.uncert sbt	5	0.0772	0.0415
17	logreg_no_interc.uncert m1	5	0.0772	0.0415
18	logreg_sweep_C.uncert sbt	5	0.0771	0.0414
19	cal_average.uncert sbt	5	0.0771	0.0414
20	logreg_no_interc_sweep_C.uncert m2	5	0.0769	0.0412

21	logreg.uncert bc	5	0.0769	0.0412
22	logreg_no_interc.uncert bc	5	0.0769	0.0412
23	logreg.uncert m1	5	0.0768	0.0411
24	logreg_sweep_C.uncert m2	5	0.0768	0.0411
25	cal_average.uncert m2	5	0.0768	0.0411
26	logreg_no_interc_sweep_C.uncert m1	5	0.0766	0.0409
27	logreg_sweep_C.uncert m1	5	0.0766	0.0409
28	logreg_no_interc_sweep_C.uncert bc	5	0.0766	0.0408
29	logreg_sweep_C.uncert bc	5	0.0765	0.0408
30	cal_average.uncert m1	5	0.0765	0.0407
31	cal_average.uncert bc	5	0.0763	0.0406
32	prob_average.uncert bc	5	0.0761	0.0403
33	prob_average.uncert m1	5	0.0760	0.0403
34	prob_average.uncert m2	5	0.0759	0.0402
35	prob_average.uncert sbt	5	0.0759	0.0402
36	average of TemperatureScaling	5	0.0756	0.0399
37	cal_prob_average.uncert m2	5	0.0755	0.0398
38	cal_prob_average.uncert bc	5	0.0755	0.0398
39	cal_prob_average.uncert m1	5	0.0755	0.0398
40	logreg m2	5	0.0755	0.0398
41	cal_prob_average.uncert sbt	5	0.0754	0.0397
42	grad_m1.uncert bc	5	0.0751	0.0394
43	grad_bc.uncert bc	5	0.0751	0.0393
44	average.uncert bc	5	0.0750	0.0393
45	grad_m2.uncert bc	5	0.0750	0.0393
46	average.uncert m1	5	0.0750	0.0393
47	grad_bc.uncert m1	5	0.0750	0.0393
48	grad_m1.uncert m1	5	0.0750	0.0393
49	grad_m2.uncert m1	5	0.0750	0.0393
50	grad_m1.uncert sbt	5	0.0749	0.0392
51	grad_m2.uncert sbt	5	0.0749	0.0392
52	grad_bc.uncert sbt	5	0.0749	0.0392
53	logreg_no_interc m2	5	0.0749	0.0392
54	logreg m1	5	0.0749	0.0392
55	average.uncert m2	5	0.0749	0.0392
56	grad_m2.uncert m2	5	0.0749	0.0392
57	average.uncert sbt	5	0.0749	0.0392
58	grad_bc.uncert m2	5	0.0749	0.0392
59	grad_m1.uncert m2	5	0.0749	0.0392
60	grad_m2 m2	5	0.0749	0.0391
61	grad_m2 sbt	5	0.0745	0.0388
62	grad_m1 sbt	5	0.0744	0.0387
63	grad_m1 m1	5	0.0744	0.0387
64	grad_m1 m2	5	0.0743	0.0386
65	grad_m2 m1	5	0.0742	0.0385
66	logreg sbt	5	0.0736	0.0379
67	logreg_no_interc bc	5	0.0732	0.0375
68	logreg_no_interc_sweep_C bc	5	0.0727	0.0370
69	grad_m2 bc	5	0.0691	0.0334
70	grad_bc m1	5	0.0690	0.0333
71	grad_bc m2	5	0.0689	0.0332
72	grad_bc bc	5	0.0689	0.0332
73	grad_bc sbt	5	0.0688	0.0331
74	grad_m1 bc	5	0.0685	0.0328

75	average m1	5	0.0678	0.0321
76	average m2	5	0.0678	0.0321
77	average sbt	5	0.0678	0.0321
78	average bc	5	0.0678	0.0320
79	cal_prob_average m2	5	0.0669	0.0312
80	prob_average sbt	5	0.0648	0.0291
81	prob_average m2	5	0.0645	0.0288
82	prob_average m1	5	0.0633	0.0276
83	cal_prob_average sbt	5	0.0626	0.0269
84	cal_prob_average m1	5	0.0605	0.0248
85	logreg_sweep_C bc	5	0.0582	0.0224
86	logreg bc	5	0.0557	0.0200
87	prob_average bc	5	0.0538	0.0181
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 % Wed Feb 09 10:34:54 2022

	method	combination_size	imp_o_avg	imp_o_max
1	logreg_no_interc_sweep_C sbt	5	0.0811	0.0454
2	logreg_no_interc_sweep_C m1	5	0.0808	0.0451
3	logreg_no_interc_sweep_C m2	5	0.0798	0.0441
4	logreg_sweep_C m2	5	0.0794	0.0437
5	logreg_sweep_C sbt	5	0.0792	0.0435
6	logreg_no_interc.uncert sbt	5	0.0783	0.0426
7	cal_average bc	5	0.0782	0.0425
8	cal_average m1	5	0.0781	0.0424
9	cal_average m2	5	0.0781	0.0424
10	logreg_no_interc m1	5	0.0780	0.0423
11	logreg_no_interc.uncert m2	5	0.0780	0.0423
12	logreg_sweep_C m1	5	0.0777	0.0420
13	logreg.uncert sbt	5	0.0775	0.0418
14	logreg_no_interc sbt	5	0.0774	0.0417
15	logreg.uncert m2	5	0.0774	0.0417
16	logreg_no_interc_sweep_C.uncert sbt	5	0.0772	0.0415
17	logreg_no_interc.uncert m1	5	0.0772	0.0415
18	logreg_sweep_C.uncert sbt	5	0.0771	0.0414
19	cal_average.uncert sbt	5	0.0771	0.0414
20	logreg_no_interc_sweep_C.uncert m2	5	0.0769	0.0412
21	logreg.uncert bc	5	0.0769	0.0412
22	logreg_no_interc.uncert bc	5	0.0769	0.0412
23	logreg.uncert m1	5	0.0768	0.0411
24	logreg_sweep_C.uncert m2	5	0.0768	0.0411
25	cal_average.uncert m2	5	0.0768	0.0411
26	logreg_no_interc_sweep_C.uncert m1	5	0.0766	0.0409
27	logreg_sweep_C.uncert m1	5	0.0766	0.0409
28	logreg_no_interc_sweep_C.uncert bc	5	0.0766	0.0408
29	logreg_sweep_C.uncert bc	5	0.0765	0.0408
30	cal_average.uncert m1	5	0.0765	0.0407
31	cal_average.uncert bc	5	0.0763	0.0406
32	prob_average.uncert bc	5	0.0761	0.0403
33	prob_average.uncert m1	5	0.0760	0.0403
34	prob_average.uncert m2	5	0.0759	0.0402

35	prob_average.uncert sbt	5	0.0759	0.0402
36	average of TemperatureScaling	5	0.0756	0.0399
37	cal_prob_average.uncert m2	5	0.0755	0.0398
38	cal_prob_average.uncert bc	5	0.0755	0.0398
39	cal_prob_average.uncert m1	5	0.0755	0.0398
40	logreg m2	5	0.0755	0.0398
41	cal_prob_average.uncert sbt	5	0.0754	0.0397
42	grad_m1.uncert bc	5	0.0751	0.0394
43	grad_bc.uncert bc	5	0.0751	0.0393
44	average.uncert bc	5	0.0750	0.0393
45	grad_m2.uncert bc	5	0.0750	0.0393
46	average.uncert m1	5	0.0750	0.0393
47	grad_bc.uncert m1	5	0.0750	0.0393
48	grad_m1.uncert m1	5	0.0750	0.0393
49	grad_m2.uncert m1	5	0.0750	0.0393
50	grad_m1.uncert sbt	5	0.0749	0.0392
51	grad_m2.uncert sbt	5	0.0749	0.0392
52	grad_bc.uncert sbt	5	0.0749	0.0392
53	logreg_no_interc m2	5	0.0749	0.0392
54	logreg m1	5	0.0749	0.0392
55	average.uncert m2	5	0.0749	0.0392
56	grad_m2.uncert m2	5	0.0749	0.0392
57	average.uncert sbt	5	0.0749	0.0392
58	grad_bc.uncert m2	5	0.0749	0.0392
59	grad_m1.uncert m2	5	0.0749	0.0392
60	grad_m2 m2	5	0.0749	0.0391
61	grad_m2 sbt	5	0.0745	0.0388
62	grad_m1 sbt	5	0.0744	0.0387
63	grad_m1 m1	5	0.0744	0.0387
64	grad_m1 m2	5	0.0743	0.0386
65	grad_m2 m1	5	0.0742	0.0385
66	logreg sbt	5	0.0736	0.0379
67	logreg_no_interc bc	5	0.0732	0.0375
68	logreg_no_interc_sweep_C bc	5	0.0727	0.0370
69	grad_m2 bc	5	0.0691	0.0334
70	grad_bc m1	5	0.0690	0.0333
71	grad_bc m2	5	0.0689	0.0332
72	grad_bc bc	5	0.0689	0.0332
73	grad_bc sbt	5	0.0688	0.0331
74	grad_m1 bc	5	0.0685	0.0328
75	average m1	5	0.0678	0.0321
76	average m2	5	0.0678	0.0321
77	average sbt	5	0.0678	0.0321
78	average bc	5	0.0678	0.0320
79	cal_prob_average m2	5	0.0669	0.0312
80	prob_average sbt	5	0.0648	0.0291
81	prob_average m2	5	0.0645	0.0288
82	prob_average m1	5	0.0633	0.0276
83	cal_prob_average sbt	5	0.0626	0.0269
84	cal_prob_average m1	5	0.0605	0.0248
85	logreg_sweep_C bc	5	0.0582	0.0224
86	logreg bc	5	0.0557	0.0200
87	prob_average bc	5	0.0538	0.0181
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 % Wed Feb 09 10:34:54 2022

	method	combination_size	imp_o_avg	imp_o_max
1	logreg_no_interc_sweep_C m1	6	0.0862	0.0465
2	logreg_no_interc_sweep_C sbt	6	0.0851	0.0454
3	logreg_no_interc_sweep_C m2	6	0.0832	0.0435
4	logreg_sweep_C sbt	6	0.0830	0.0433
5	logreg_no_interc.uncert sbt	6	0.0828	0.0431
6	logreg_sweep_C m2	6	0.0825	0.0428
7	logreg.uncert sbt	6	0.0822	0.0425
8	logreg_no_interc.uncert m2	6	0.0821	0.0424
9	logreg_no_interc.uncert m1	6	0.0820	0.0423
10	logreg_sweep_C m1	6	0.0820	0.0423
11	cal_average bc	6	0.0819	0.0422
12	cal_average m1	6	0.0819	0.0422
13	cal_average m2	6	0.0819	0.0422
14	logreg_no_interc.uncert bc	6	0.0817	0.0420
15	cal_average.uncert sbt	6	0.0816	0.0419
16	logreg_no_interc_sweep_C.uncert m2	6	0.0815	0.0418
17	logreg_no_interc_sweep_C.uncert sbt	6	0.0815	0.0418
18	logreg.uncert m2	6	0.0813	0.0416
19	logreg_no_interc_sweep_C.uncert m1	6	0.0812	0.0415
20	logreg_sweep_C.uncert m2	6	0.0811	0.0414
21	logreg.uncert bc	6	0.0809	0.0412
22	logreg.uncert m1	6	0.0808	0.0411
23	logreg_sweep_C.uncert sbt	6	0.0808	0.0411
24	cal_average.uncert m2	6	0.0807	0.0410
25	logreg_no_interc_sweep_C.uncert bc	6	0.0807	0.0410
26	logreg_sweep_C.uncert m1	6	0.0807	0.0410
27	logreg_no_interc m1	6	0.0803	0.0406
28	logreg_sweep_C.uncert bc	6	0.0802	0.0405
29	cal_average.uncert m1	6	0.0799	0.0402
30	cal_average.uncert bc	6	0.0796	0.0399
31	average of TemperatureScaling	6	0.0793	0.0396
32	prob_average.uncert m1	6	0.0790	0.0393
33	logreg_no_interc sbt	6	0.0788	0.0391
34	prob_average.uncert bc	6	0.0788	0.0391
35	prob_average.uncert sbt	6	0.0788	0.0391
36	average.uncert bc	6	0.0787	0.0390
37	grad_bc.uncert bc	6	0.0787	0.0390
38	grad_m1.uncert bc	6	0.0787	0.0390
39	prob_average.uncert m2	6	0.0787	0.0390
40	grad_m2.uncert bc	6	0.0786	0.0389
41	average.uncert m1	6	0.0785	0.0388
42	grad_bc.uncert m1	6	0.0785	0.0388
43	grad_m1.uncert m1	6	0.0785	0.0388
44	grad_m2.uncert m1	6	0.0785	0.0388
45	grad_bc.uncert m2	6	0.0784	0.0387
46	grad_m1.uncert m2	6	0.0784	0.0387
47	grad_m2.uncert m2	6	0.0784	0.0387
48	average.uncert m2	6	0.0783	0.0386

49	average.uncert sbt	6	0.0783	0.0386
50	grad_bc.uncert sbt	6	0.0783	0.0386
51	grad_m1.uncert sbt	6	0.0783	0.0386
52	grad_m2.uncert sbt	6	0.0783	0.0386
53	cal_prob_average.uncert bc	6	0.0782	0.0385
54	cal_prob_average.uncert m1	6	0.0781	0.0384
55	cal_prob_average.uncert sbt	6	0.0781	0.0384
56	grad_m2 m2	6	0.0781	0.0384
57	logreg m1	6	0.0780	0.0383
58	cal_prob_average.uncert m2	6	0.0778	0.0381
59	grad_m2 sbt	6	0.0778	0.0381
60	logreg m2	6	0.0775	0.0378
61	grad_m1 m2	6	0.0774	0.0377
62	grad_m1 sbt	6	0.0771	0.0374
63	grad_m2 m1	6	0.0771	0.0374
64	grad_m1 m1	6	0.0770	0.0373
65	logreg_no_interc bc	6	0.0769	0.0372
66	logreg_no_interc m2	6	0.0764	0.0367
67	logreg_no_interc_sweep_C bc	6	0.0757	0.0360
68	logreg sbt	6	0.0747	0.0350
69	grad_bc bc	6	0.0722	0.0325
70	grad_bc m1	6	0.0713	0.0316
71	grad_m1 bc	6	0.0713	0.0316
72	grad_m2 bc	6	0.0710	0.0313
73	grad_bc m2	6	0.0708	0.0311
74	grad_bc sbt	6	0.0704	0.0307
75	average bc	6	0.0697	0.0300
76	average m1	6	0.0697	0.0300
77	average m2	6	0.0697	0.0300
78	average sbt	6	0.0697	0.0300
79	cal_prob_average m2	6	0.0683	0.0286
80	prob_average sbt	6	0.0670	0.0273
81	prob_average m2	6	0.0665	0.0268
82	prob_average m1	6	0.0664	0.0267
83	cal_prob_average sbt	6	0.0649	0.0252
84	cal_prob_average m1	6	0.0626	0.0229
85	logreg_sweep_C bc	6	0.0621	0.0224
86	logreg bc	6	0.0601	0.0204
87	prob_average bc	6	0.0567	0.0170
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 % Wed Feb 09 10:34:54 2022

	method	combination_size	imp_o_avg	imp_o_max
1	logreg_no_interc_sweep_C m1	6	0.0862	0.0465
2	logreg_no_interc_sweep_C sbt	6	0.0851	0.0454
3	logreg_no_interc_sweep_C m2	6	0.0832	0.0435
4	logreg_sweep_C sbt	6	0.0830	0.0433
5	logreg_no_interc.uncert sbt	6	0.0828	0.0431
6	logreg_sweep_C m2	6	0.0825	0.0428
7	logreg.uncert sbt	6	0.0822	0.0425
8	logreg_no_interc.uncert m2	6	0.0821	0.0424

9	logreg_no_interc.uncert m1	6	0.0820	0.0423
10	logreg_sweep_C m1	6	0.0820	0.0423
11	cal_average bc	6	0.0819	0.0422
12	cal_average m1	6	0.0819	0.0422
13	cal_average m2	6	0.0819	0.0422
14	logreg_no_interc.uncert bc	6	0.0817	0.0420
15	cal_average.uncert sbt	6	0.0816	0.0419
16	logreg_no_interc_sweep_C.uncert m2	6	0.0815	0.0418
17	logreg_no_interc_sweep_C.uncert sbt	6	0.0815	0.0418
18	logreg.uncert m2	6	0.0813	0.0416
19	logreg_no_interc_sweep_C.uncert m1	6	0.0812	0.0415
20	logreg_sweep_C.uncert m2	6	0.0811	0.0414
21	logreg.uncert bc	6	0.0809	0.0412
22	logreg.uncert m1	6	0.0808	0.0411
23	logreg_sweep_C.uncert sbt	6	0.0808	0.0411
24	cal_average.uncert m2	6	0.0807	0.0410
25	logreg_no_interc_sweep_C.uncert bc	6	0.0807	0.0410
26	logreg_sweep_C.uncert m1	6	0.0807	0.0410
27	logreg_no_interc m1	6	0.0803	0.0406
28	logreg_sweep_C.uncert bc	6	0.0802	0.0405
29	cal_average.uncert m1	6	0.0799	0.0402
30	cal_average.uncert bc	6	0.0796	0.0399
31	average of TemperatureScaling	6	0.0793	0.0396
32	prob_average.uncert m1	6	0.0790	0.0393
33	logreg_no_interc sbt	6	0.0788	0.0391
34	prob_average.uncert bc	6	0.0788	0.0391
35	prob_average.uncert sbt	6	0.0788	0.0391
36	average.uncert bc	6	0.0787	0.0390
37	grad_bc.uncert bc	6	0.0787	0.0390
38	grad_m1.uncert bc	6	0.0787	0.0390
39	prob_average.uncert m2	6	0.0787	0.0390
40	grad_m2.uncert bc	6	0.0786	0.0389
41	average.uncert m1	6	0.0785	0.0388
42	grad_bc.uncert m1	6	0.0785	0.0388
43	grad_m1.uncert m1	6	0.0785	0.0388
44	grad_m2.uncert m1	6	0.0785	0.0388
45	grad_bc.uncert m2	6	0.0784	0.0387
46	grad_m1.uncert m2	6	0.0784	0.0387
47	grad_m2.uncert m2	6	0.0784	0.0387
48	average.uncert m2	6	0.0783	0.0386
49	average.uncert sbt	6	0.0783	0.0386
50	grad_bc.uncert sbt	6	0.0783	0.0386
51	grad_m1.uncert sbt	6	0.0783	0.0386
52	grad_m2.uncert sbt	6	0.0783	0.0386
53	cal_prob_average.uncert bc	6	0.0782	0.0385
54	cal_prob_average.uncert m1	6	0.0781	0.0384
55	cal_prob_average.uncert sbt	6	0.0781	0.0384
56	grad_m2 m2	6	0.0781	0.0384
57	logreg m1	6	0.0780	0.0383
58	cal_prob_average.uncert m2	6	0.0778	0.0381
59	grad_m2 sbt	6	0.0778	0.0381
60	logreg m2	6	0.0775	0.0378
61	grad_m1 m2	6	0.0774	0.0377
62	grad_m1 sbt	6	0.0771	0.0374

63	grad_m2 m1	6	0.0771	0.0374
64	grad_m1 m1	6	0.0770	0.0373
65	logreg_no_interc bc	6	0.0769	0.0372
66	logreg_no_interc m2	6	0.0764	0.0367
67	logreg_no_interc_sweep_C bc	6	0.0757	0.0360
68	logreg sbt	6	0.0747	0.0350
69	grad_bc bc	6	0.0722	0.0325
70	grad_bc m1	6	0.0713	0.0316
71	grad_m1 bc	6	0.0713	0.0316
72	grad_m2 bc	6	0.0710	0.0313
73	grad_bc m2	6	0.0708	0.0311
74	grad_bc sbt	6	0.0704	0.0307
75	average bc	6	0.0697	0.0300
76	average m1	6	0.0697	0.0300
77	average m2	6	0.0697	0.0300
78	average sbt	6	0.0697	0.0300
79	cal_prob_average m2	6	0.0683	0.0286
80	prob_average sbt	6	0.0670	0.0273
81	prob_average m2	6	0.0665	0.0268
82	prob_average m1	6	0.0664	0.0267
83	cal_prob_average sbt	6	0.0649	0.0252
84	cal_prob_average m1	6	0.0626	0.0229
85	logreg_sweep_C bc	6	0.0621	0.0224
86	logreg bc	6	0.0601	0.0204
87	prob_average bc	6	0.0567	0.0170
88	cal_prob_average bc	6	0.0458	0.0061
89	cal_average sbt	6	-0.2539	-0.2936