

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_c10/0/evaluation_val
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$nl1), max(cur_nets$nl1), mean(cur_nets$nl1), var(cur_nets$nl1),
    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

```

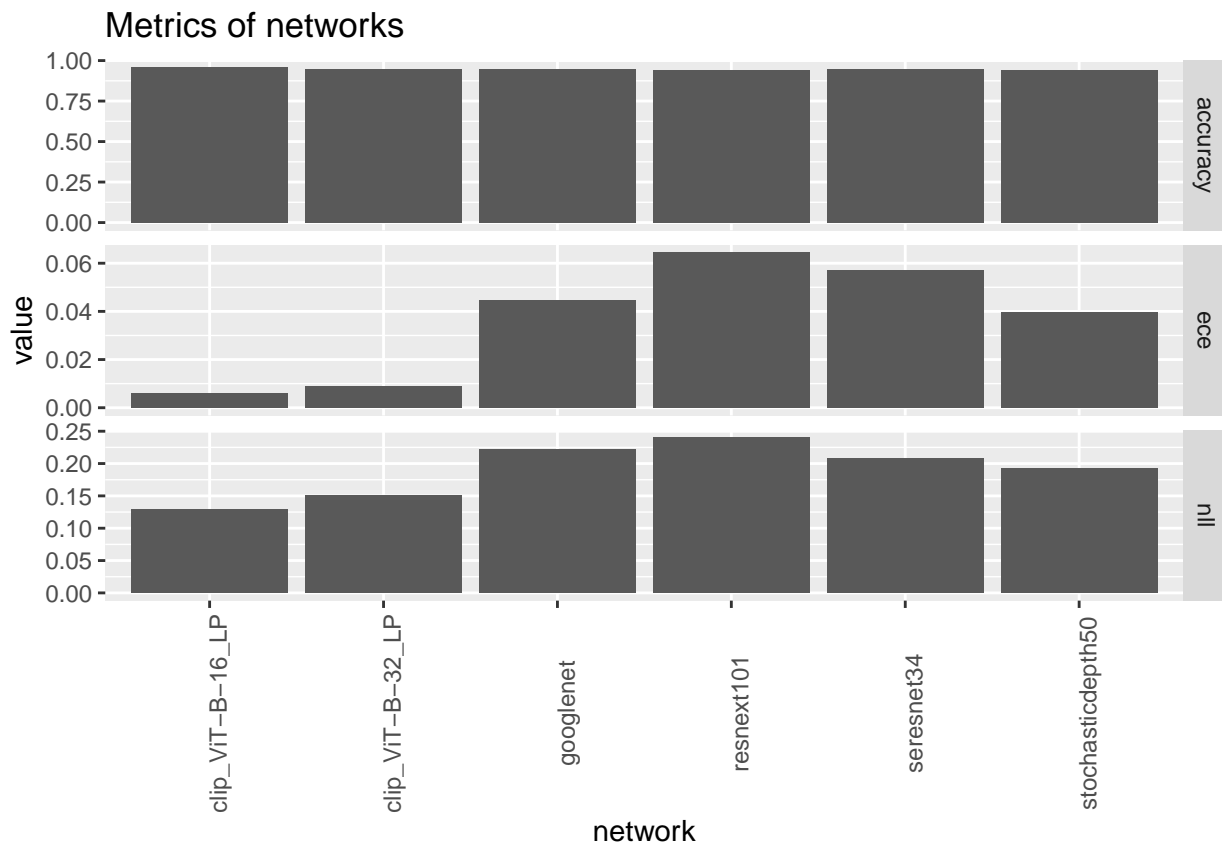
Metrics for networks.

```

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.

```

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_hline(
      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_hline(
    data = cur_ens_pwc,
    mapping = aes(
      x = combining_method, y = ece,
      colour2 = coupling_method
    ),
    size = 0.8, width = 0.11,
    position = position_dodge(width = 0.65)
  ) %>%
  rename_geom_aes(new_aes = c("colour" = "colour2"))
) +
scale_colour_brewer(
  aesthetics = "colour1", palette = 1,
  name = "network", type = "qual"
) +
scale_colour_brewer(
  aesthetics = "colour2", palette = 2,
  name = "coupling method", type = "qual"
) +
scale_color_manual(values = c("black"), name = "averaging ensemble") +
scale_y_reverse() +
theme(axis.text.x = element_text(angle = 90))

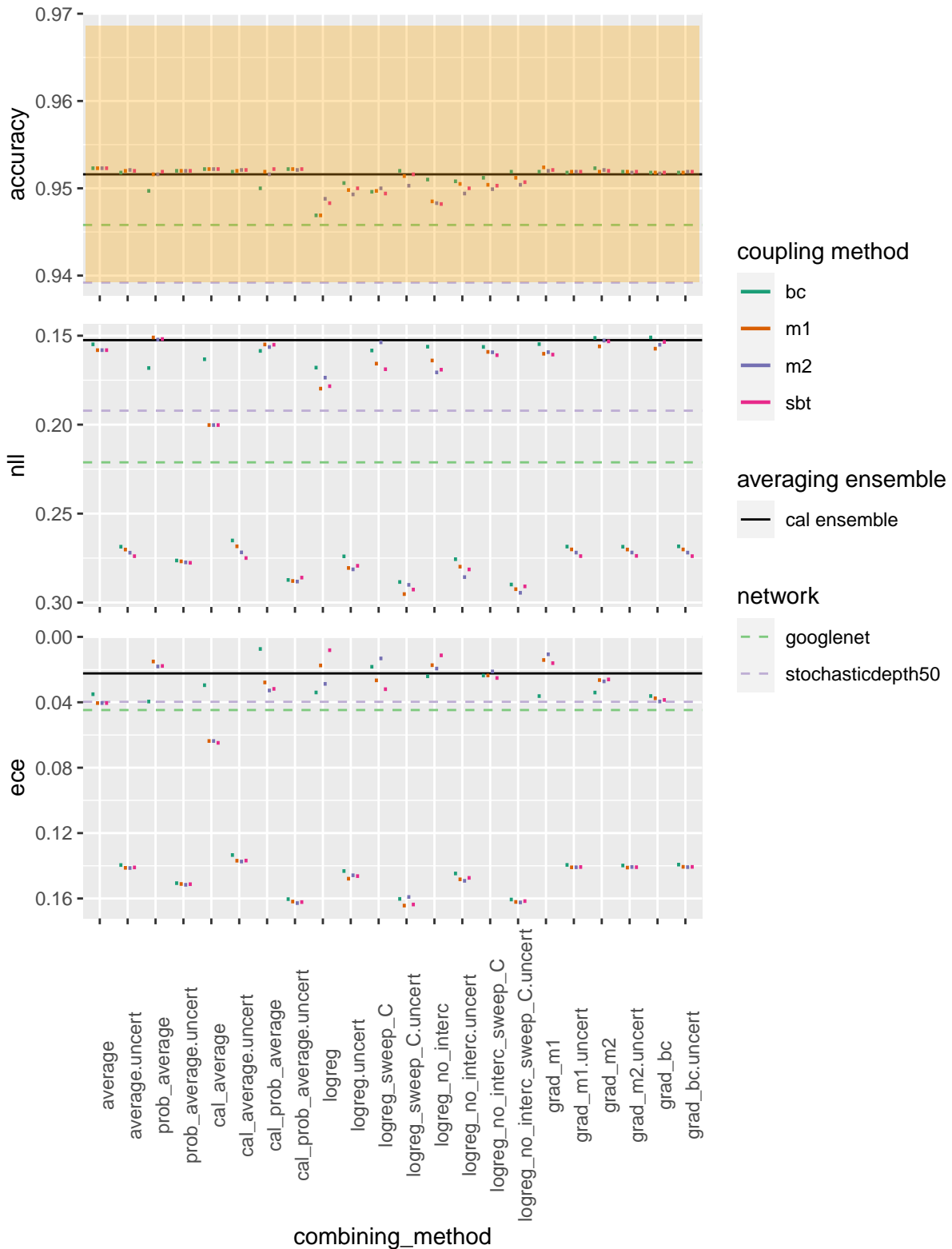
print(acc_plot / nll_plot / ece_plot + plot_layout(guides = "collect") +
  plot_annotation(title = paste(
    "Ensemble metrics",
    paste(
      c("Error inconsistency", cur_ens_cal$err_incons[[1]]),
      collapse = " "
    ),
    paste(
      c("Average pairwise accuracy variance", cur_ens_cal$mean_pwa_var[[1]]),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

```

Ensemble metrics

Error inconsistency 0.0521999970078468

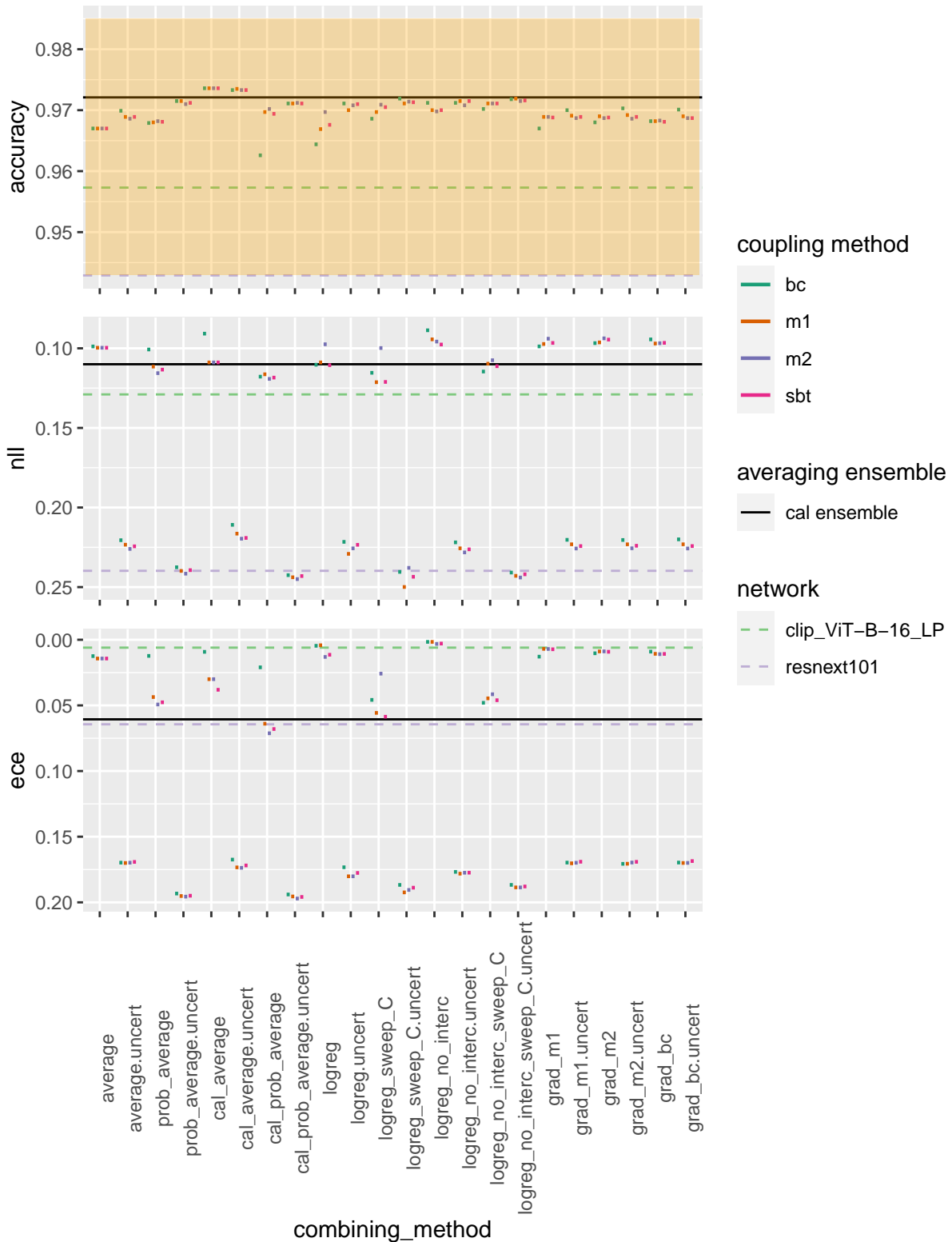
Average pairwise accuracy variance 1.78000709638582e-06



Ensemble metrics

Error inconsistency 0.0697999969124794

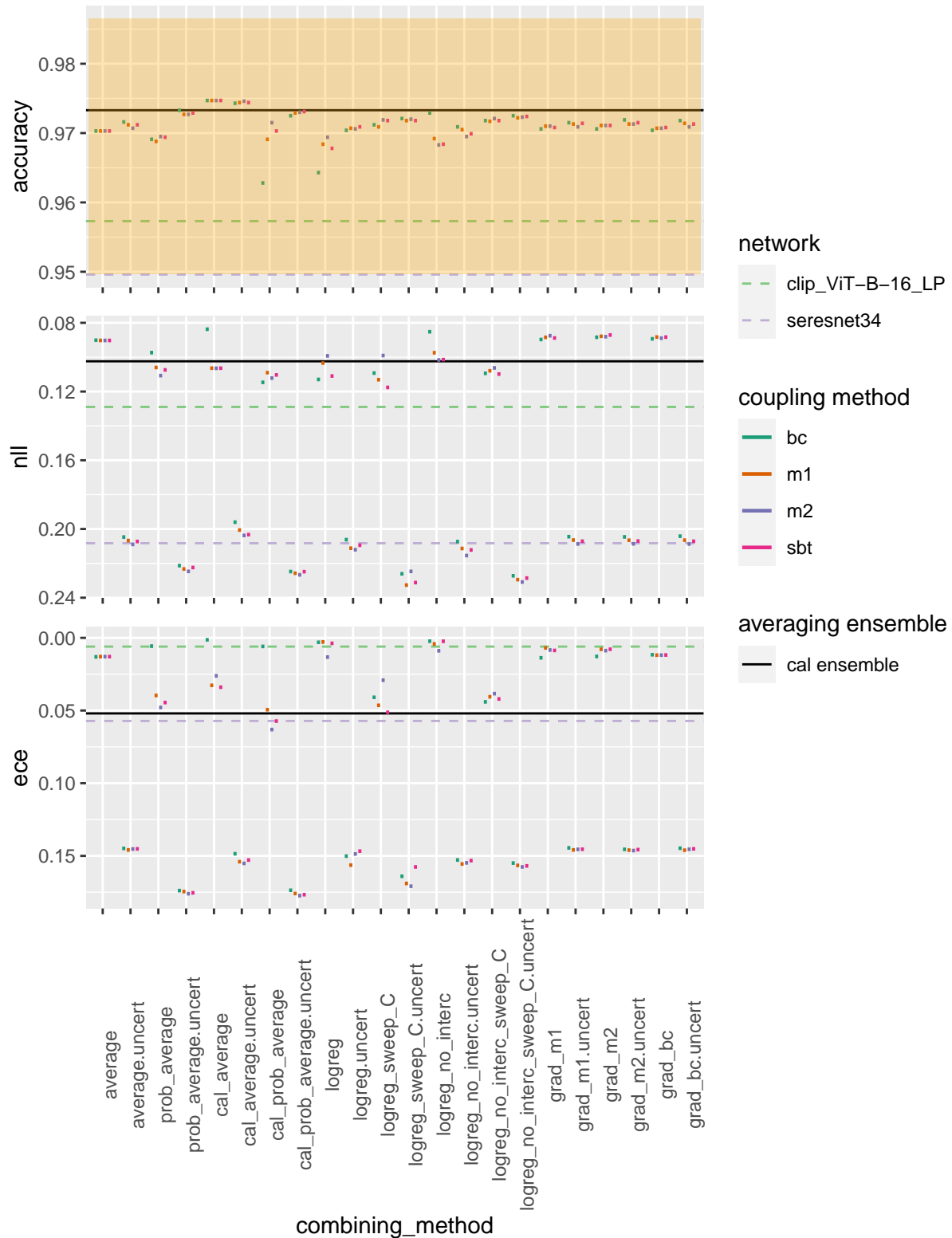
Average pairwise accuracy variance 7.22872891856241e-06



Ensemble metrics

Error inconsistency 0.066100001335144

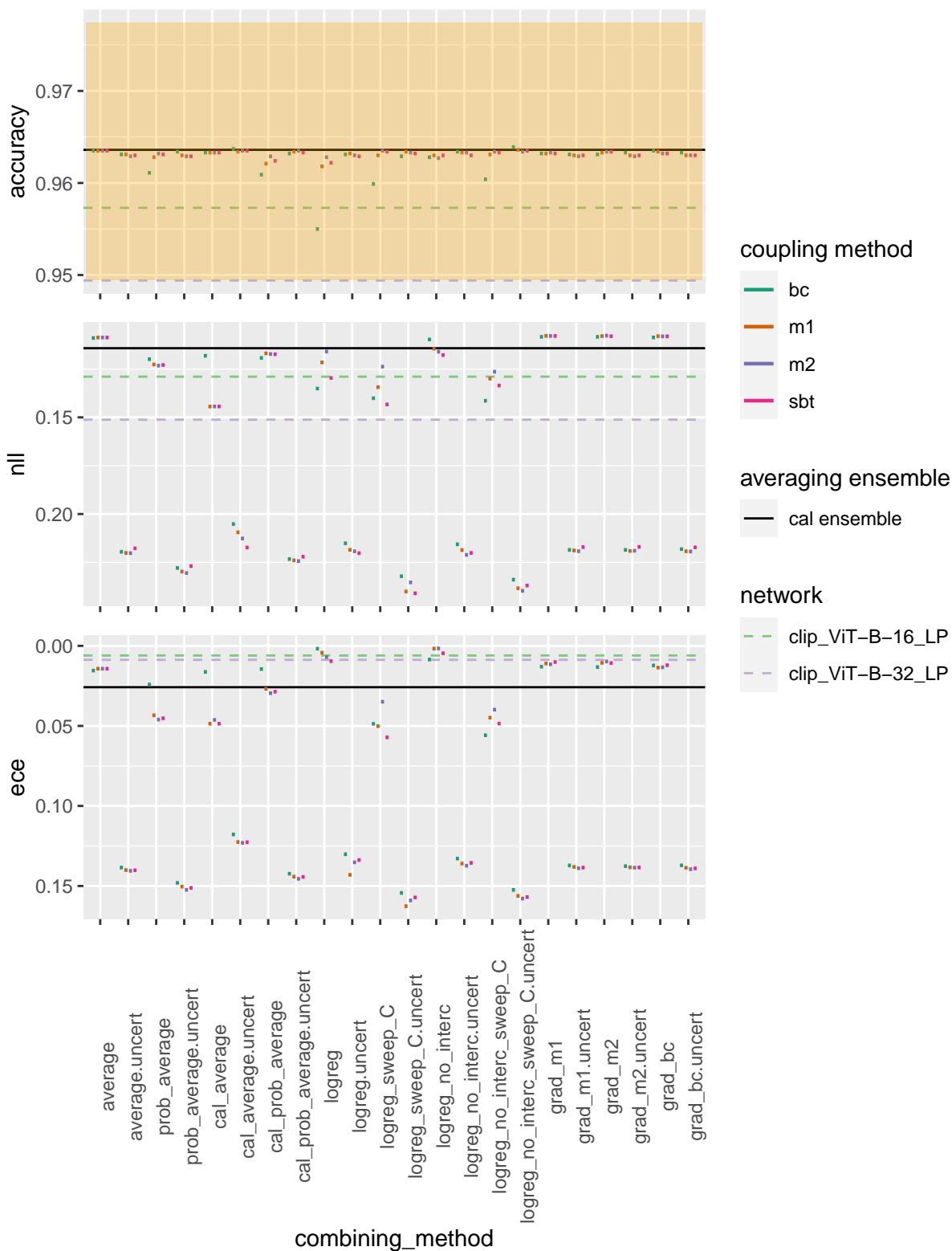
Average pairwise accuracy variance 3.92749143429683e-06



Ensemble metrics

Error inconsistency 0.0480999983847141

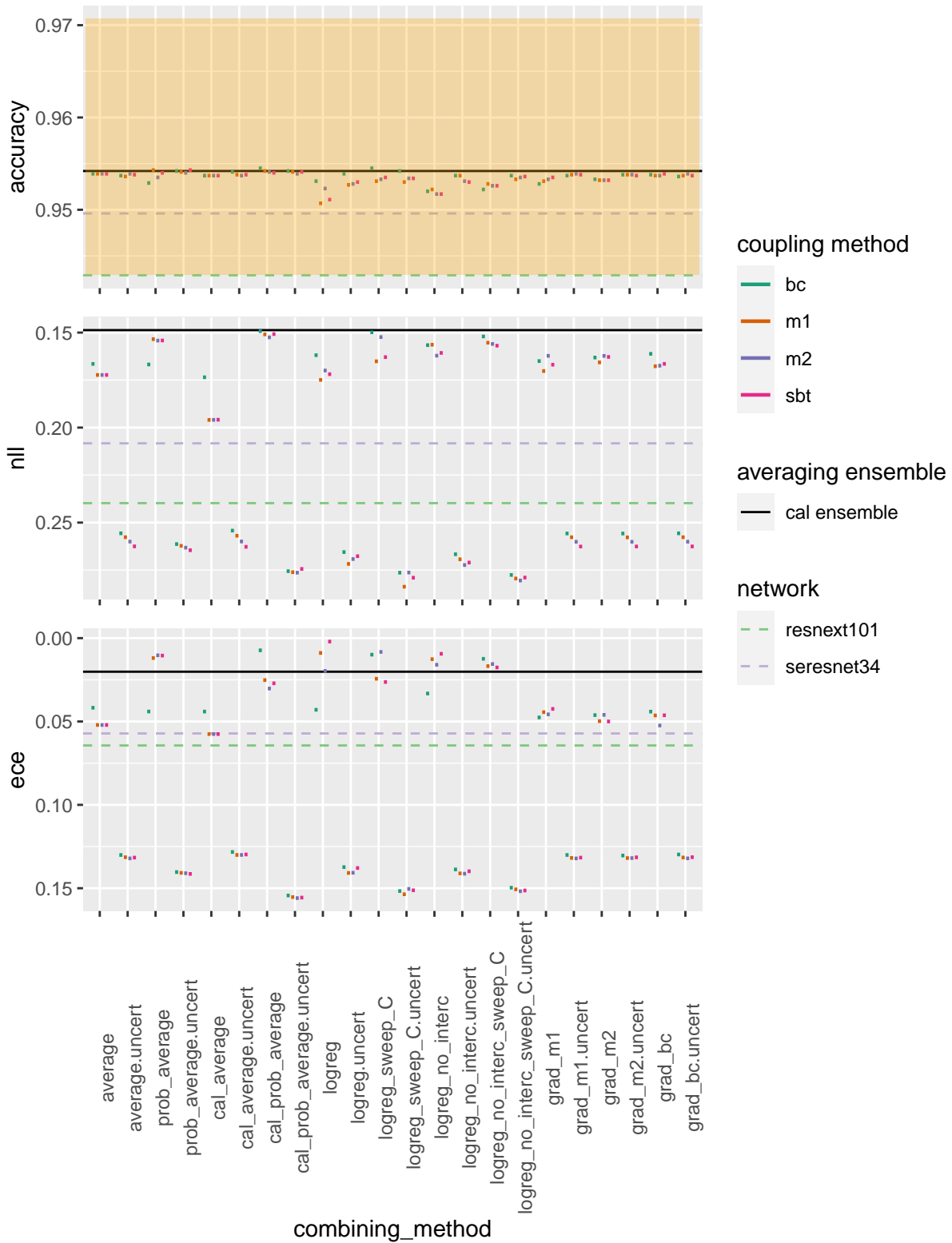
Average pairwise accuracy variance 2.50124298872834e-06



Ensemble metrics

Error inconsistency 0.0488999970257282

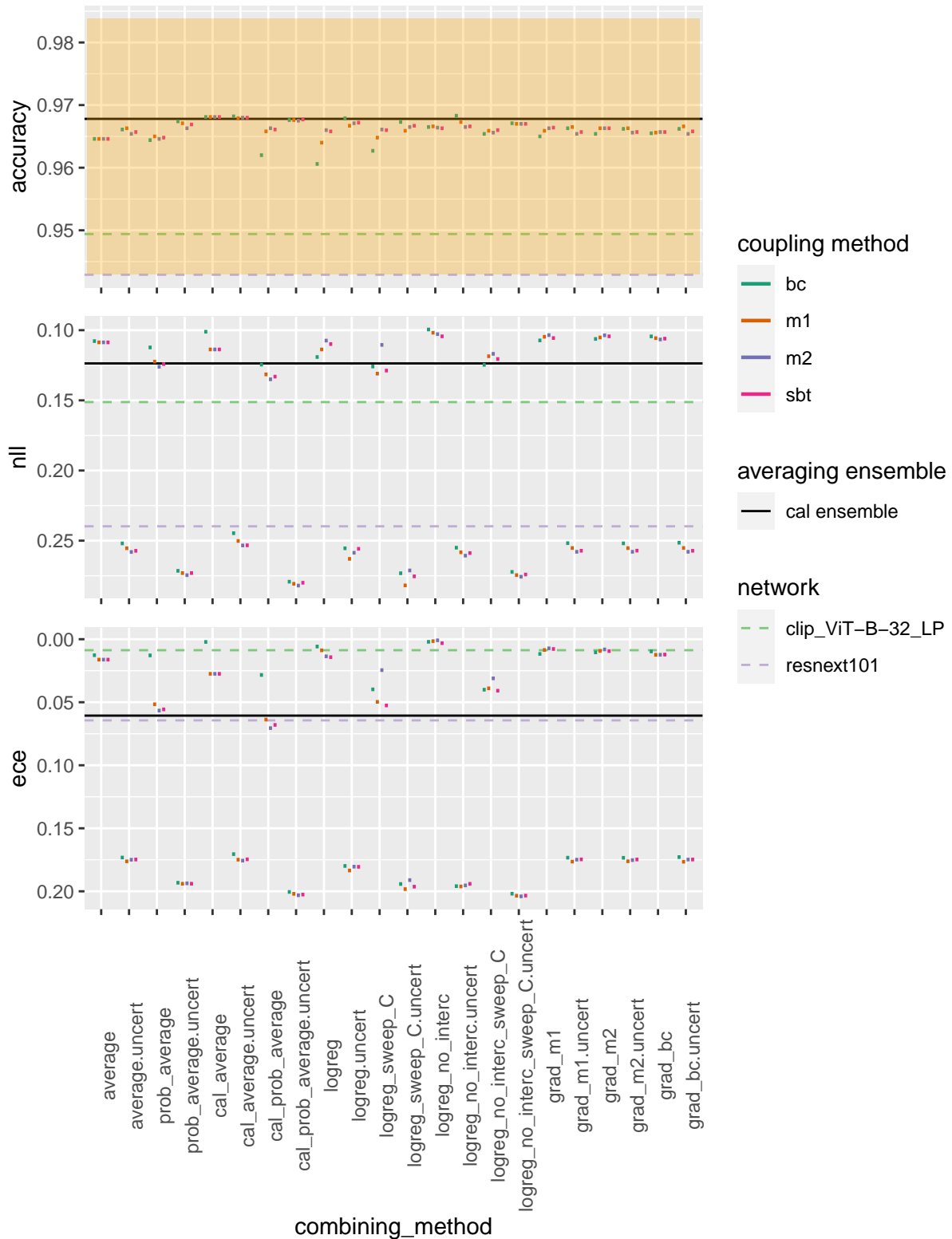
Average pairwise accuracy variance 1.72874490544928e-06



Ensemble metrics

Error inconsistency 0.0753000006079674

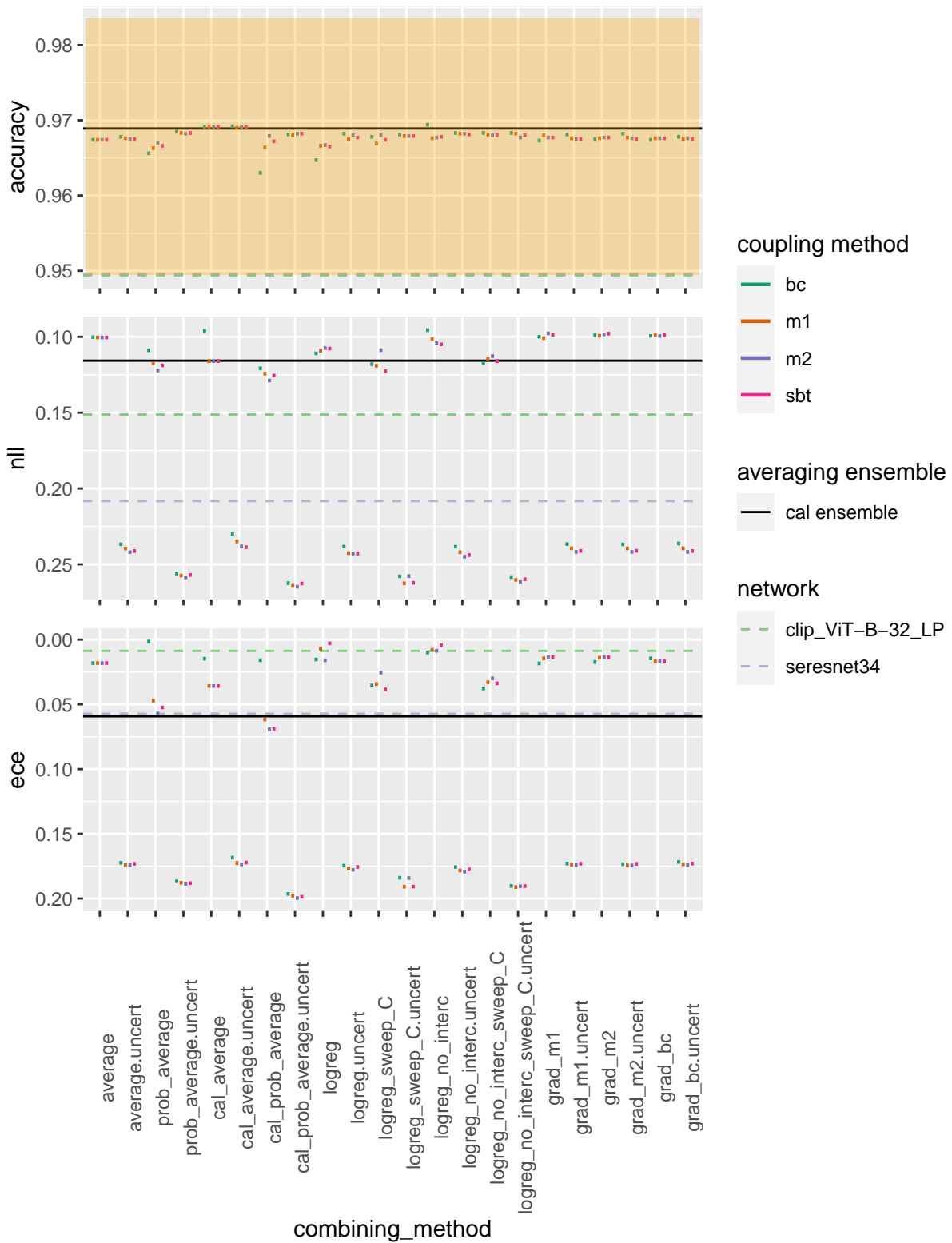
Average pairwise accuracy variance 4.93749848828884e-06



Ensemble metrics

Error inconsistency 0.0679999962449074

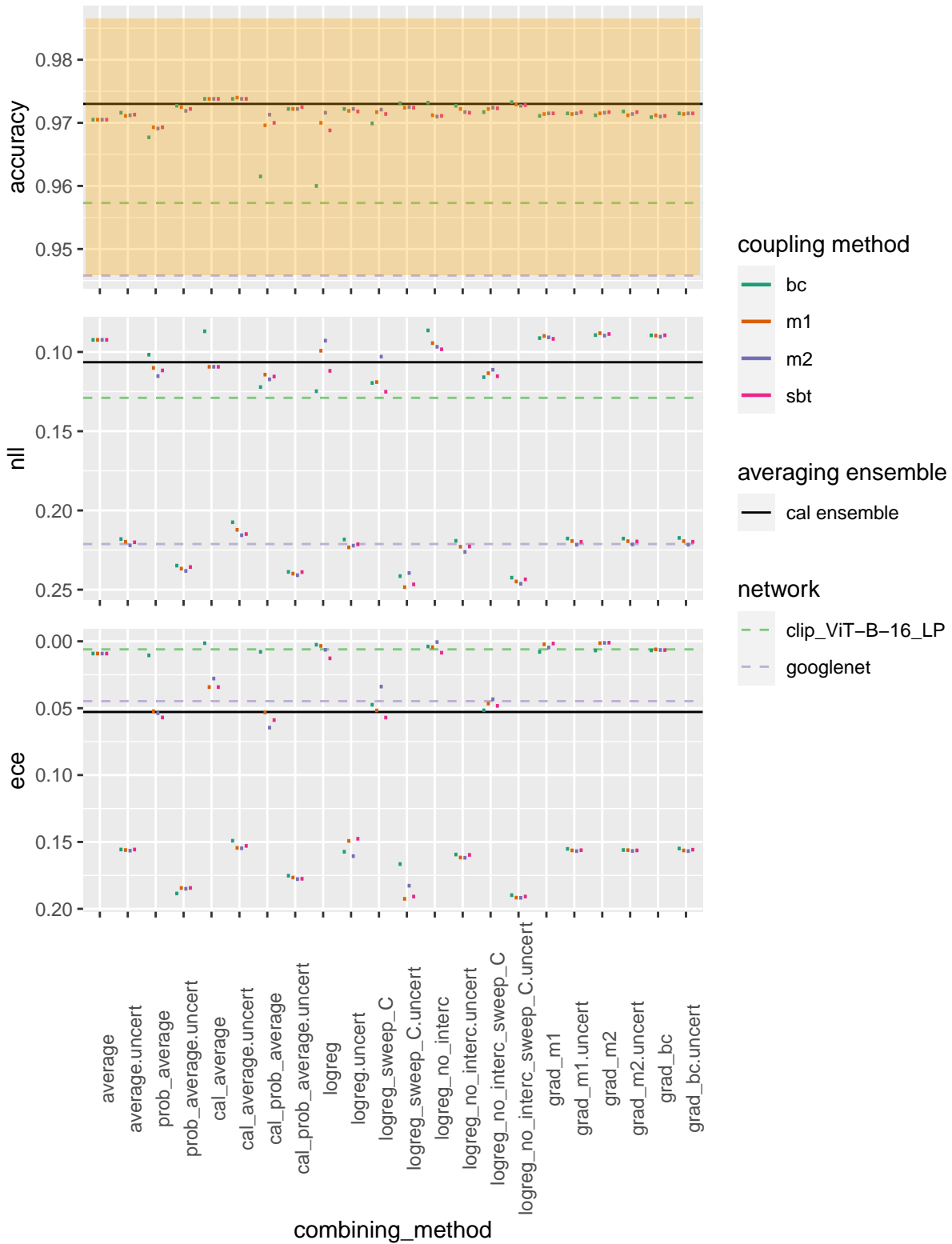
Average pairwise accuracy variance 3.35625622938096e-06



Ensemble metrics

Error inconsistency 0.0698999986052513

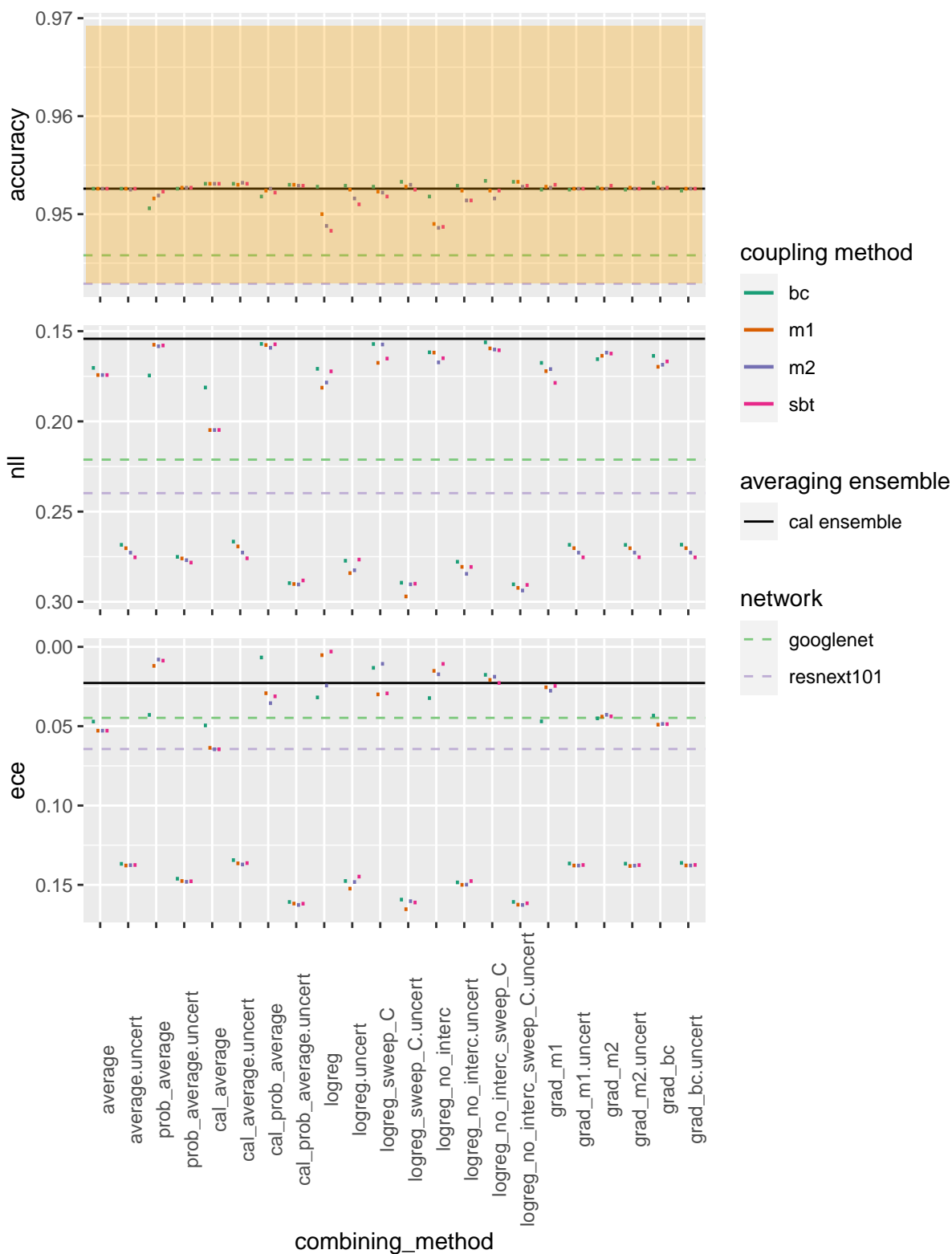
Average pairwise accuracy variance 5.5337259254884e-06



Ensemble metrics

Error inconsistency 0.0496999993920326

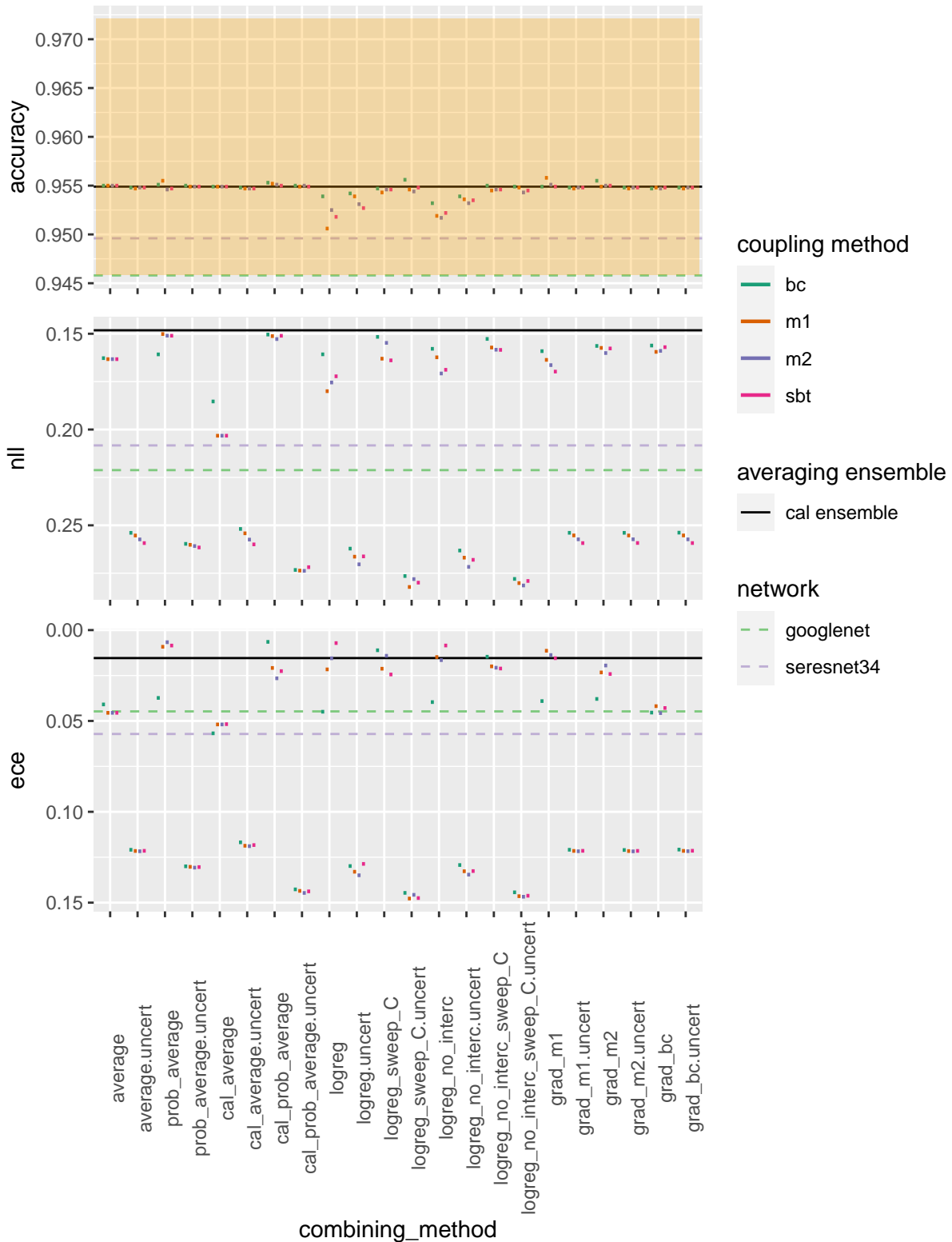
Average pairwise accuracy variance 1.15999603167438e-06



Ensemble metrics

Error inconsistency 0.0487999990582466

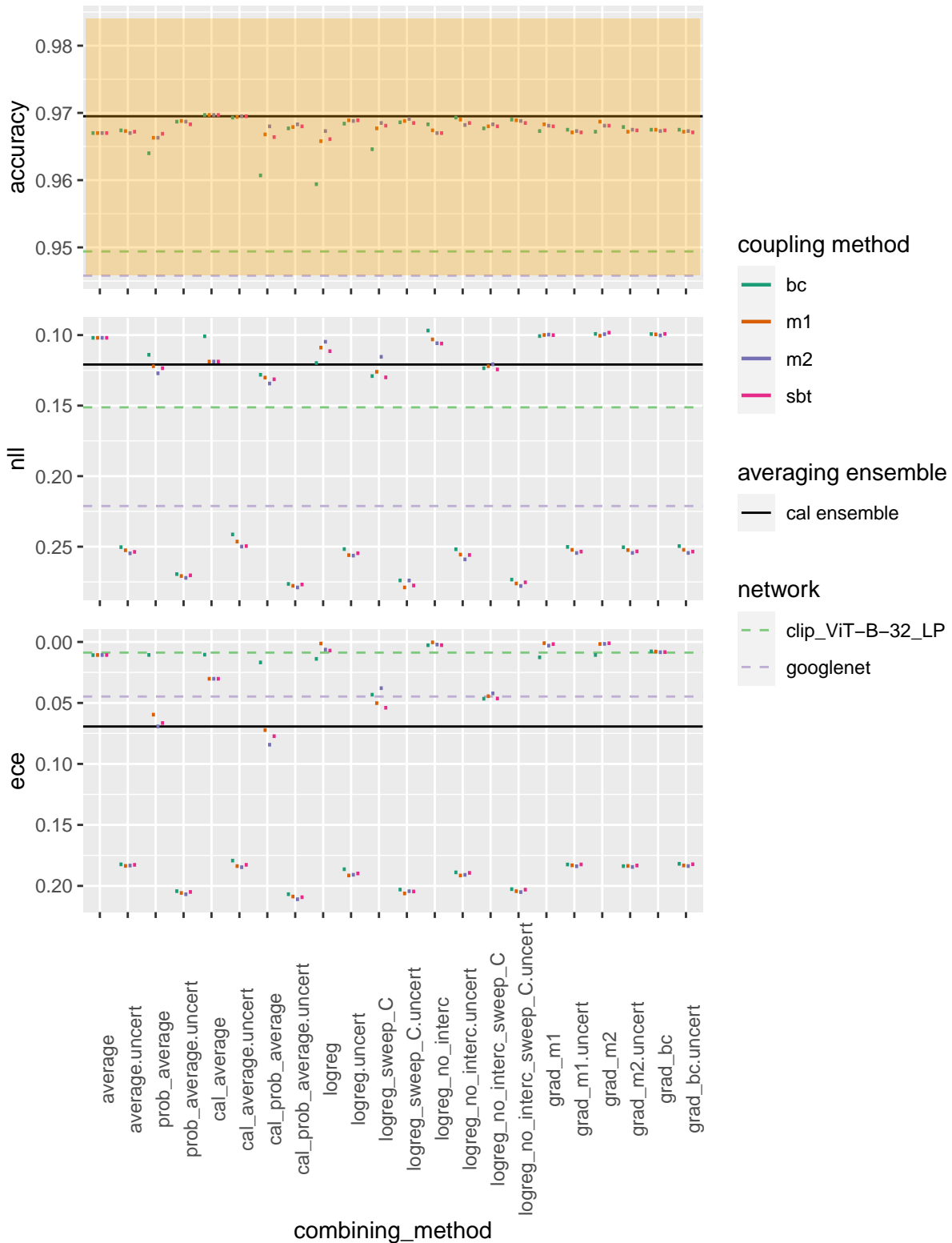
Average pairwise accuracy variance 1.13374744614703e-06



Ensemble metrics

Error inconsistency 0.0727999955415726

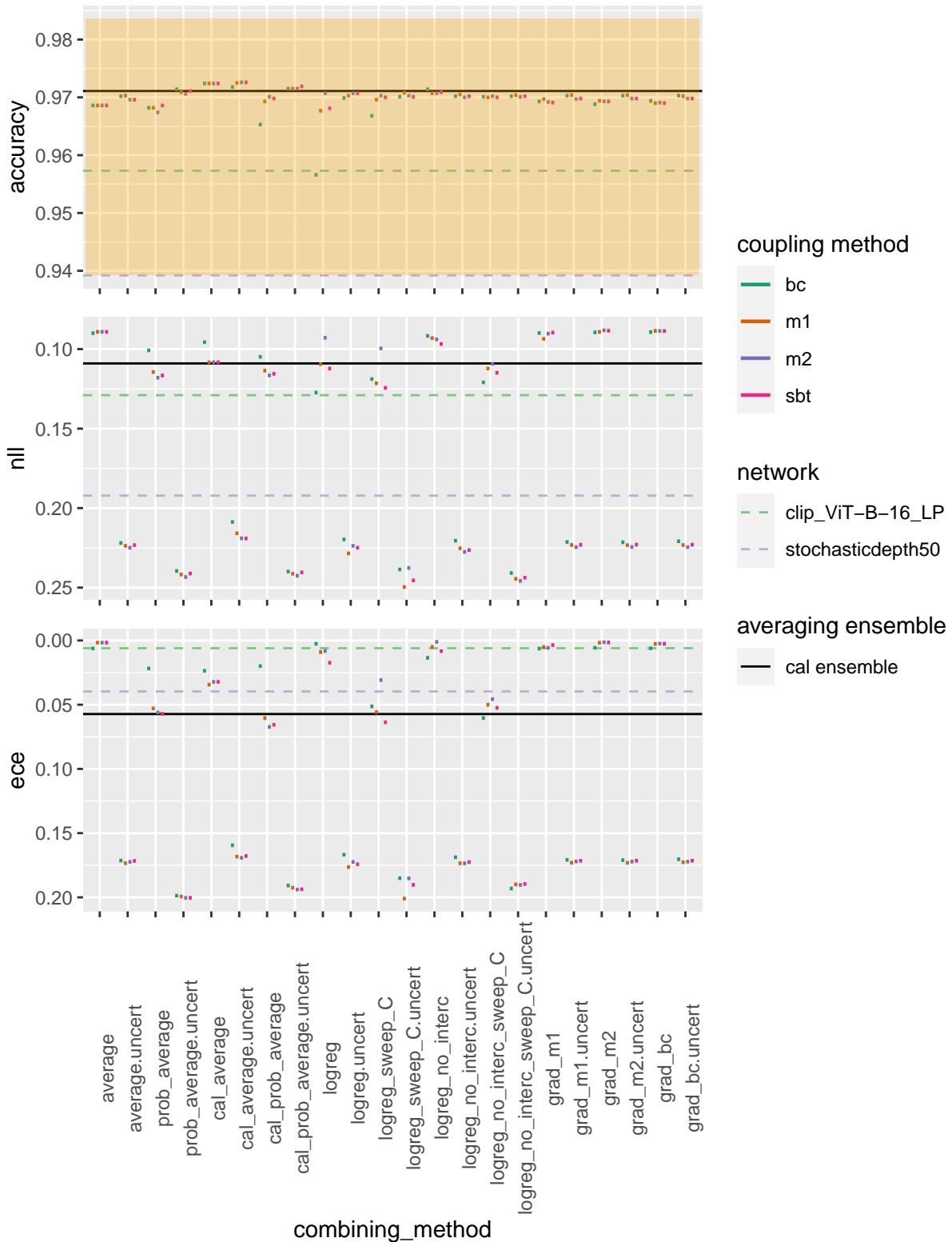
Average pairwise accuracy variance 4.57249961982598e-06



Ensemble metrics

Error inconsistency 0.0706999972462654

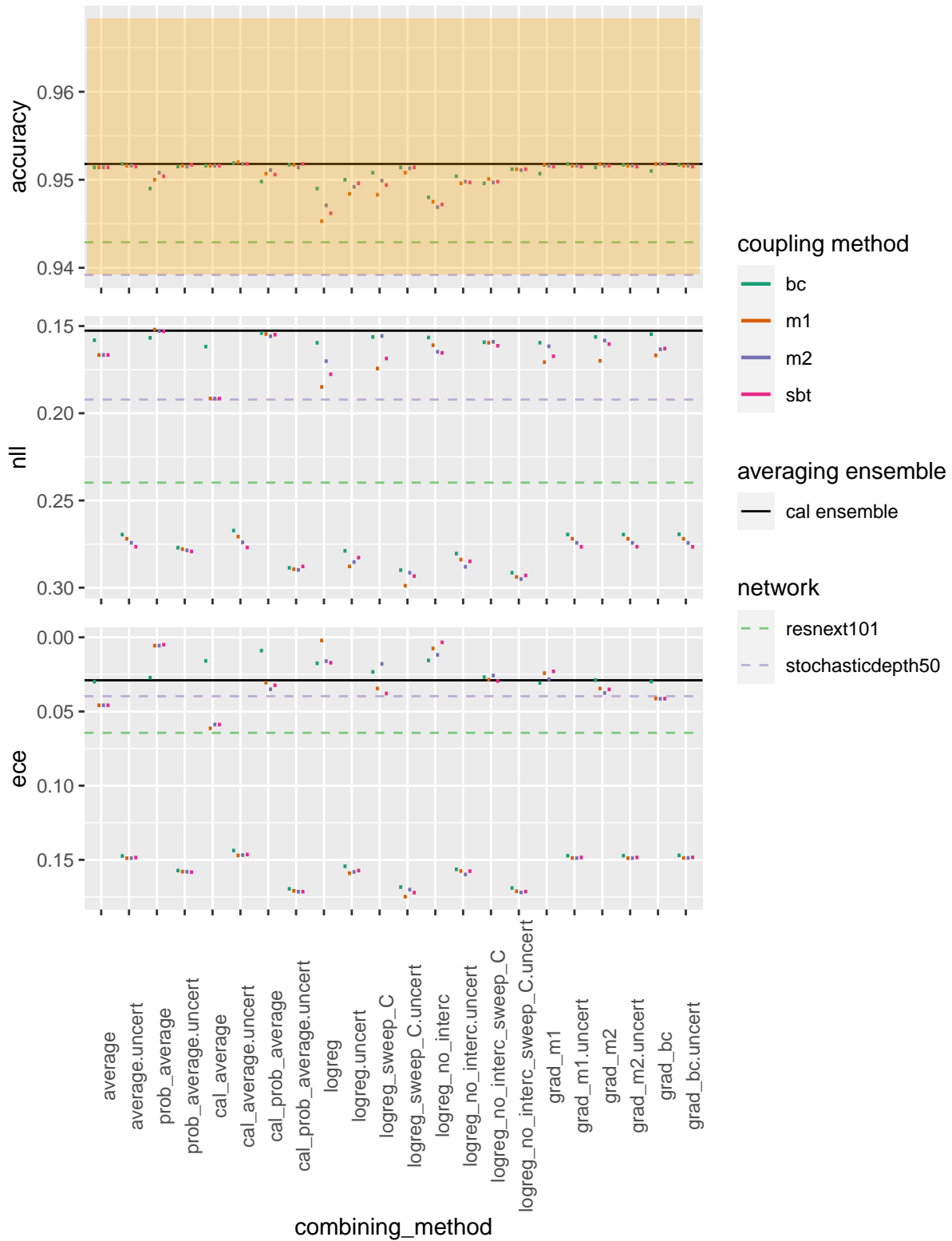
Average pairwise accuracy variance 8.63124751049327e-06



Ensemble metrics

Error inconsistency 0.0544999986886978

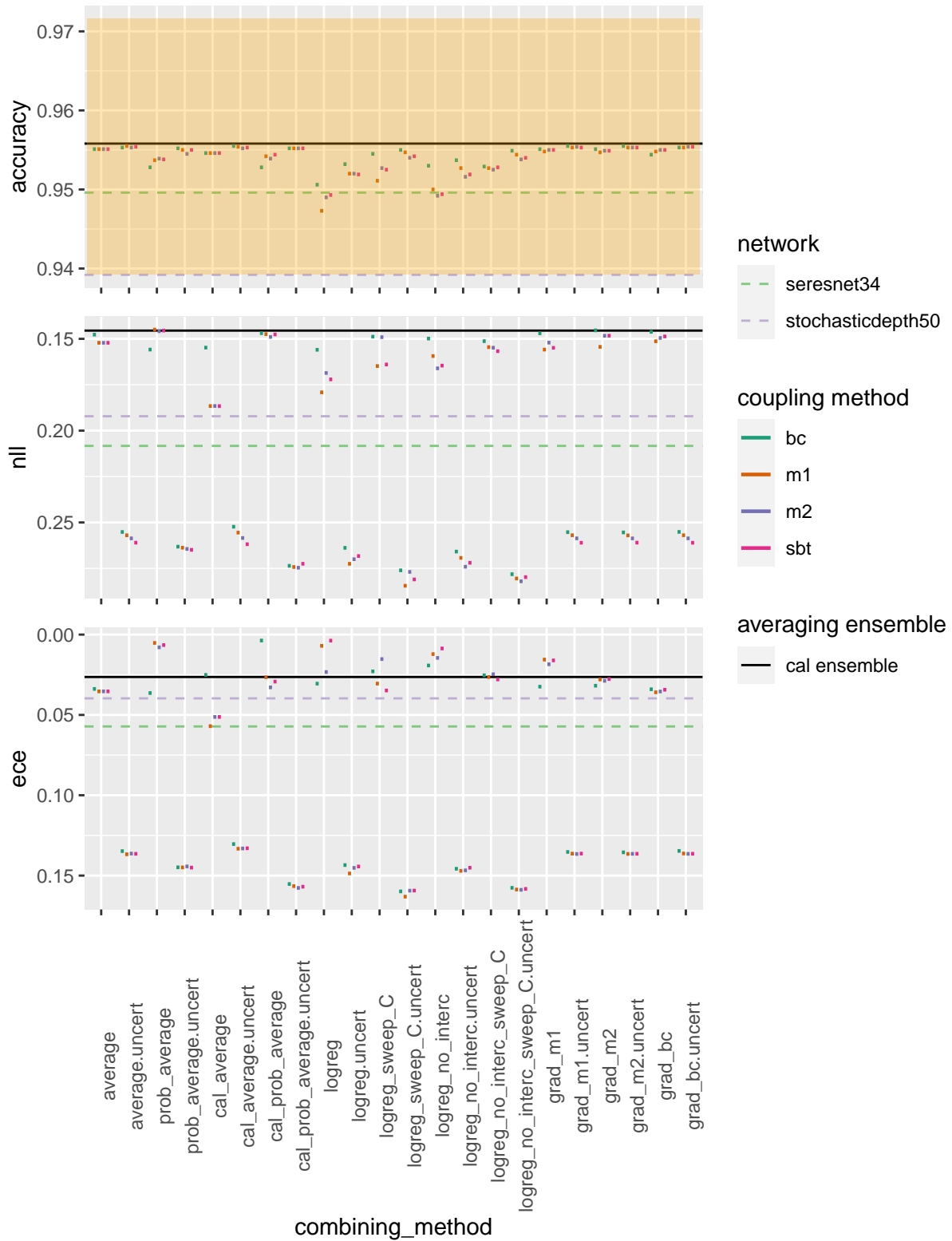
Average pairwise accuracy variance 1.10750431758788e-06



Ensemble metrics

Error inconsistency 0.0543999969959259

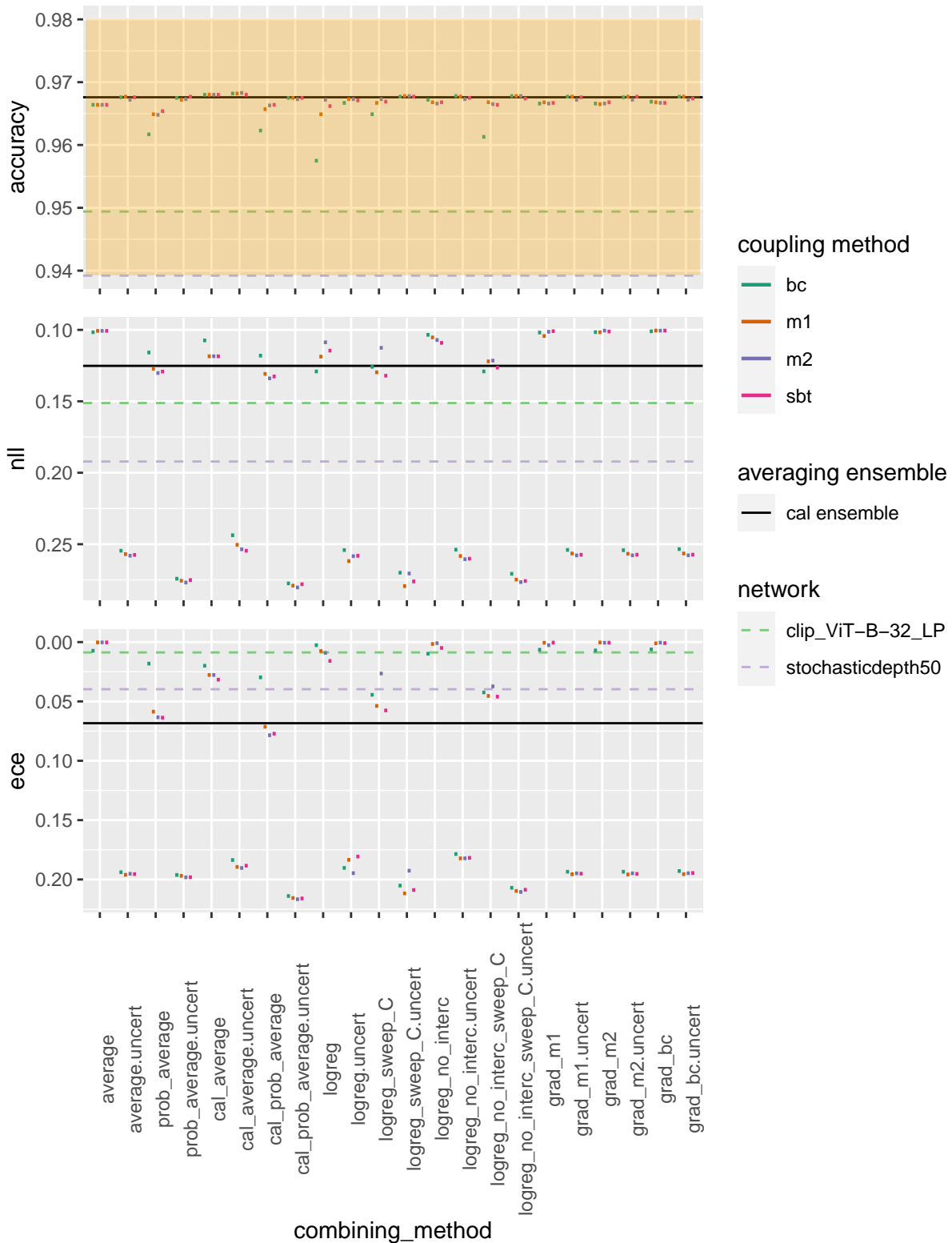
Average pairwise accuracy variance 2.44625221057504e-06



Ensemble metrics

Error inconsistency 0.0715999975800514

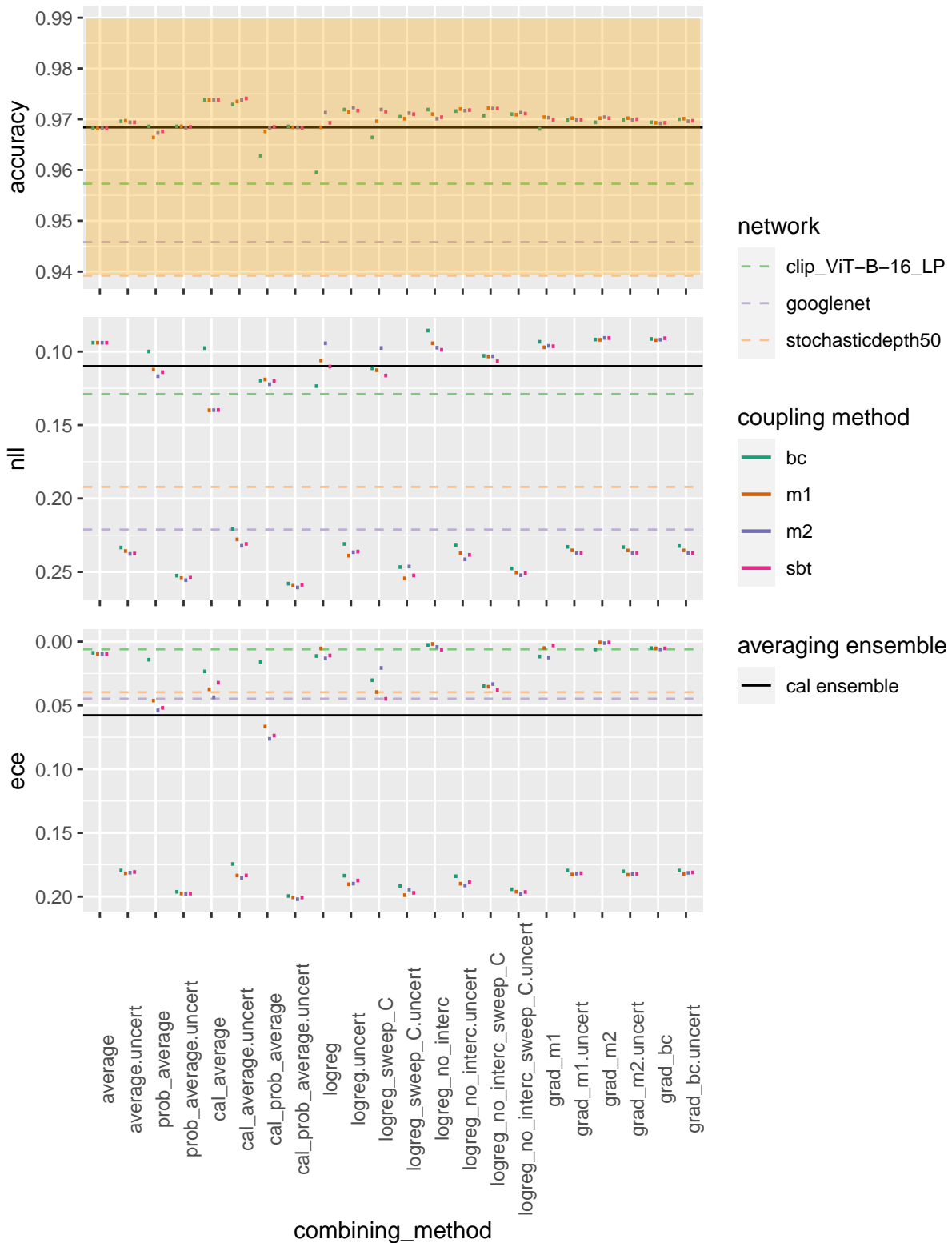
Average pairwise accuracy variance 6.37750827081618e-06



Ensemble metrics

Error inconsistency 0.0964000001549721

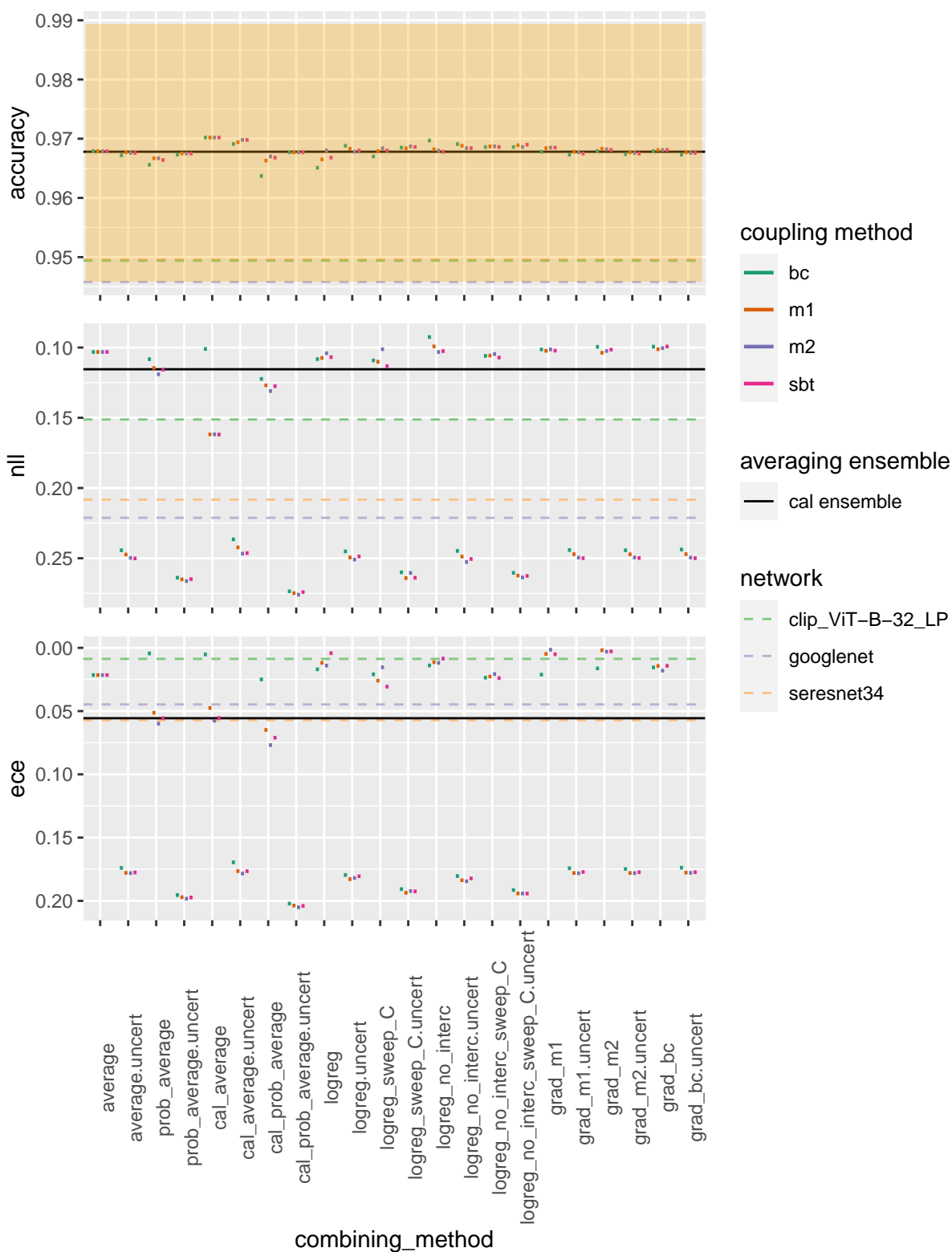
Average pairwise accuracy variance 7.08665675119846e-06



Ensemble metrics

Error inconsistency 0.0947999954223633

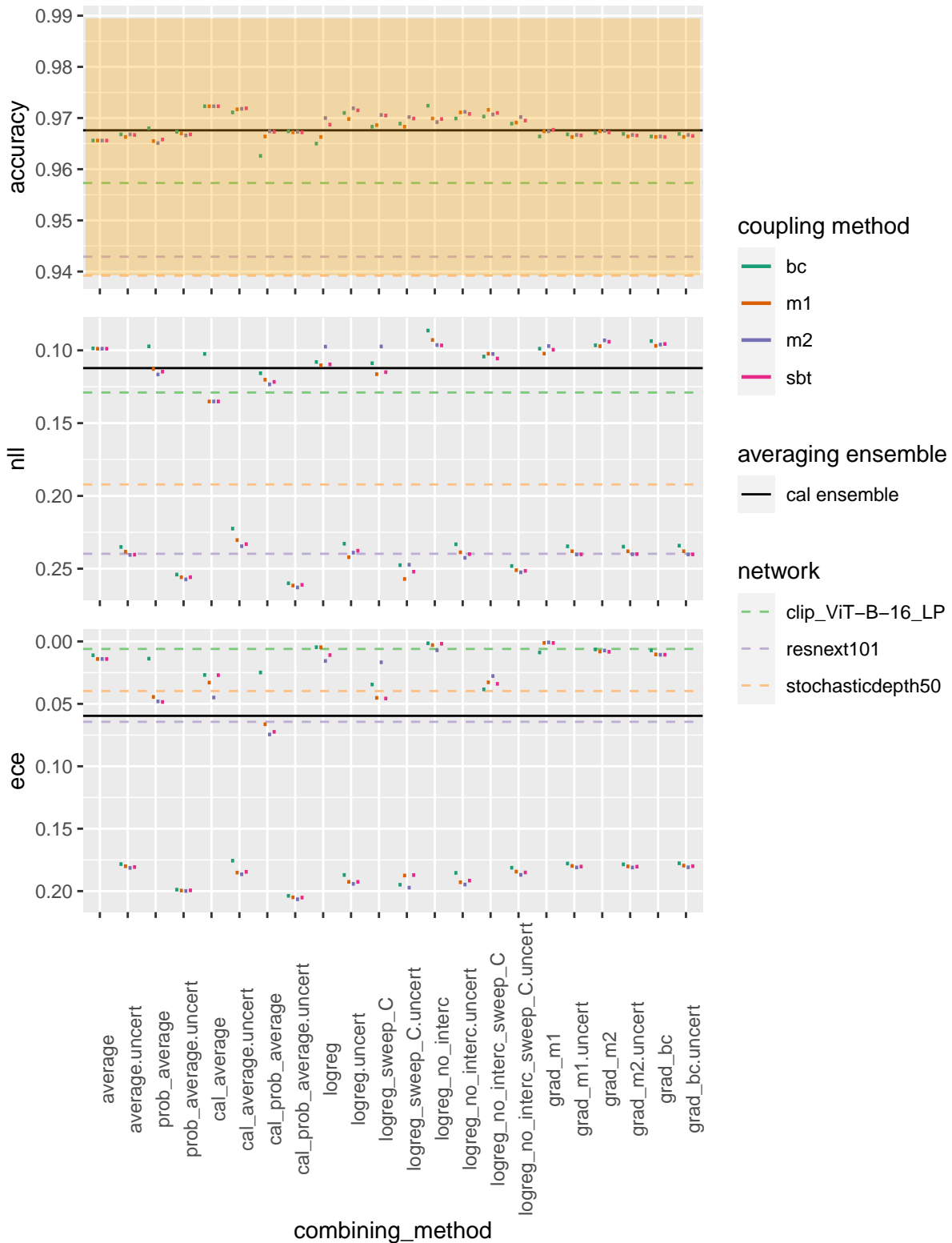
Average pairwise accuracy variance 4.02777959607192e-06



Ensemble metrics

Error inconsistency 0.0974999964237213

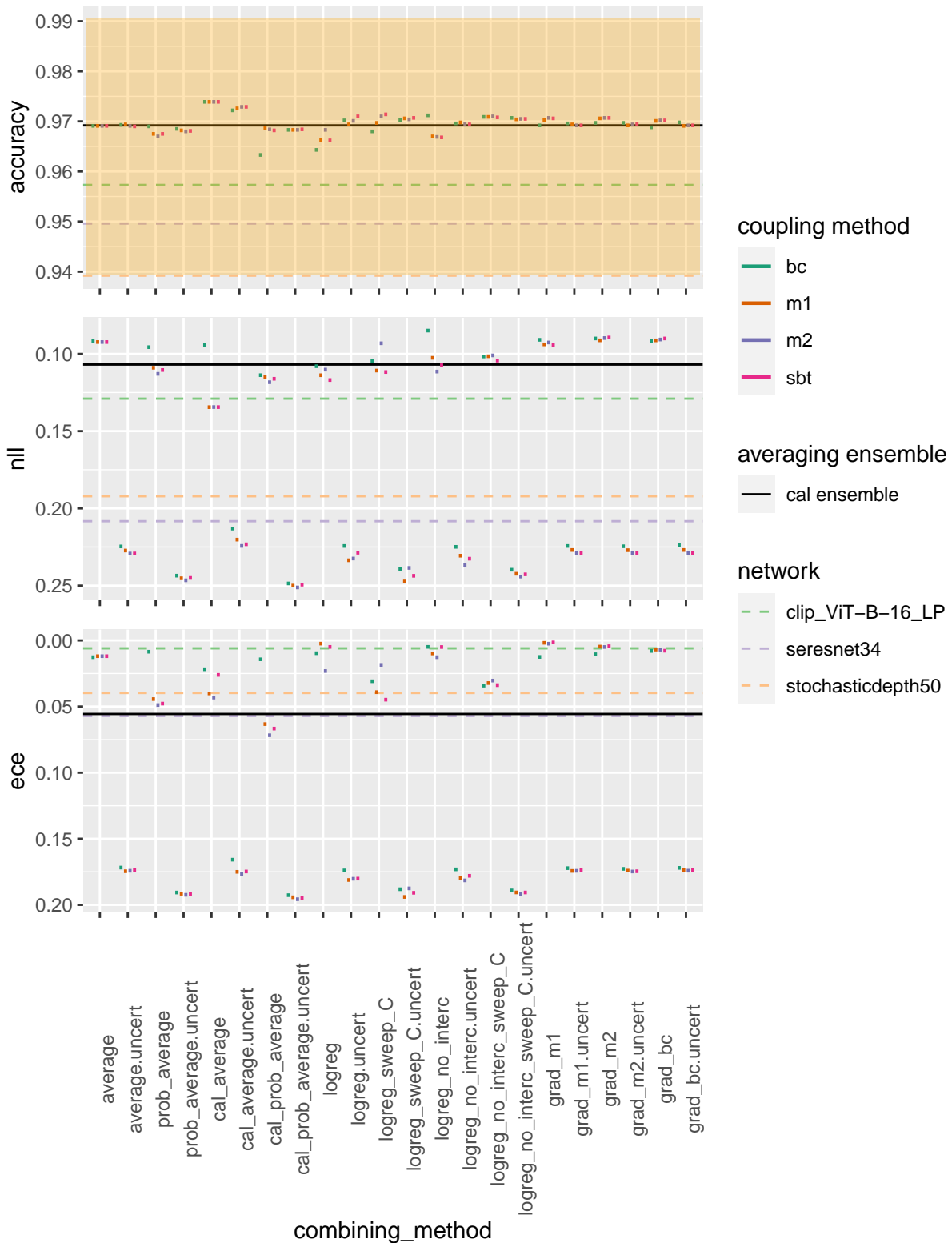
Average pairwise accuracy variance 7.54110624257009e-06



Ensemble metrics

Error inconsistency 0.0955999940633774

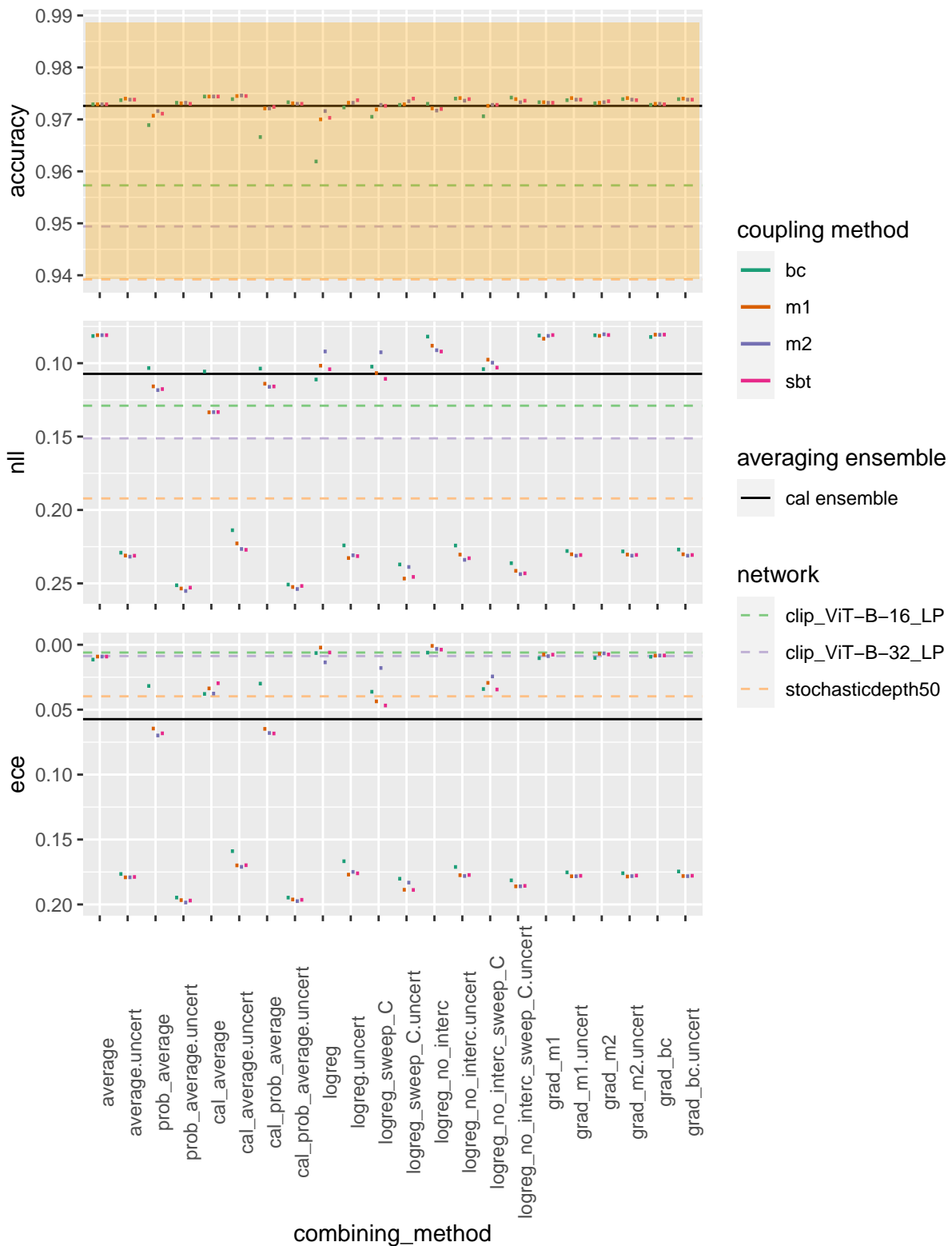
Average pairwise accuracy variance 6.66888627165463e-06



Ensemble metrics

Error inconsistency 0.0951999947428703

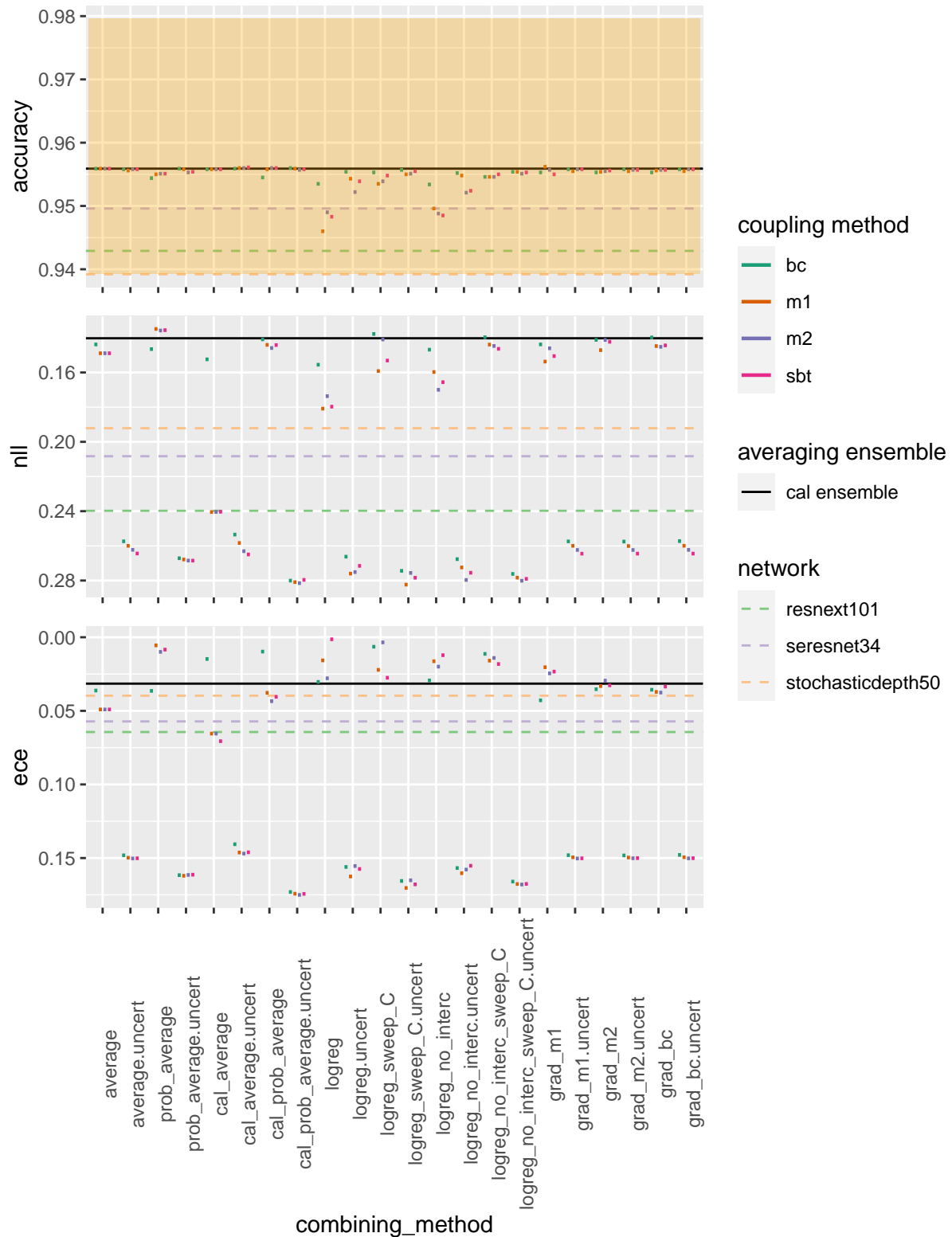
Average pairwise accuracy variance 7.78223238739884e-06



Ensemble metrics

Error inconsistency 0.0788999944925308

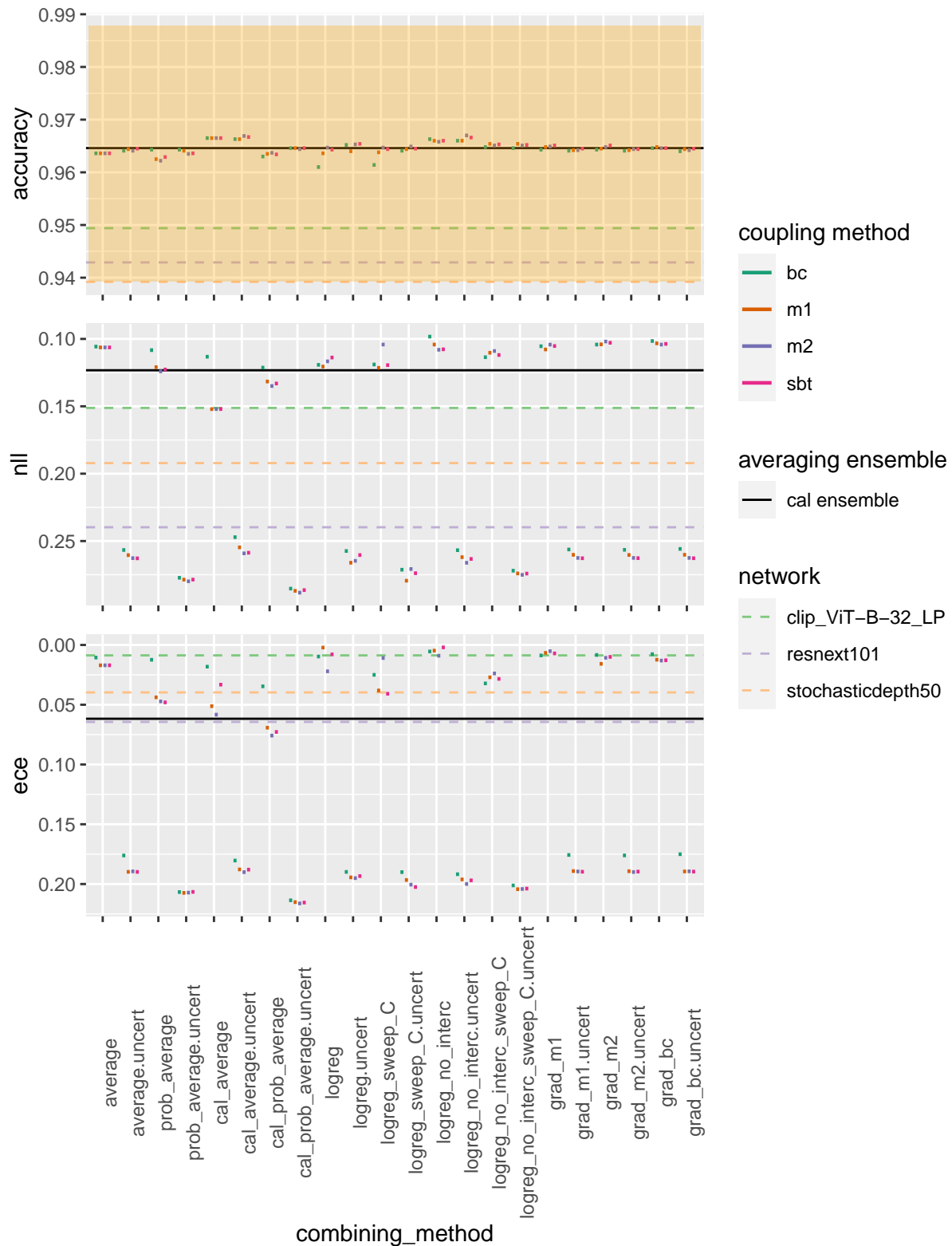
Average pairwise accuracy variance $2.3477746253775e-06$



Ensemble metrics

Error inconsistency 0.100699998438358

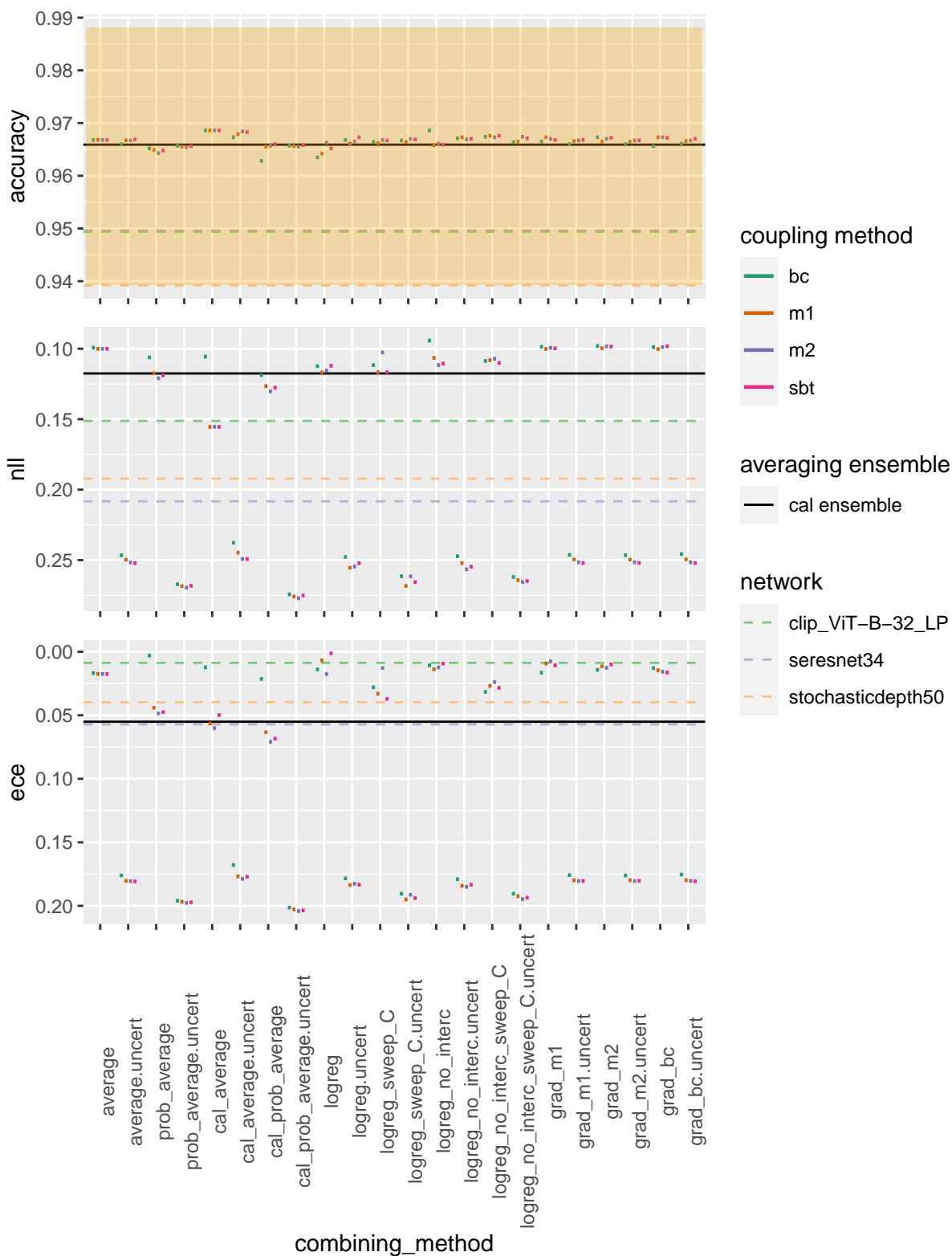
Average pairwise accuracy variance 5.52111623619567e-06



Ensemble metrics

Error inconsistency 0.0969999954104424

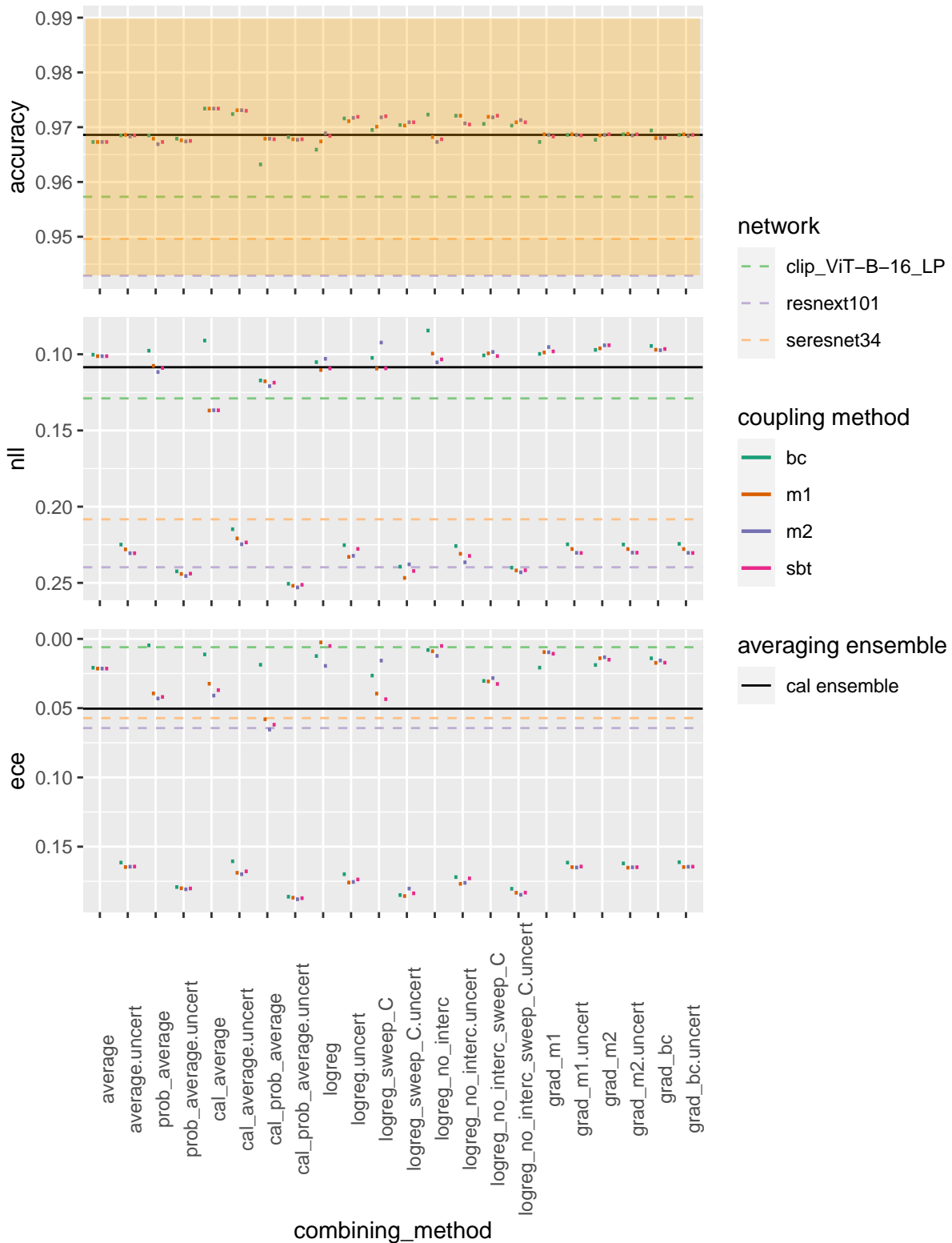
Average pairwise accuracy variance 5.41334156878293e-06



Ensemble metrics

Error inconsistency 0.092399999499321

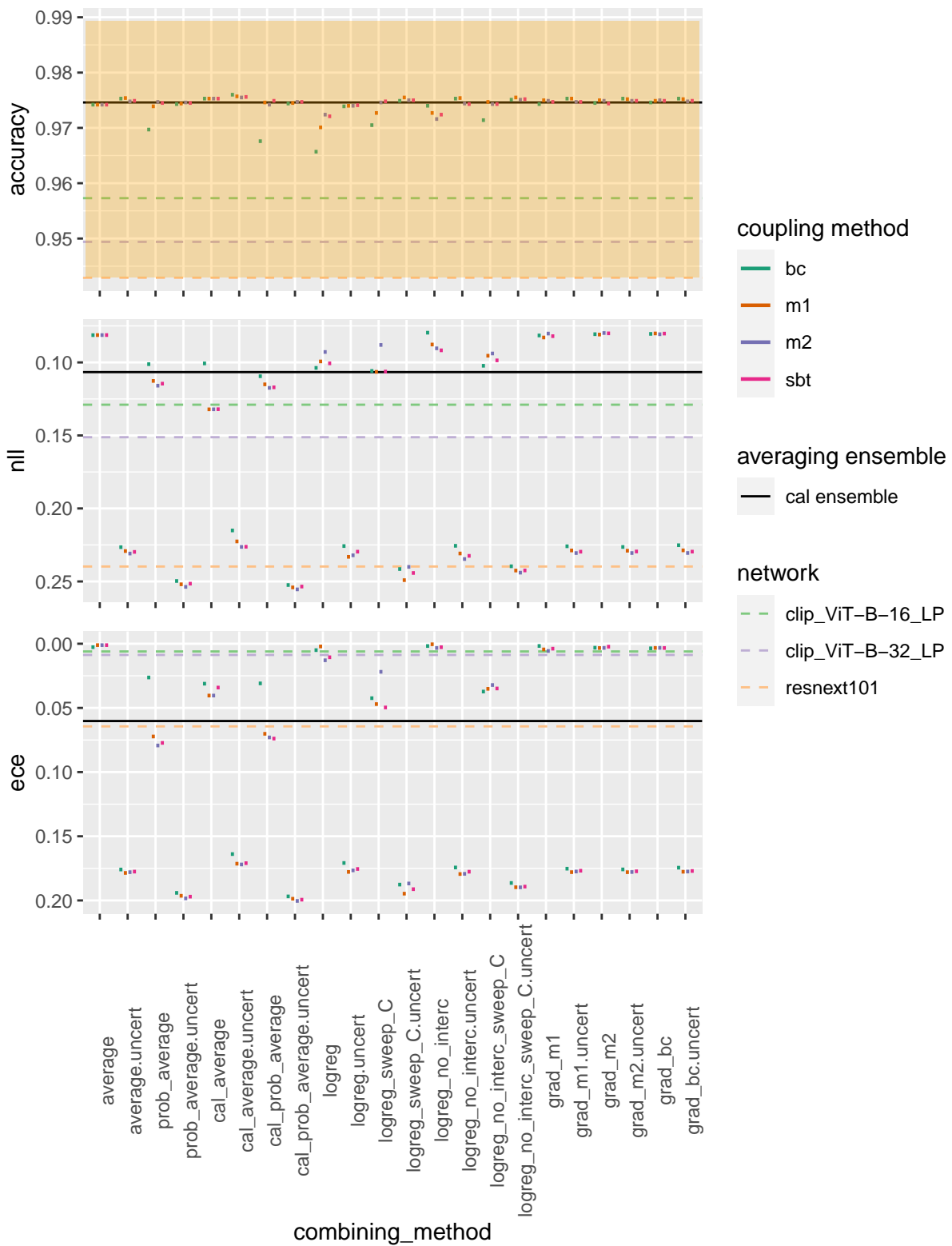
Average pairwise accuracy variance 5.7266652220278e-06



Ensemble metrics

Error inconsistency 0.0965999960899353

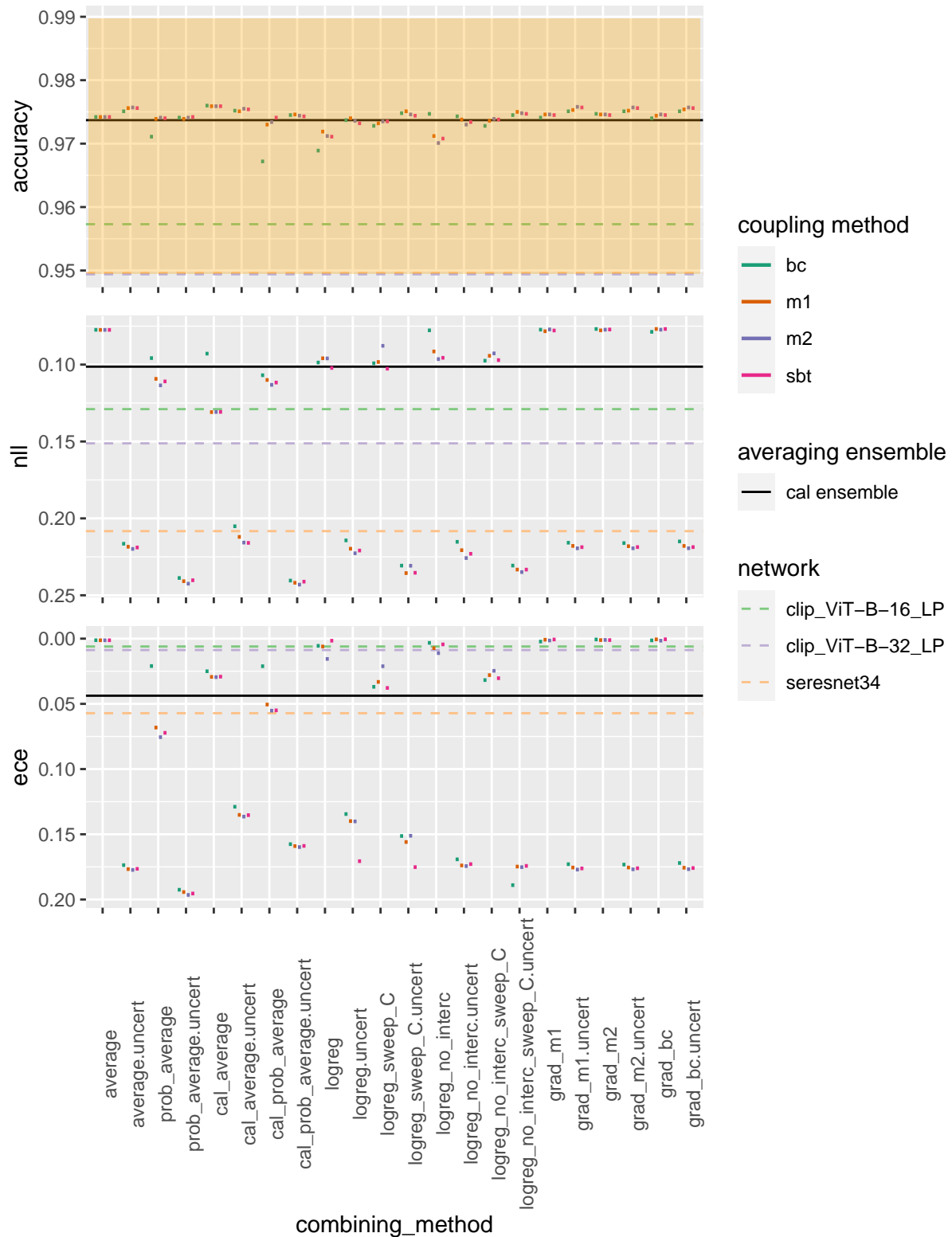
Average pairwise accuracy variance 6.51888058200711e-06



Ensemble metrics

Error inconsistency 0.0910999998450279

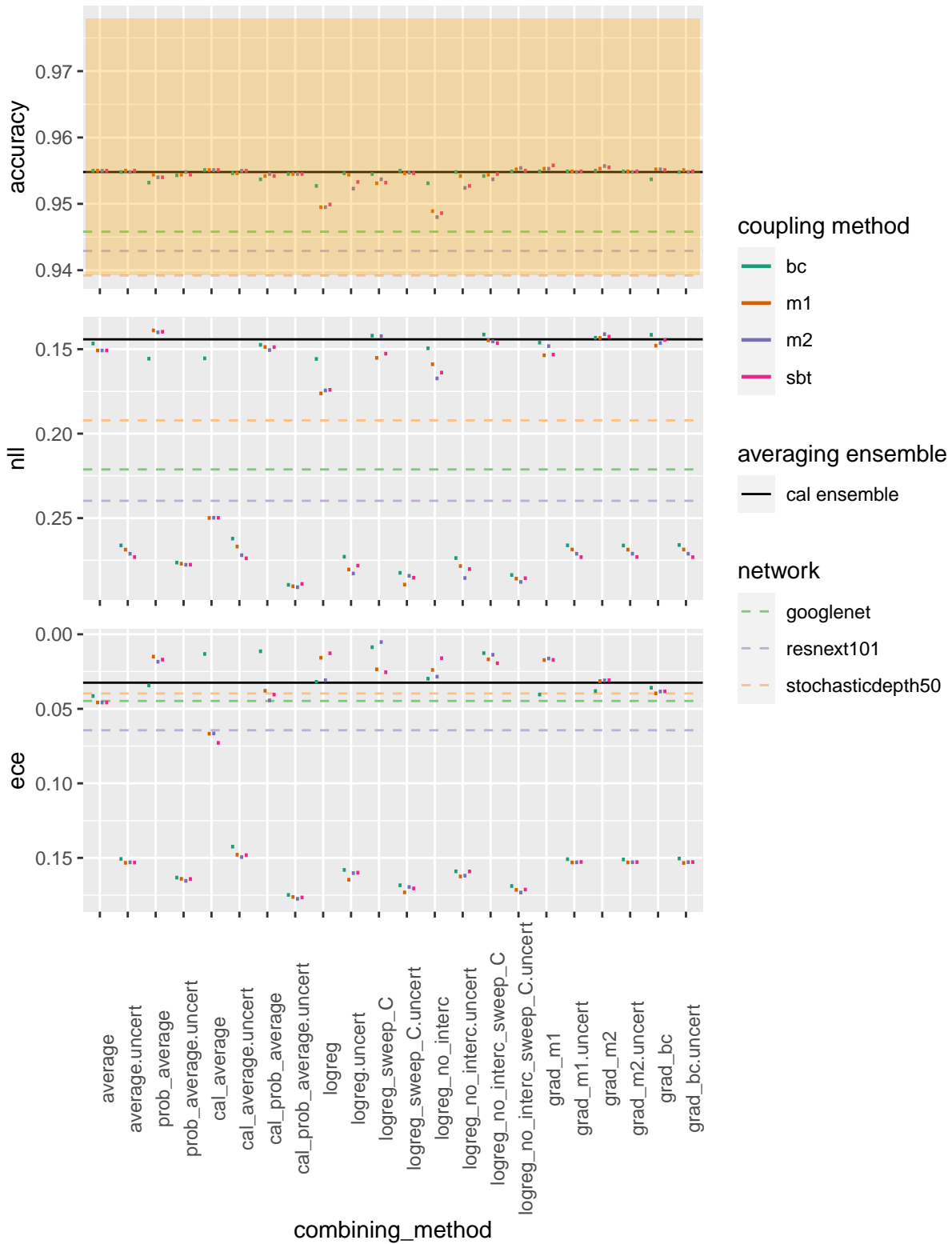
Average pairwise accuracy variance 4.34888761446928e-06



Ensemble metrics

Error inconsistency 0.0781999975442886

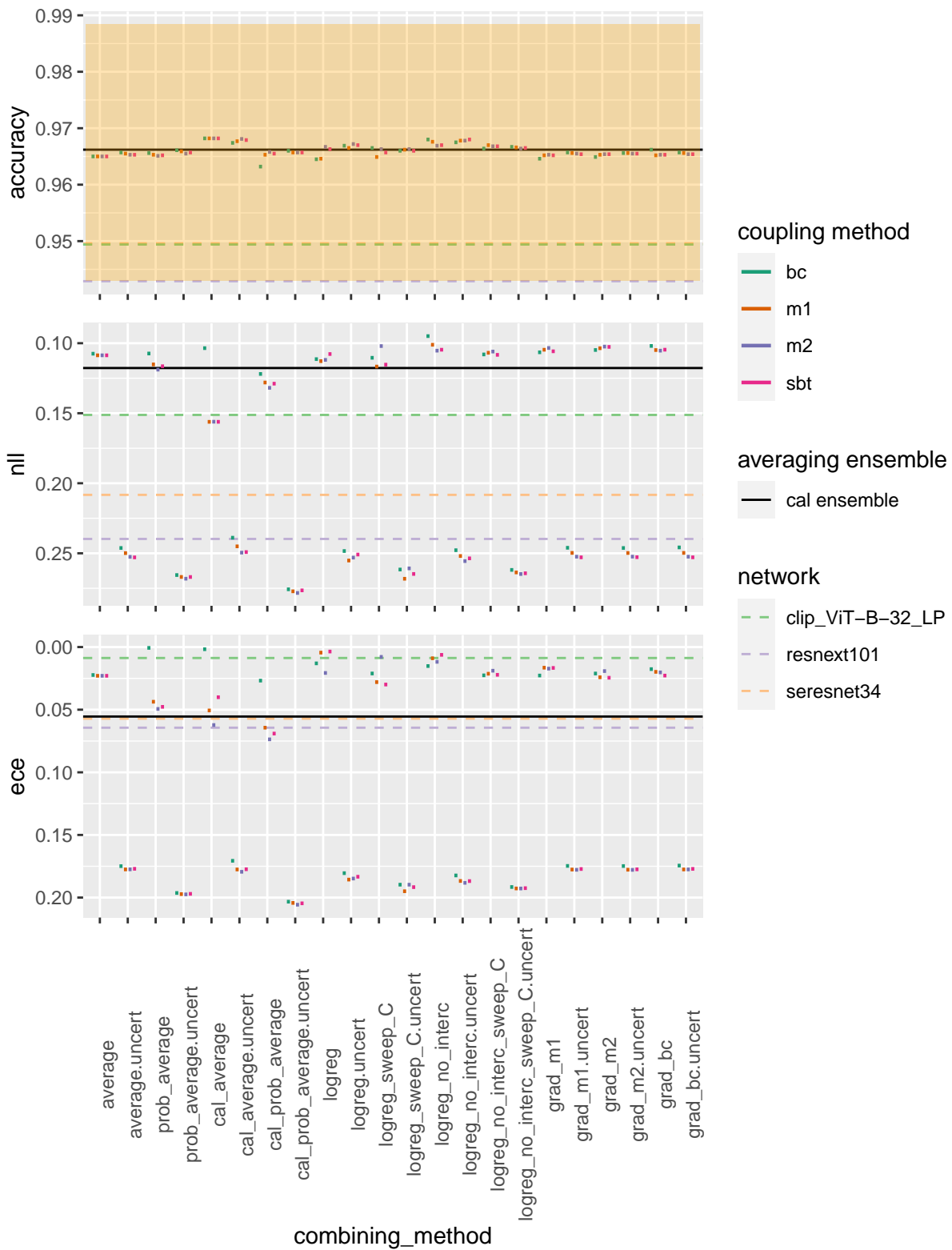
Average pairwise accuracy variance $1.79889696028113e-06$



Ensemble metrics

Error inconsistency 0.0960999950766563

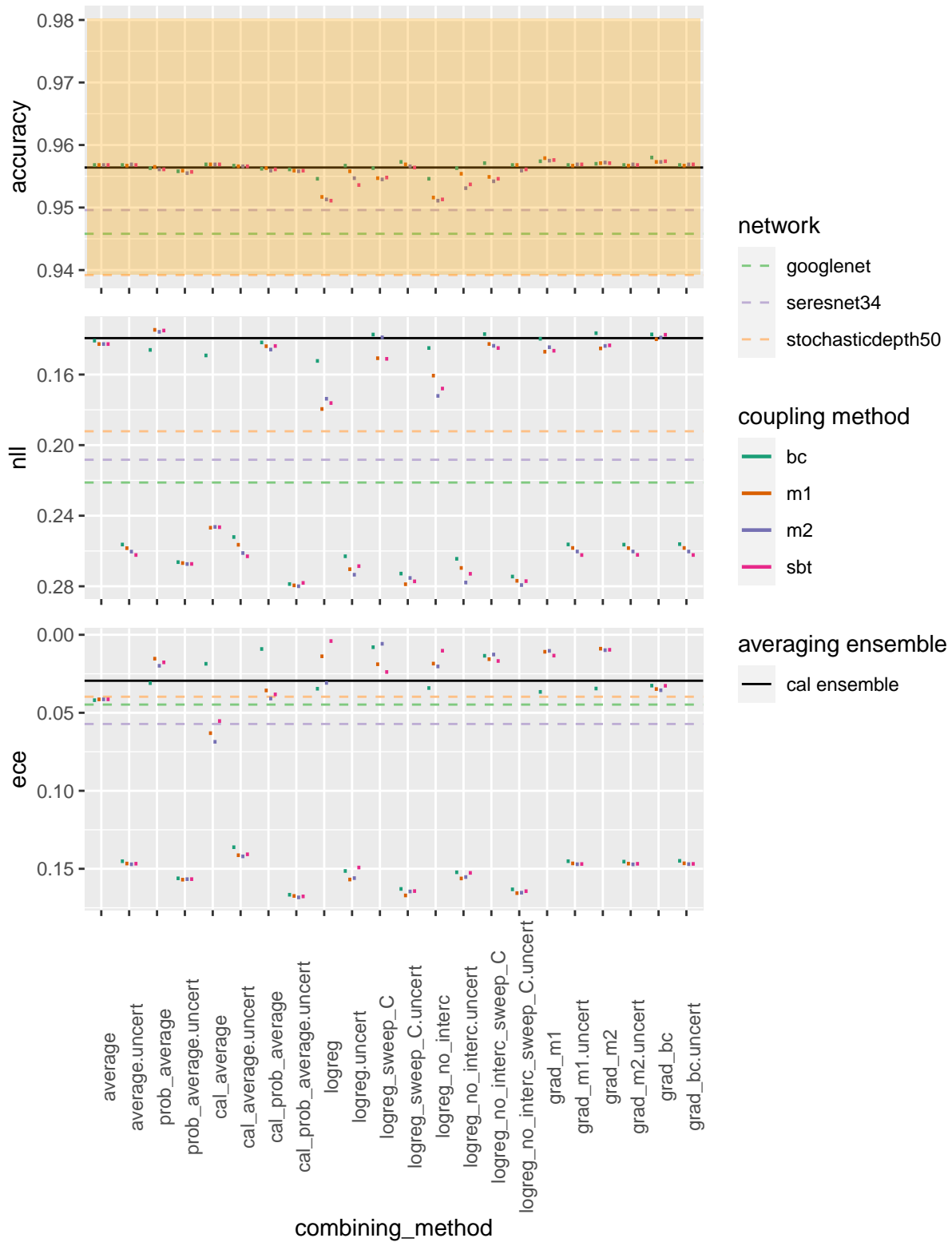
Average pairwise accuracy variance 4.45443765784148e-06



Ensemble metrics

Error inconsistency 0.0776999965310097

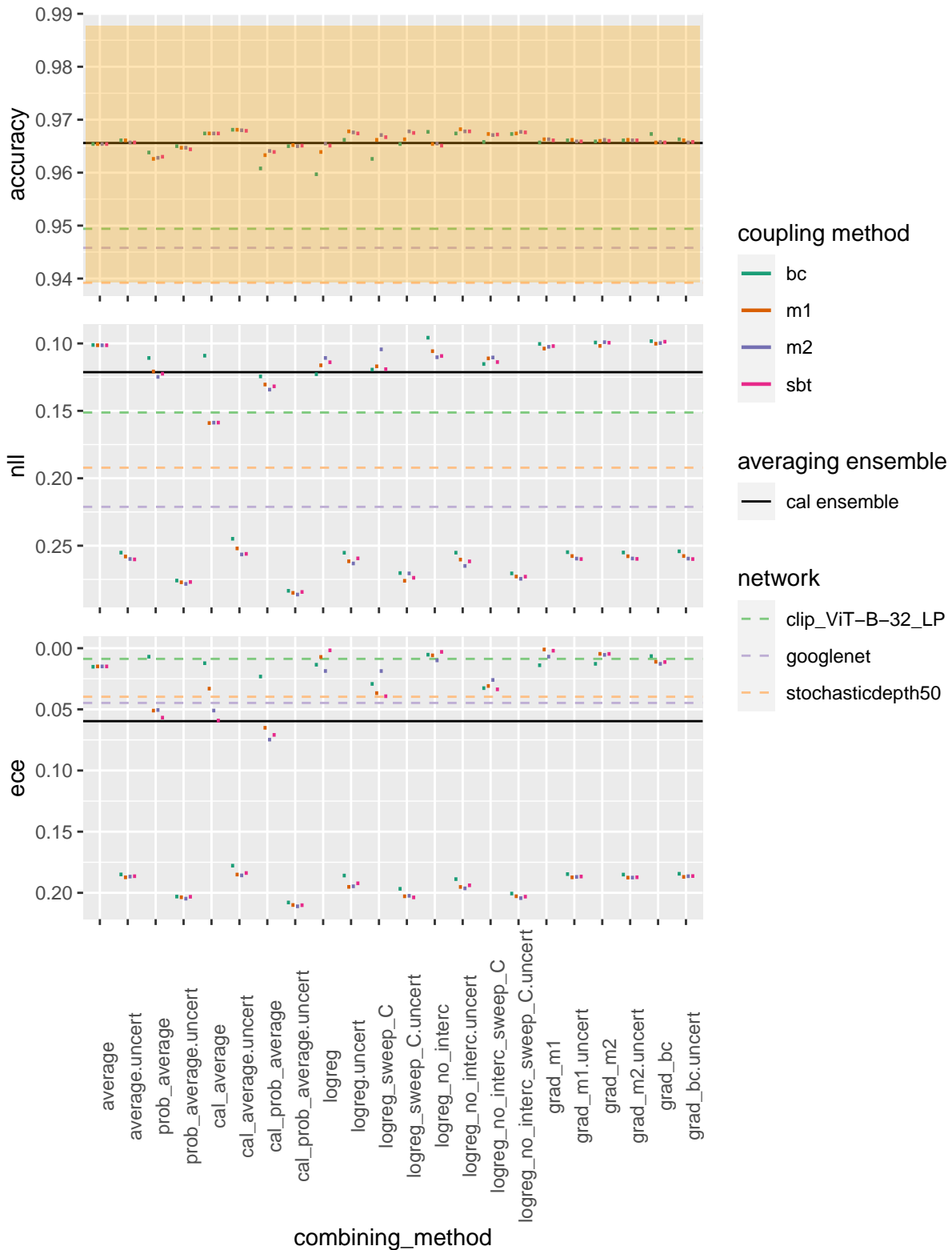
Average pairwise accuracy variance 2.38222264670185e-06



Ensemble metrics

Error inconsistency 0.0982999950647354

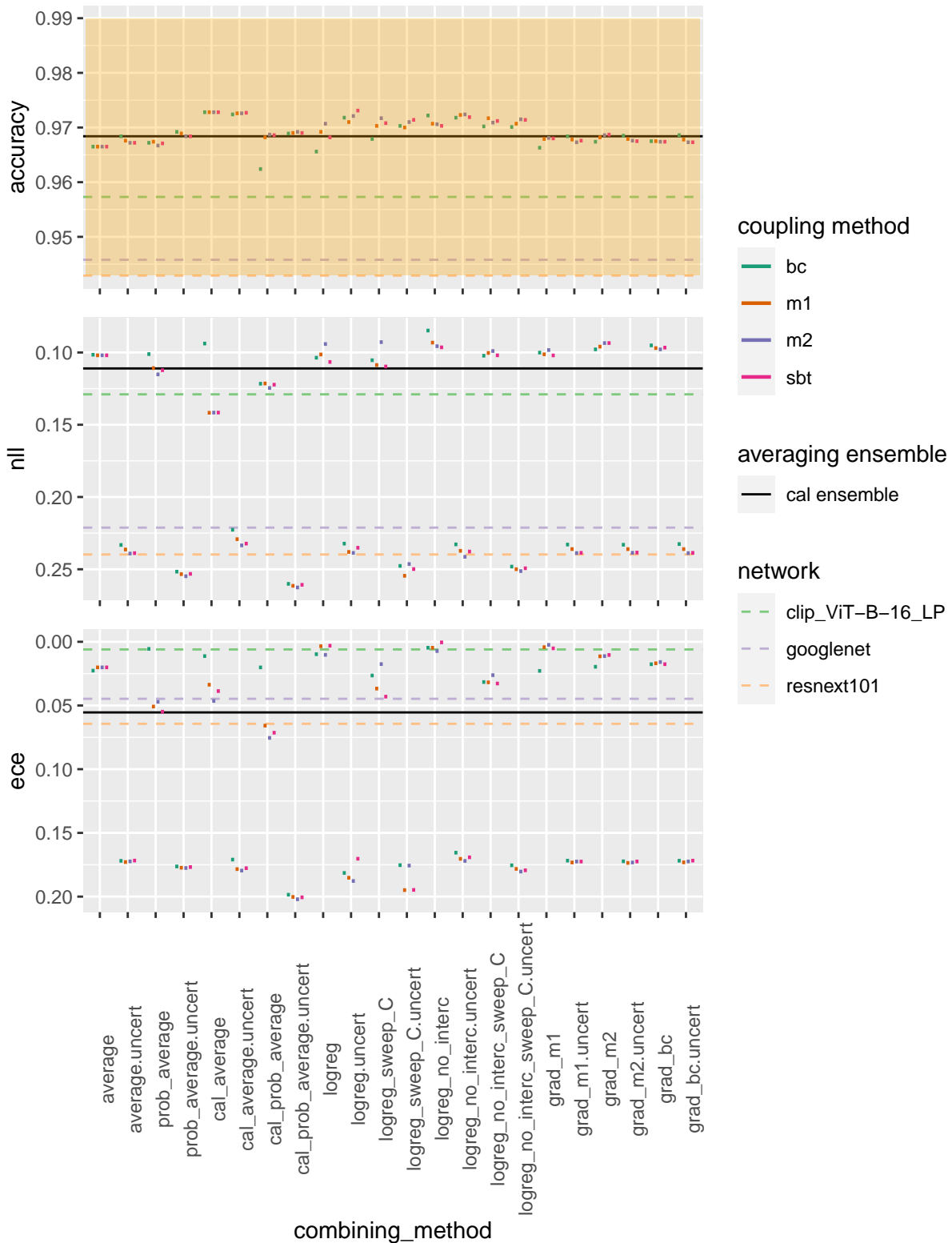
Average pairwise accuracy variance 5.65778691452579e-06



Ensemble metrics

Error inconsistency 0.094700001180172

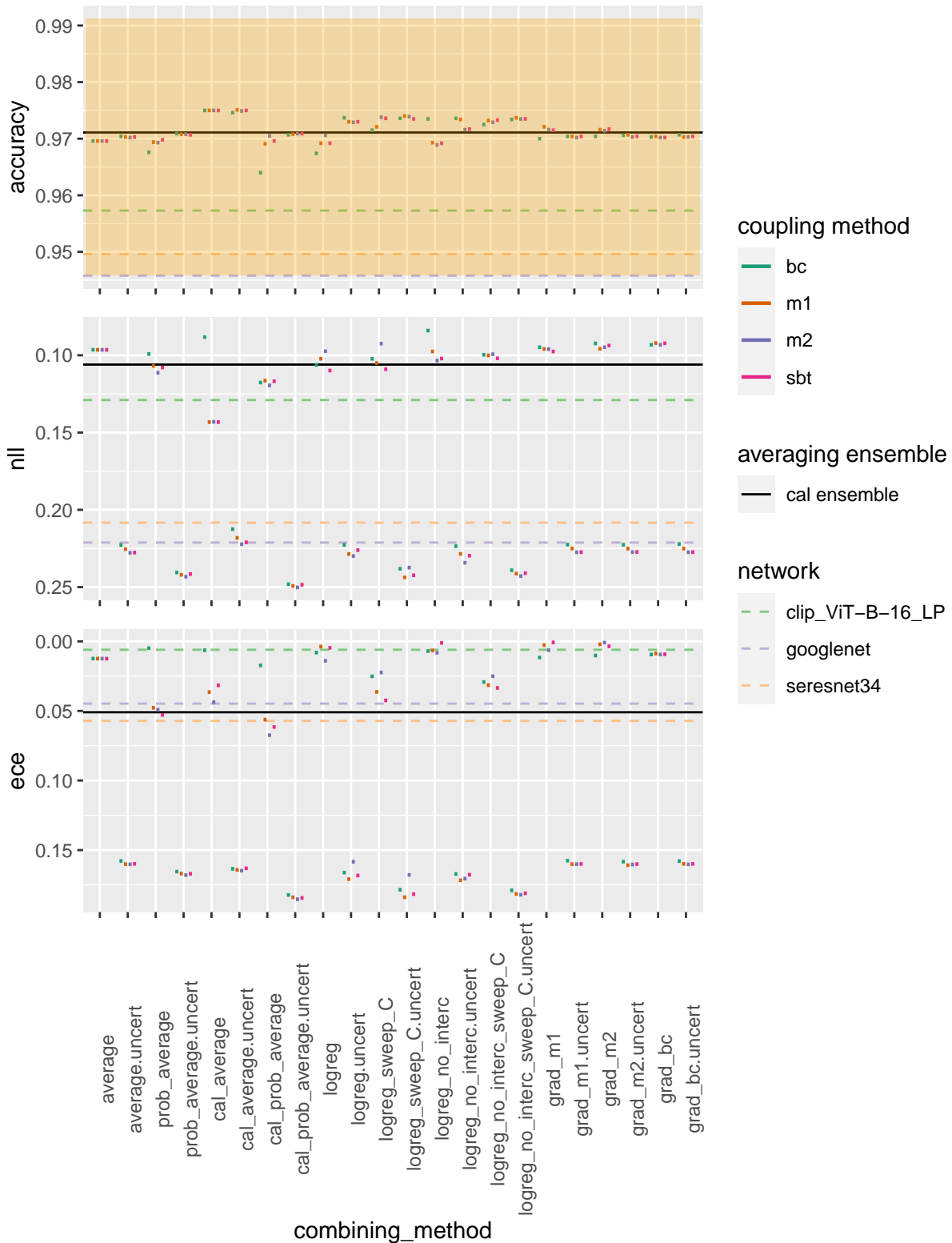
Average pairwise accuracy variance 6.18774447502801e-06



Ensemble metrics

Error inconsistency 0.092399999499321

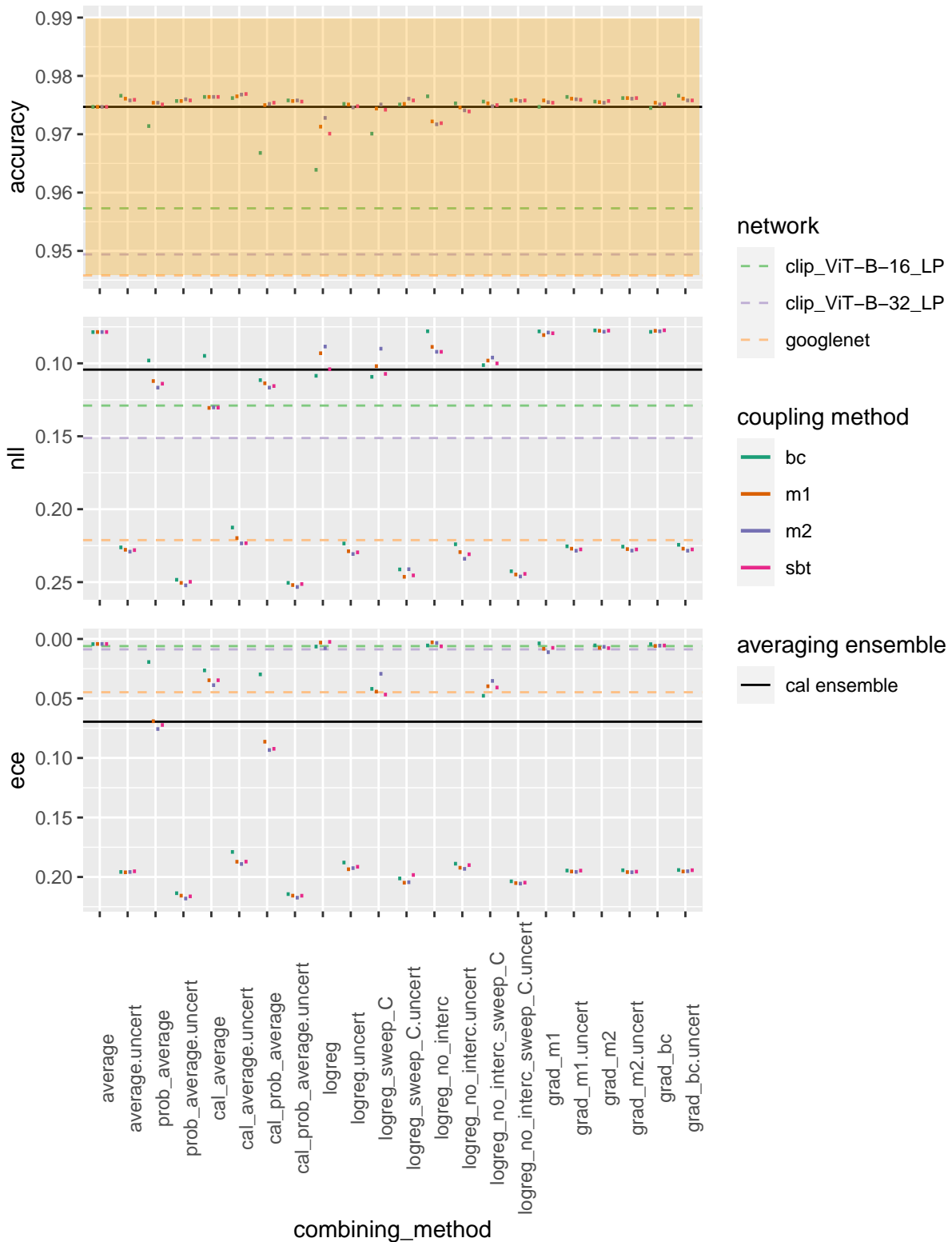
Average pairwise accuracy variance 4.70888926429325e-06



Ensemble metrics

Error inconsistency 0.0953999981284142

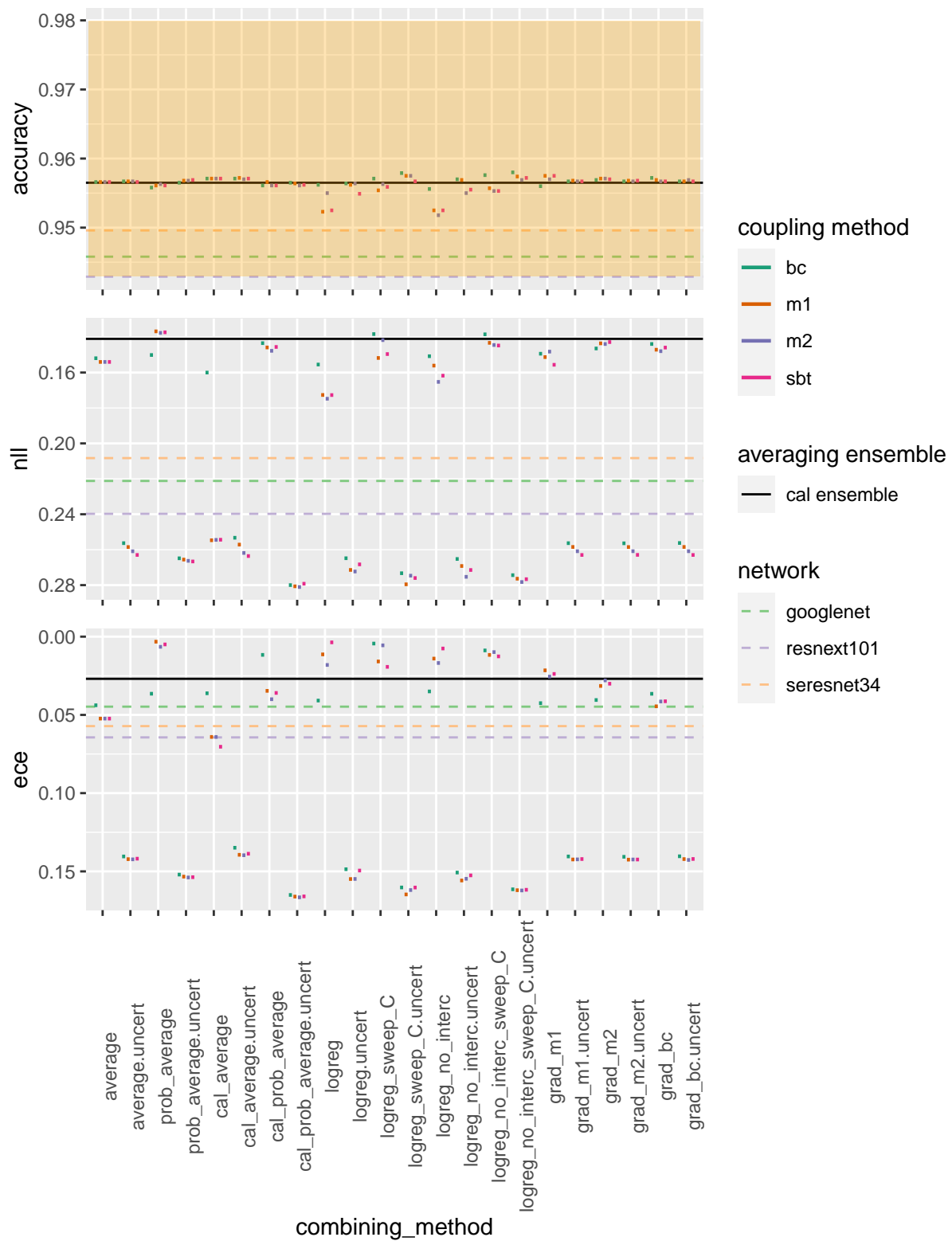
Average pairwise accuracy variance 5.60332182431011e-06



Ensemble metrics

Error inconsistency 0.0736999958753586

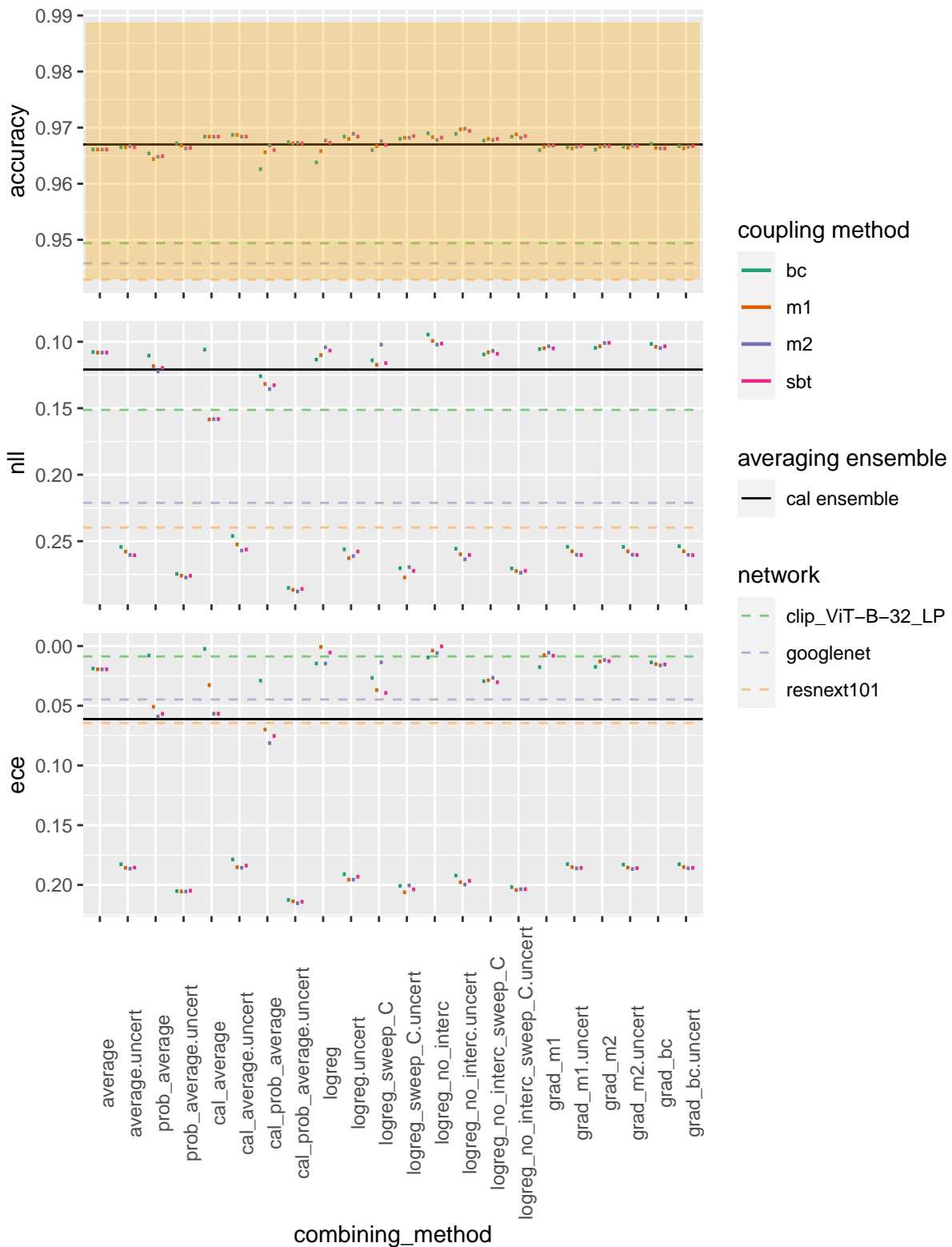
Average pairwise accuracy variance 1.78777349901793e-06



Ensemble metrics

Error inconsistency 0.0988999977707863

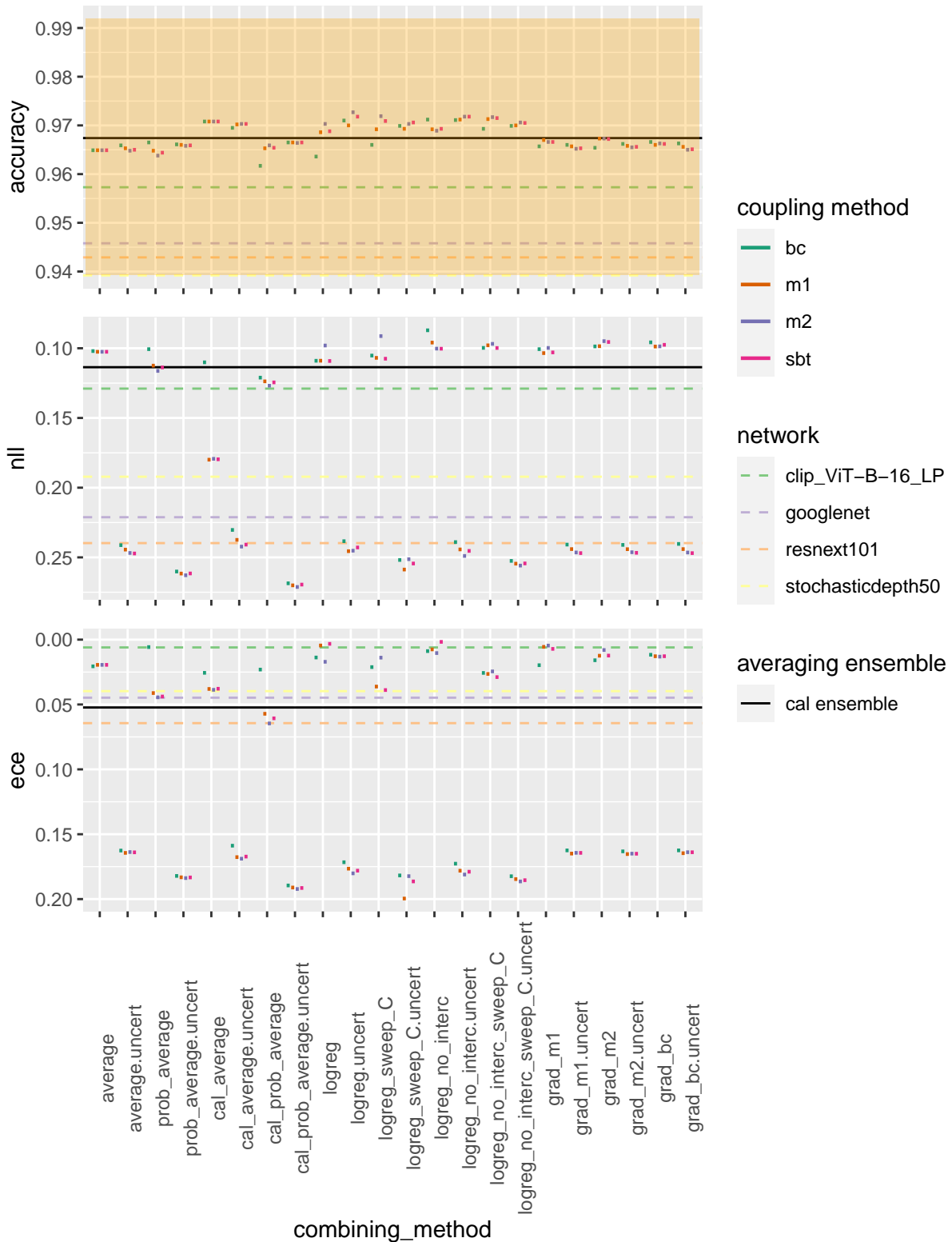
Average pairwise accuracy variance 4.74222088087117e-06



Ensemble metrics

Error inconsistency 0.112799994647503

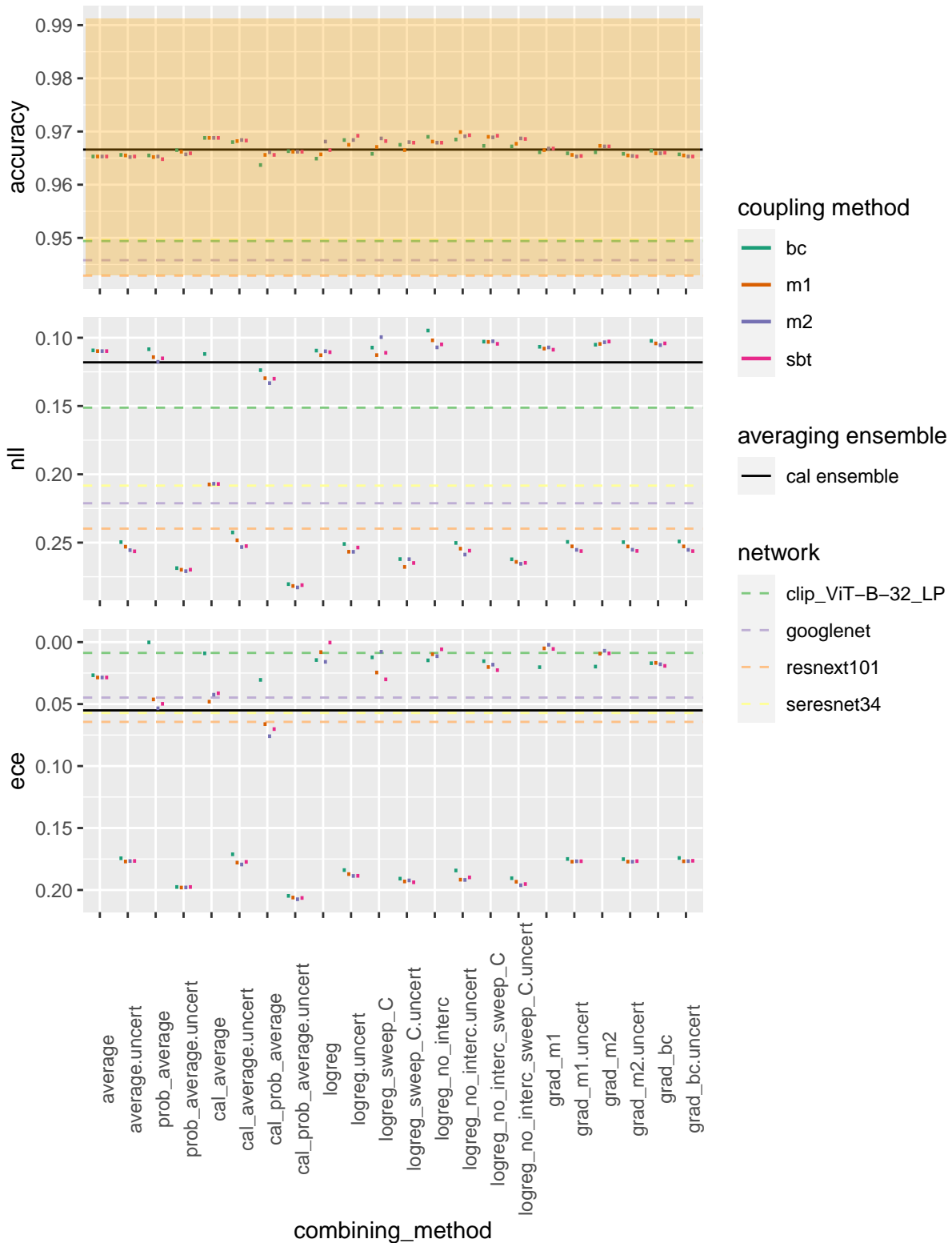
Average pairwise accuracy variance 6.3603065427742e-06



Ensemble metrics

Error inconsistency 0.111699998378754

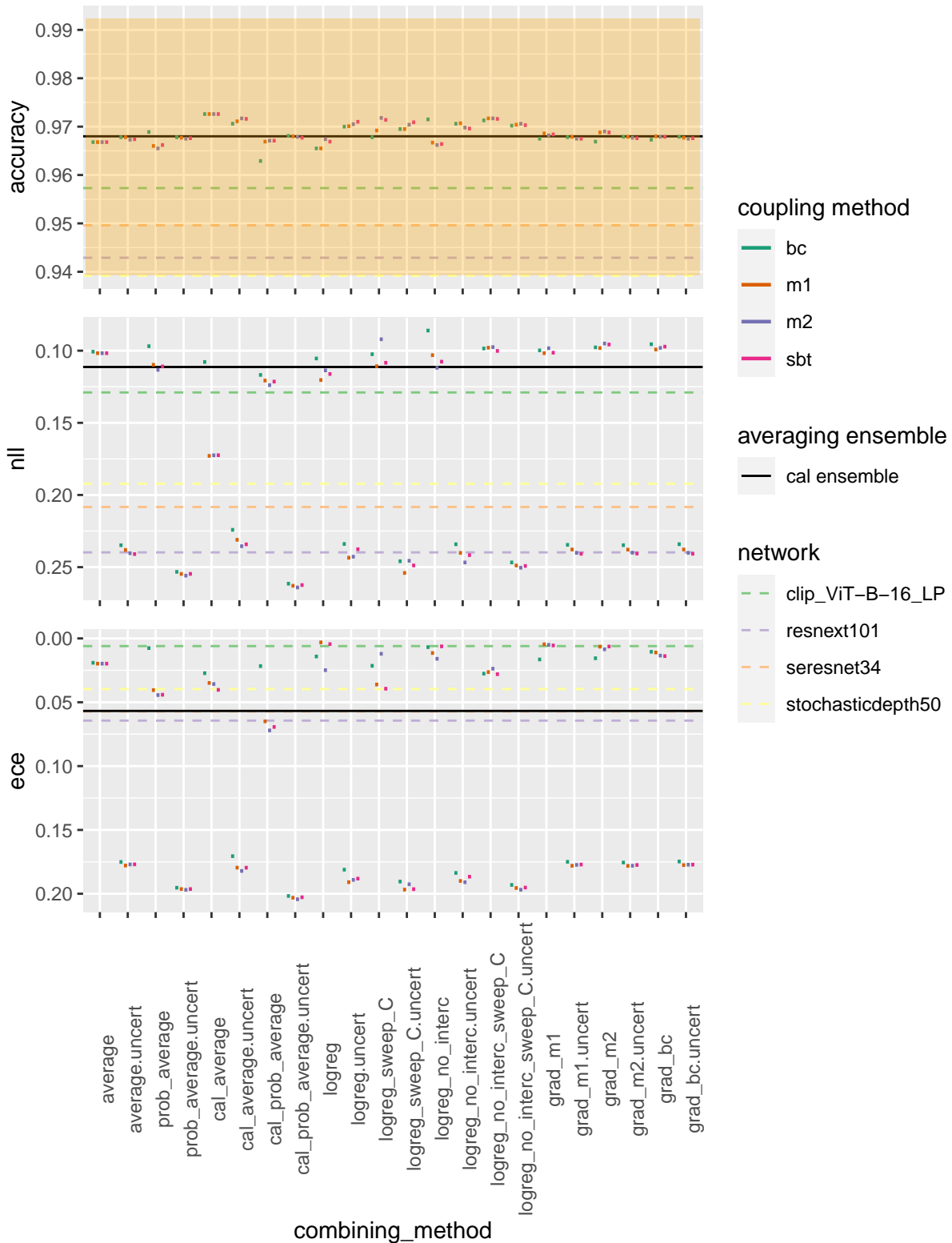
Average pairwise accuracy variance 4.2221904550388e-06



Ensemble metrics

Error inconsistency 0.111999996006489

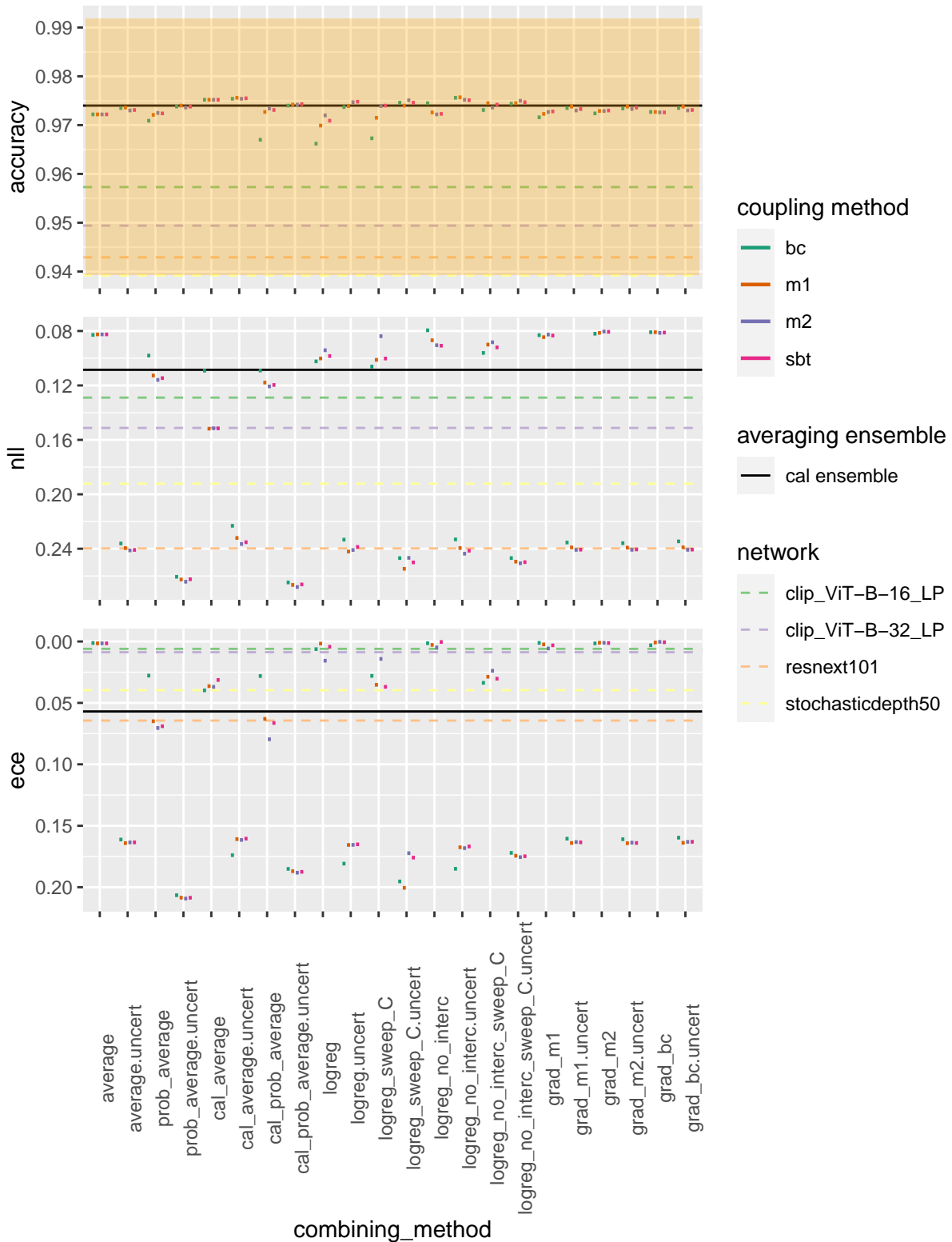
Average pairwise accuracy variance 6.26751216259436e-06



Ensemble metrics

Error inconsistency 0.117899999022484

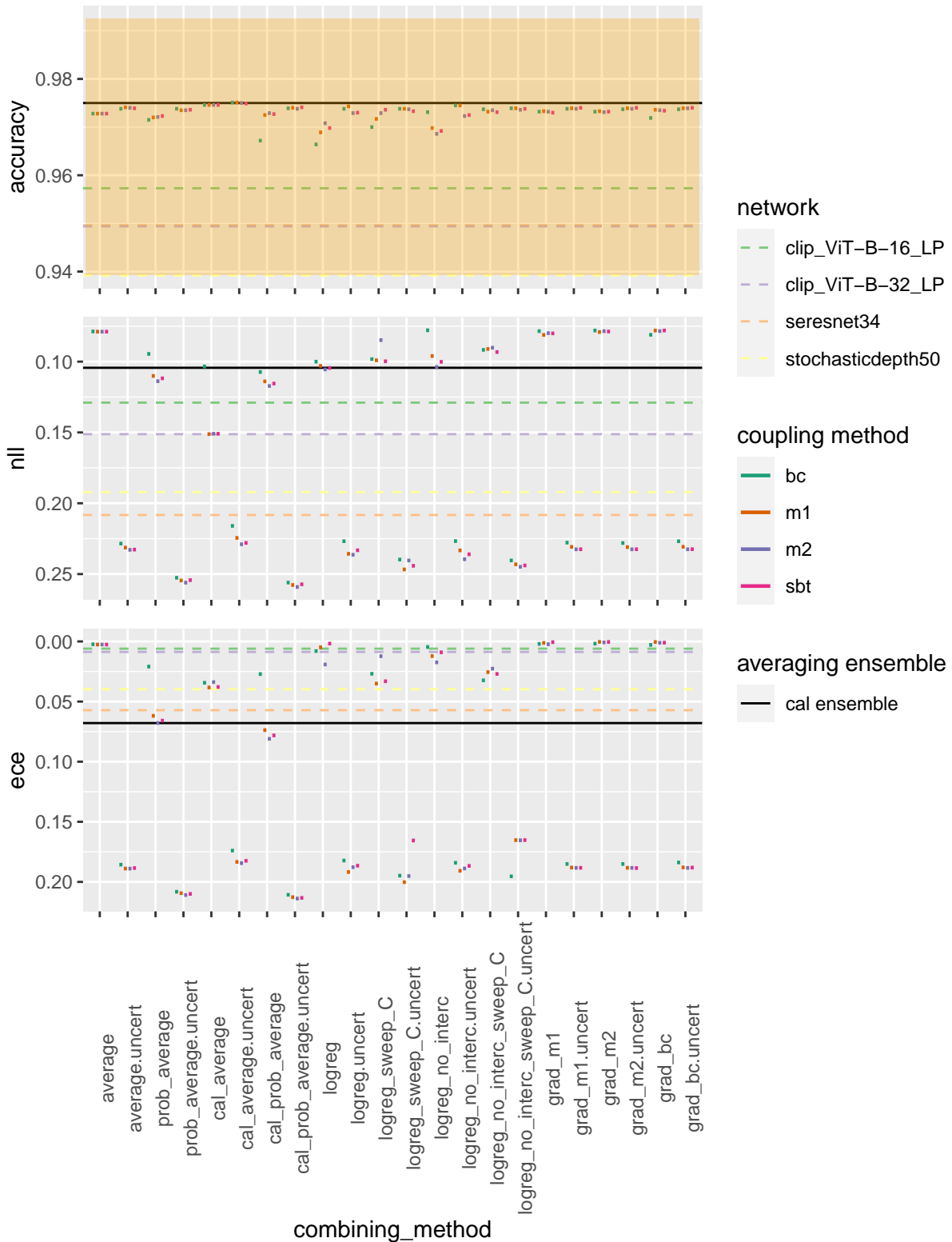
Average pairwise accuracy variance 7.69594407756813e-06



Ensemble metrics

Error inconsistency 0.115299999713898

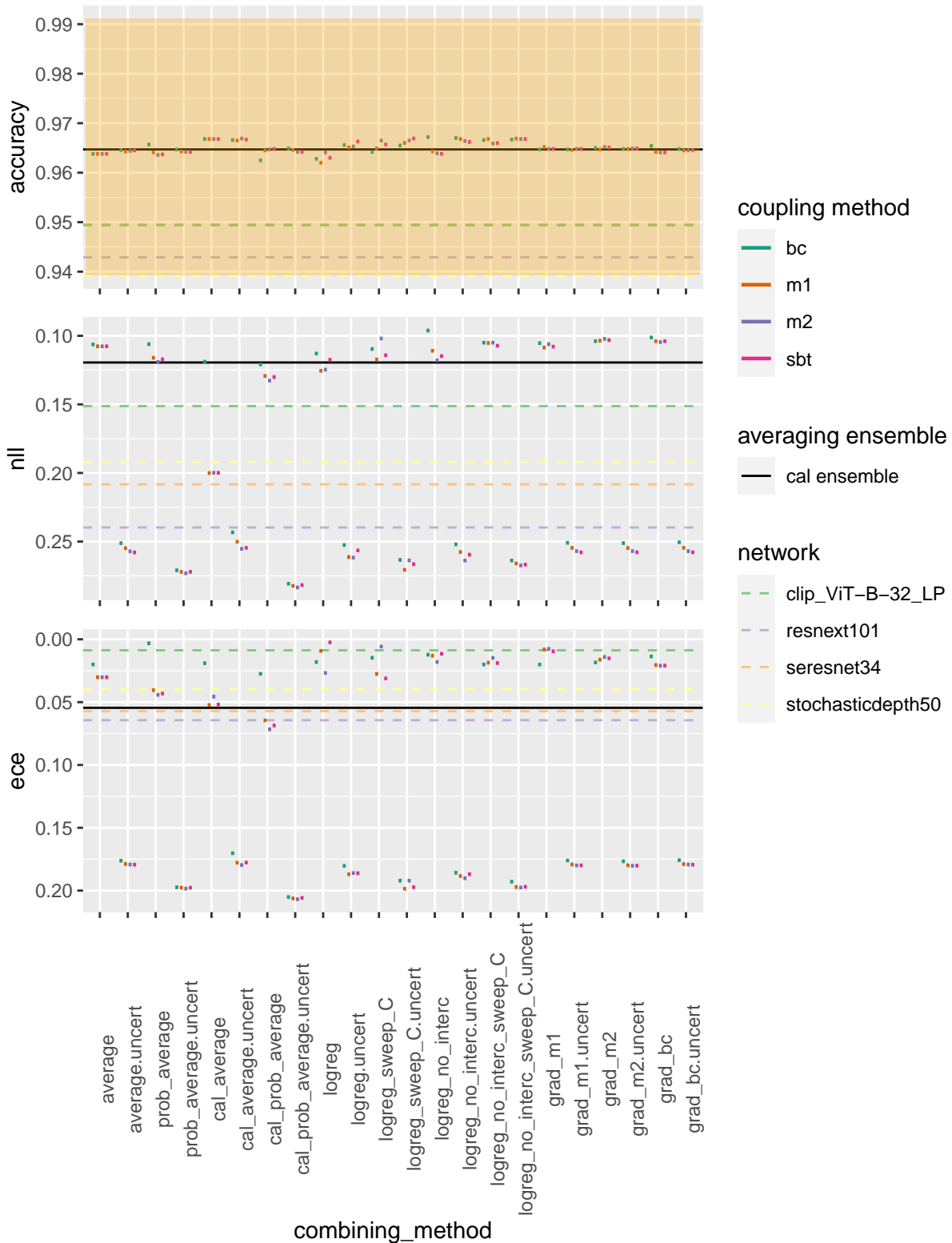
Average pairwise accuracy variance 6.81000392432907e-06



Ensemble metrics

Error inconsistency 0.115199998021126

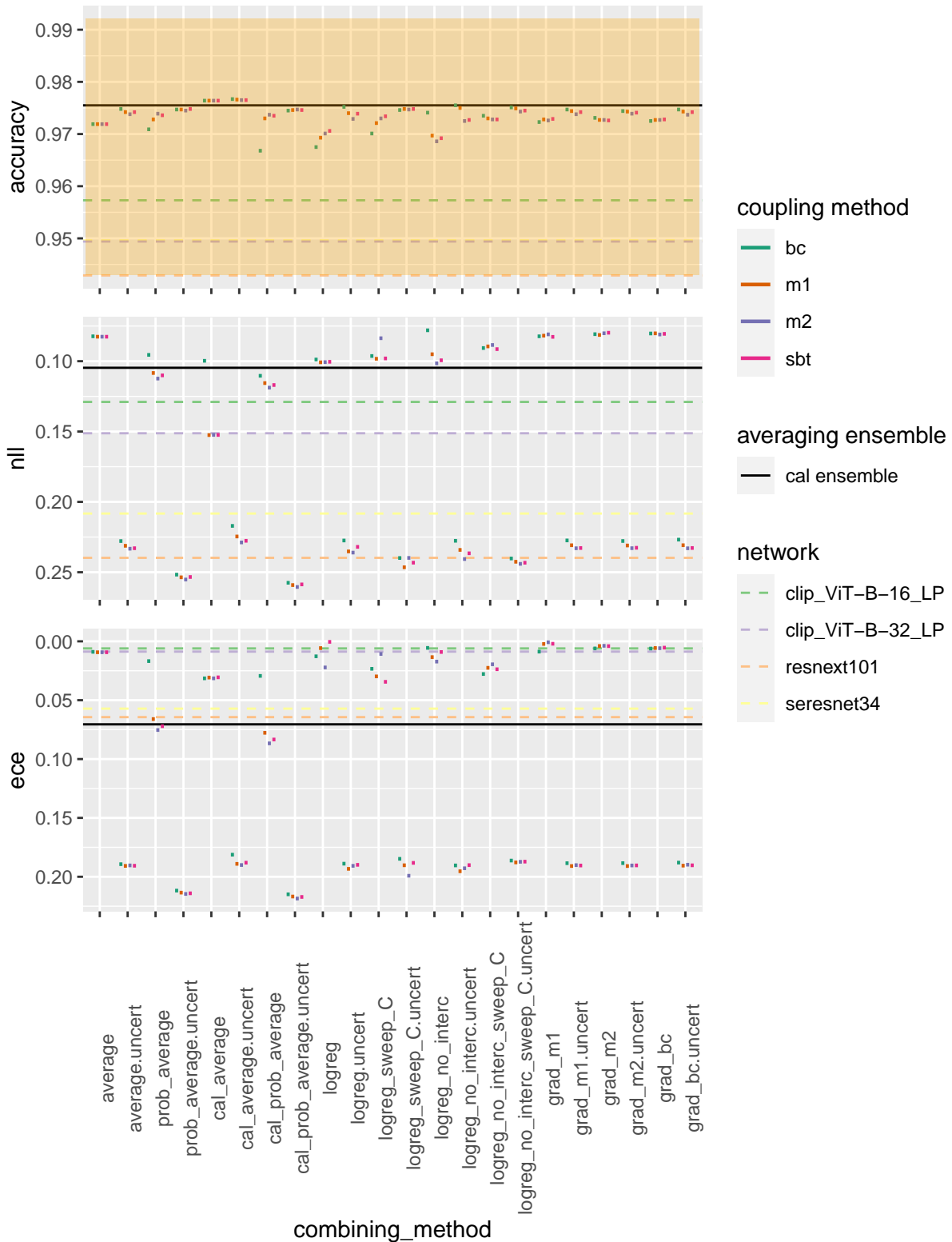
Average pairwise accuracy variance 4.98843292007223e-06



Ensemble metrics

Error inconsistency 0.114699997007847

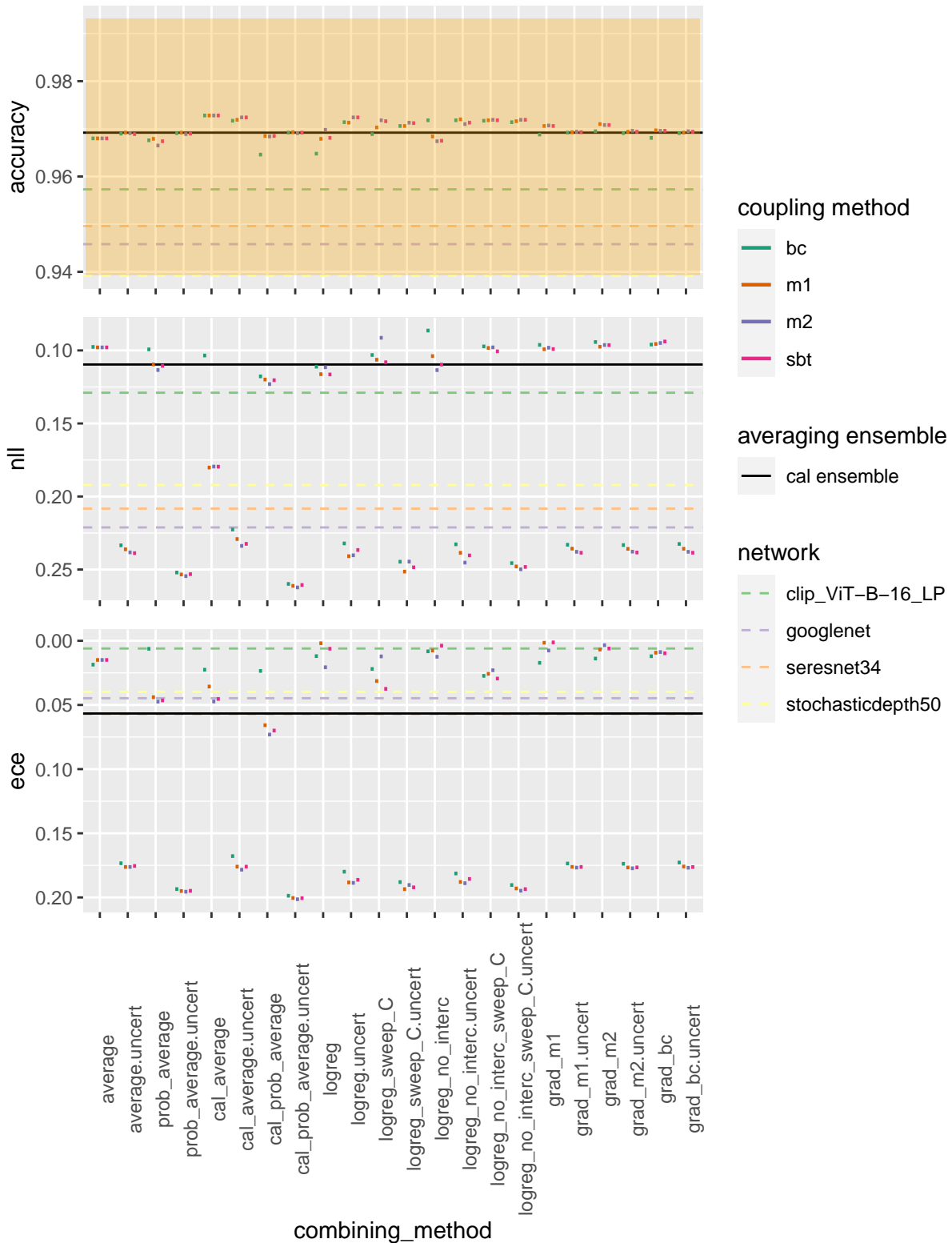
Average pairwise accuracy variance 5.91999651078368e-06



Ensemble metrics

Error inconsistency 0.11149999499321

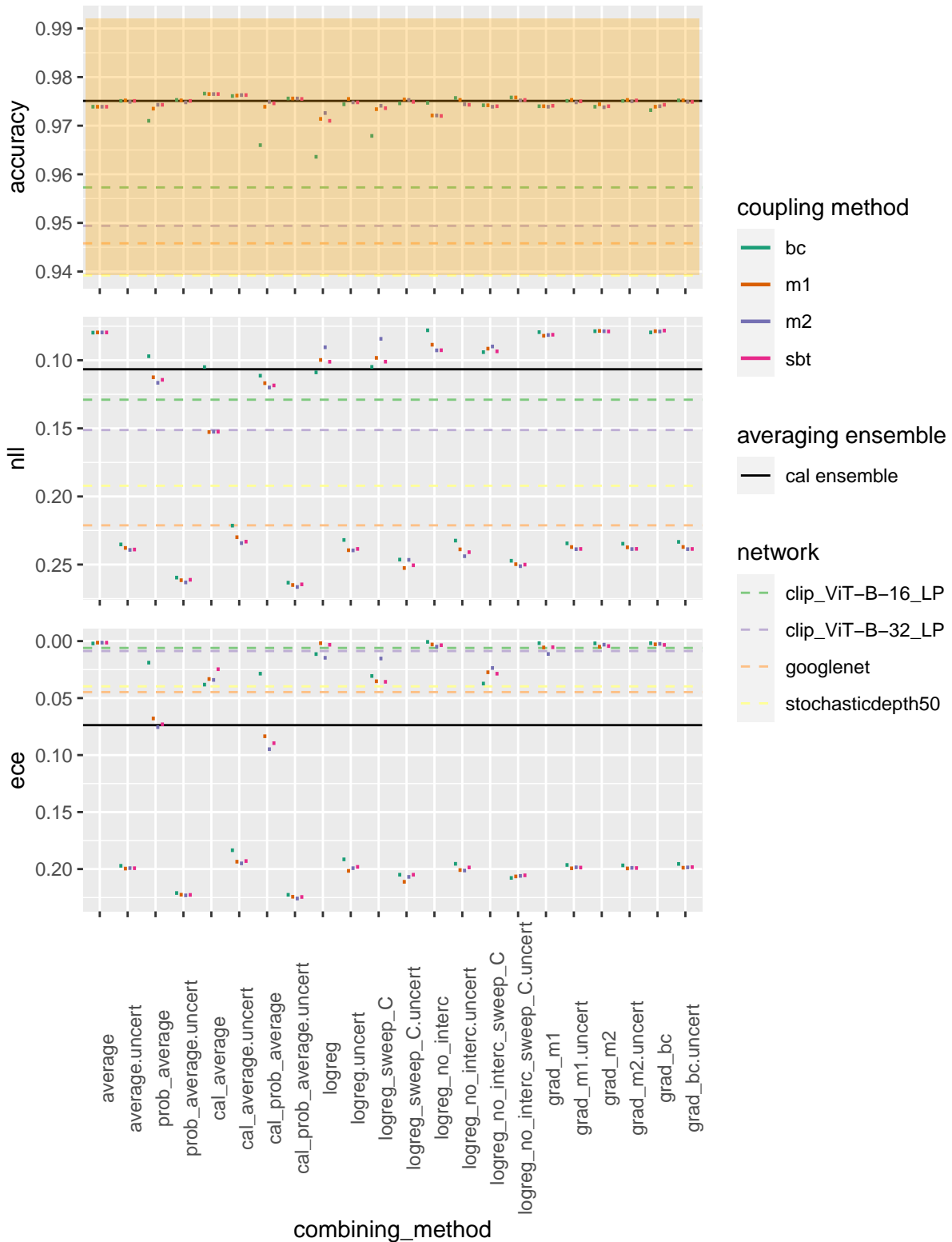
Average pairwise accuracy variance 5.86311534789274e-06



Ensemble metrics

Error inconsistency 0.116599999368191

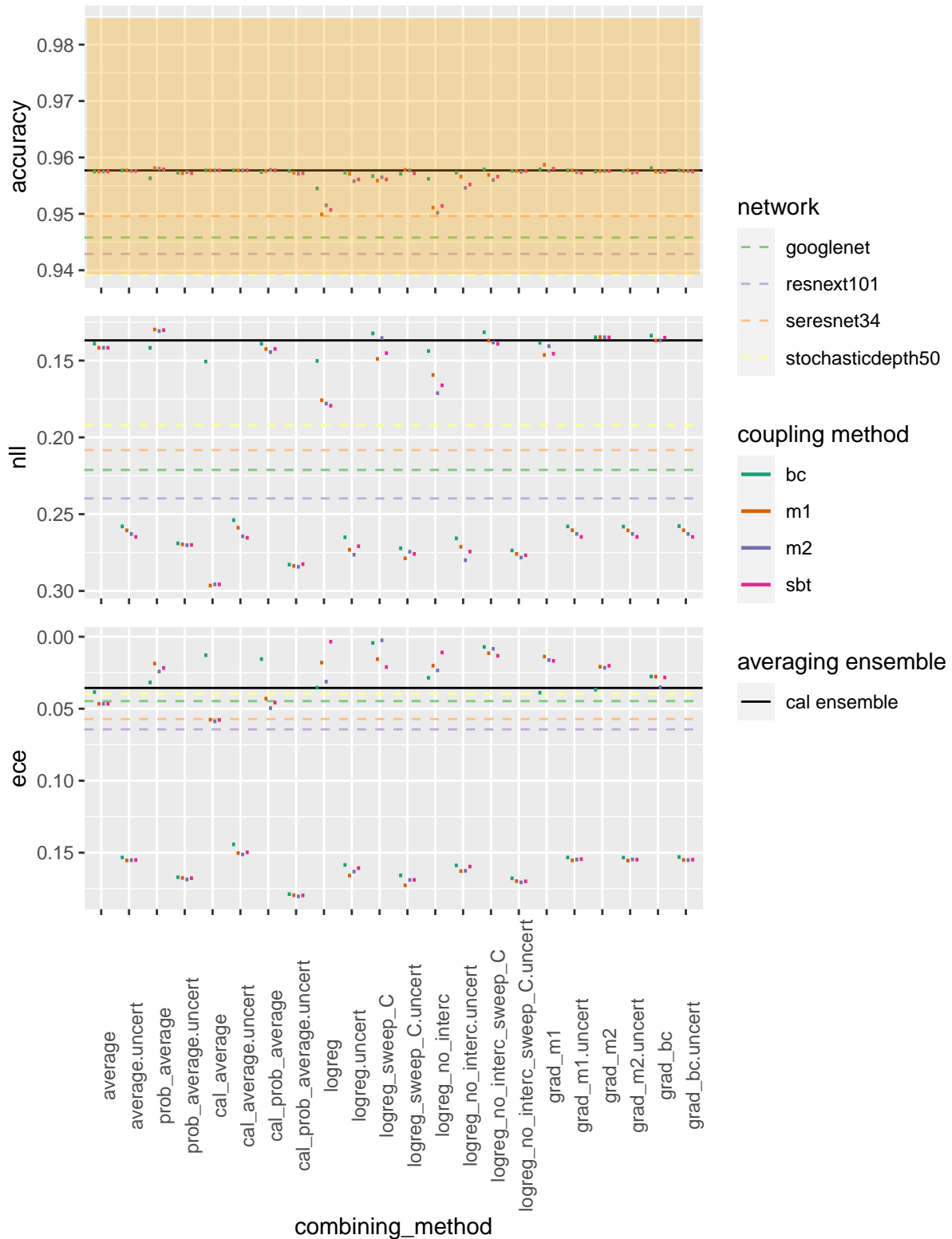
Average pairwise accuracy variance 7.34906052457518e-06



Ensemble metrics

Error inconsistency 0.0945999994874001

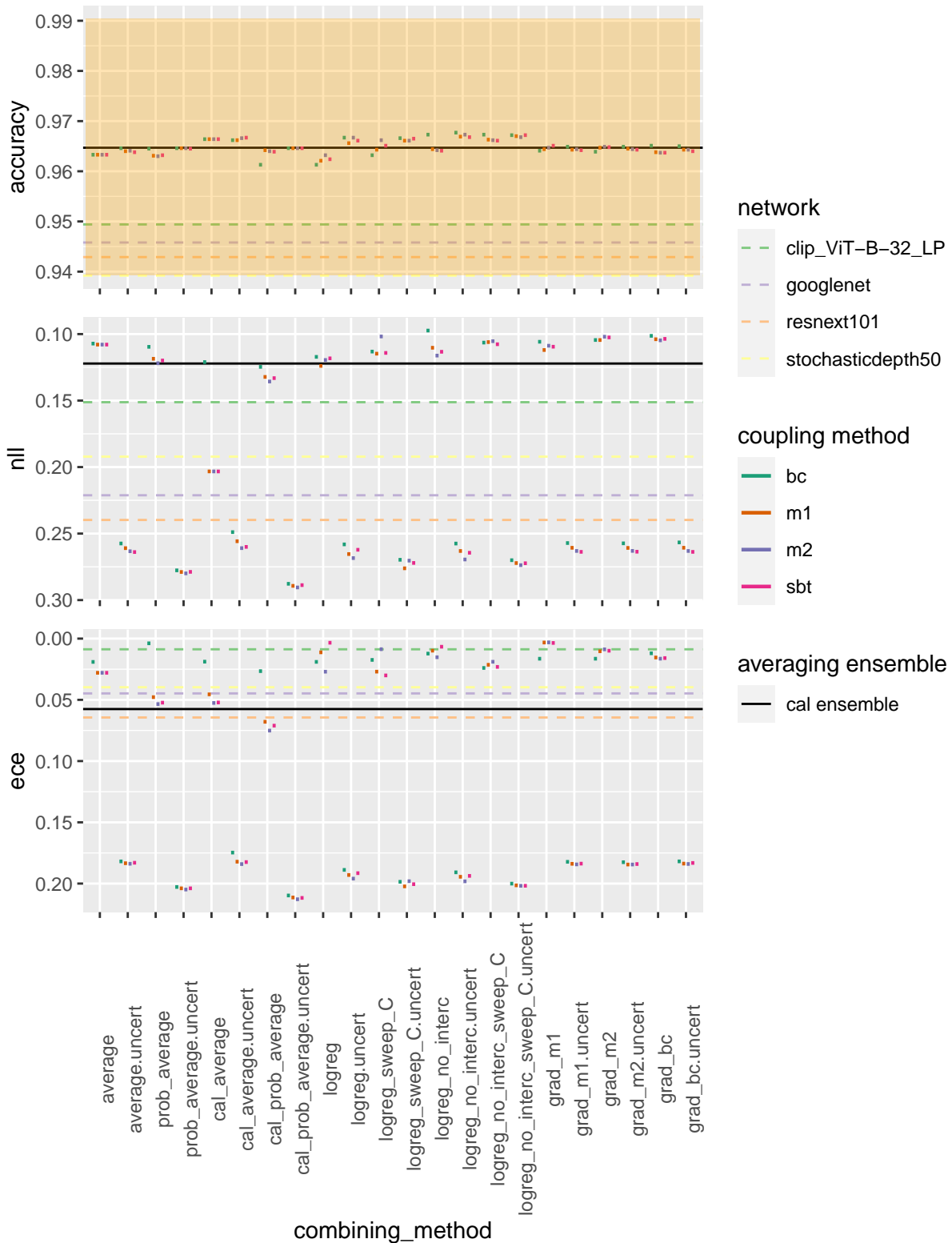
Average pairwise accuracy variance 2.33906143876084e-06



Ensemble metrics

Error inconsistency 0.115399993956089

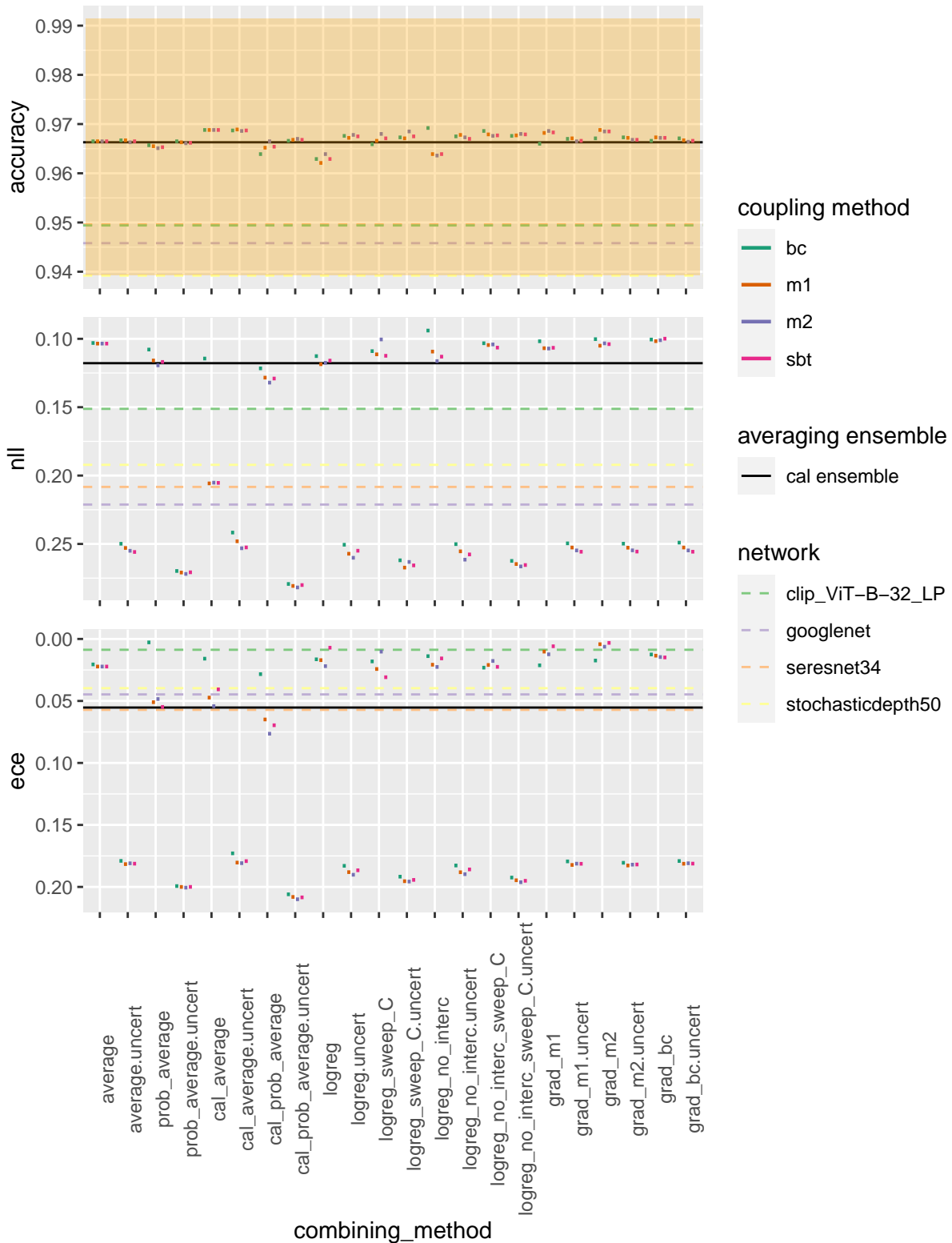
Average pairwise accuracy variance 4.98375038660015e-06



Ensemble metrics

Error inconsistency 0.113399997353554

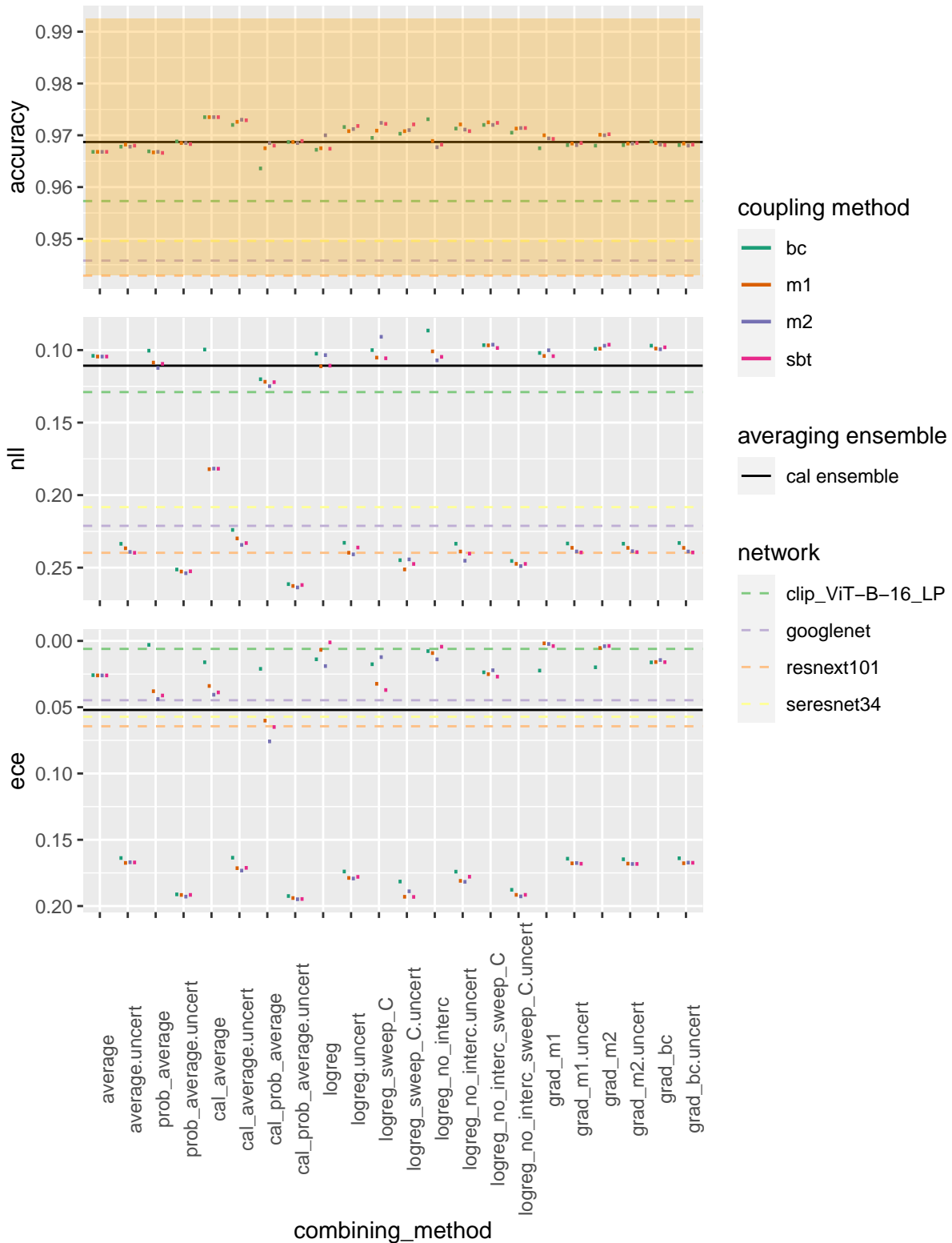
Average pairwise accuracy variance 4.91657146994839e-06



Ensemble metrics

Error inconsistency 0.108599998056889

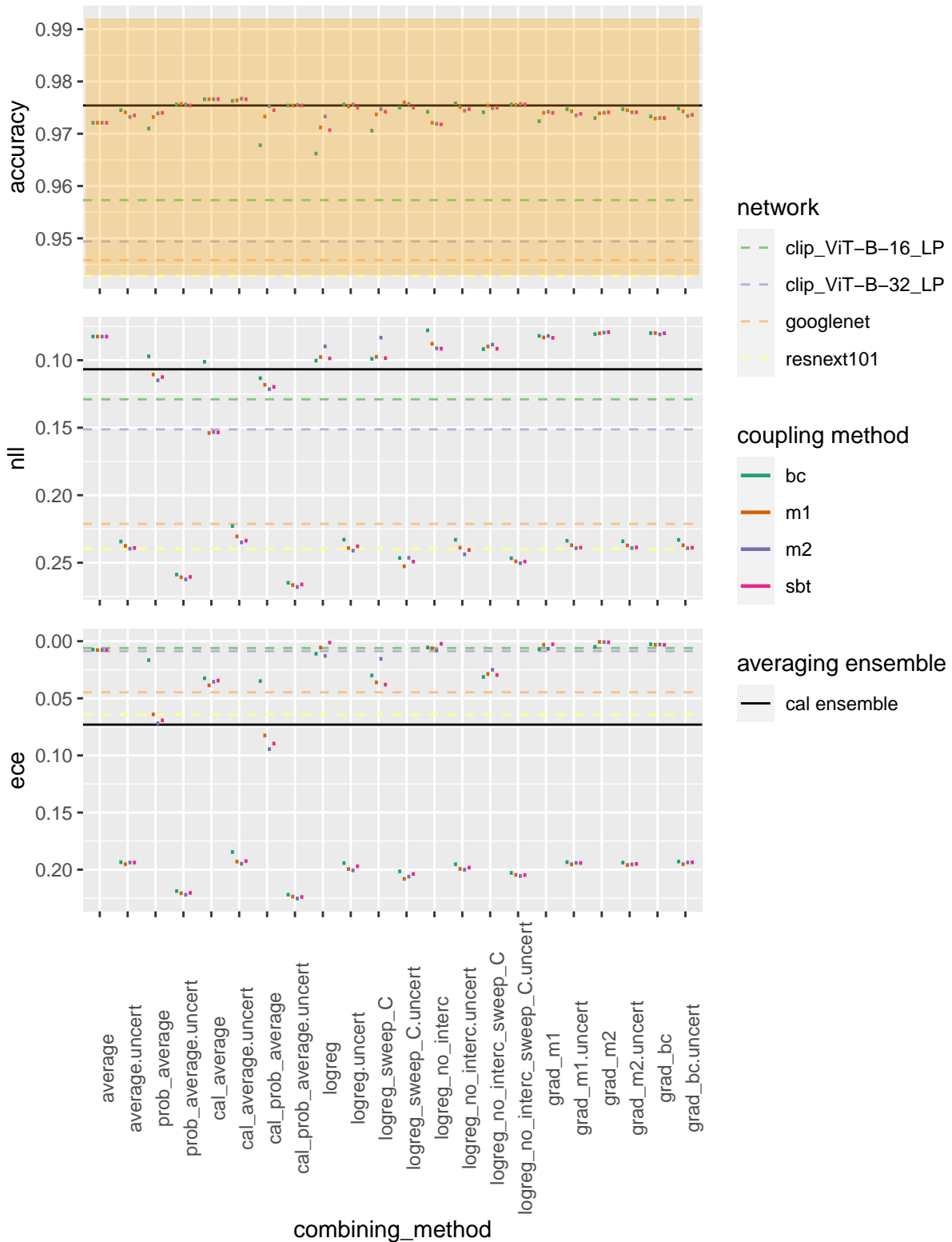
Average pairwise accuracy variance 5.17811304234783e-06



Ensemble metrics

Error inconsistency 0.116799995303154

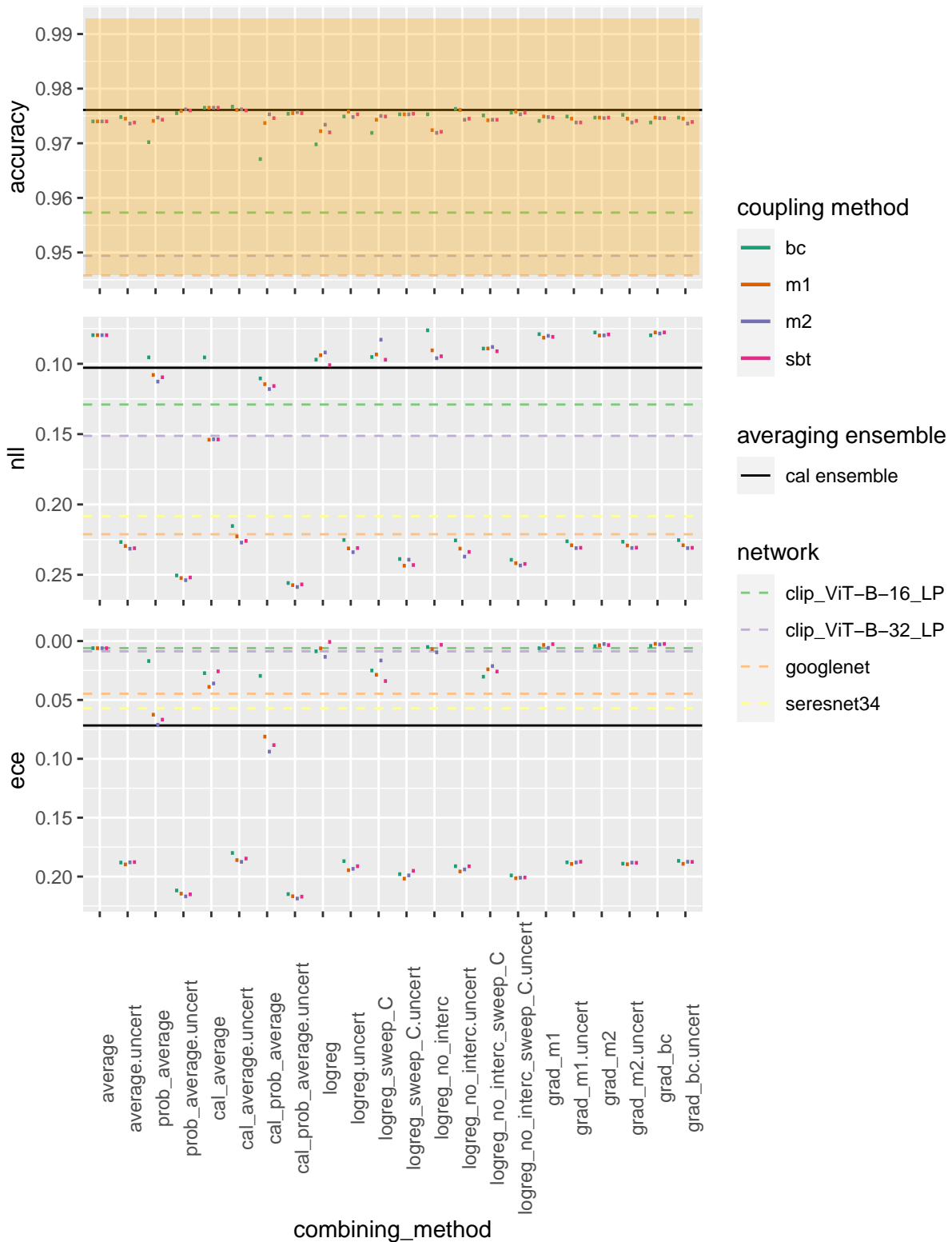
Average pairwise accuracy variance 6.48341710984823e-06



Ensemble metrics

Error inconsistency 0.113499999046326

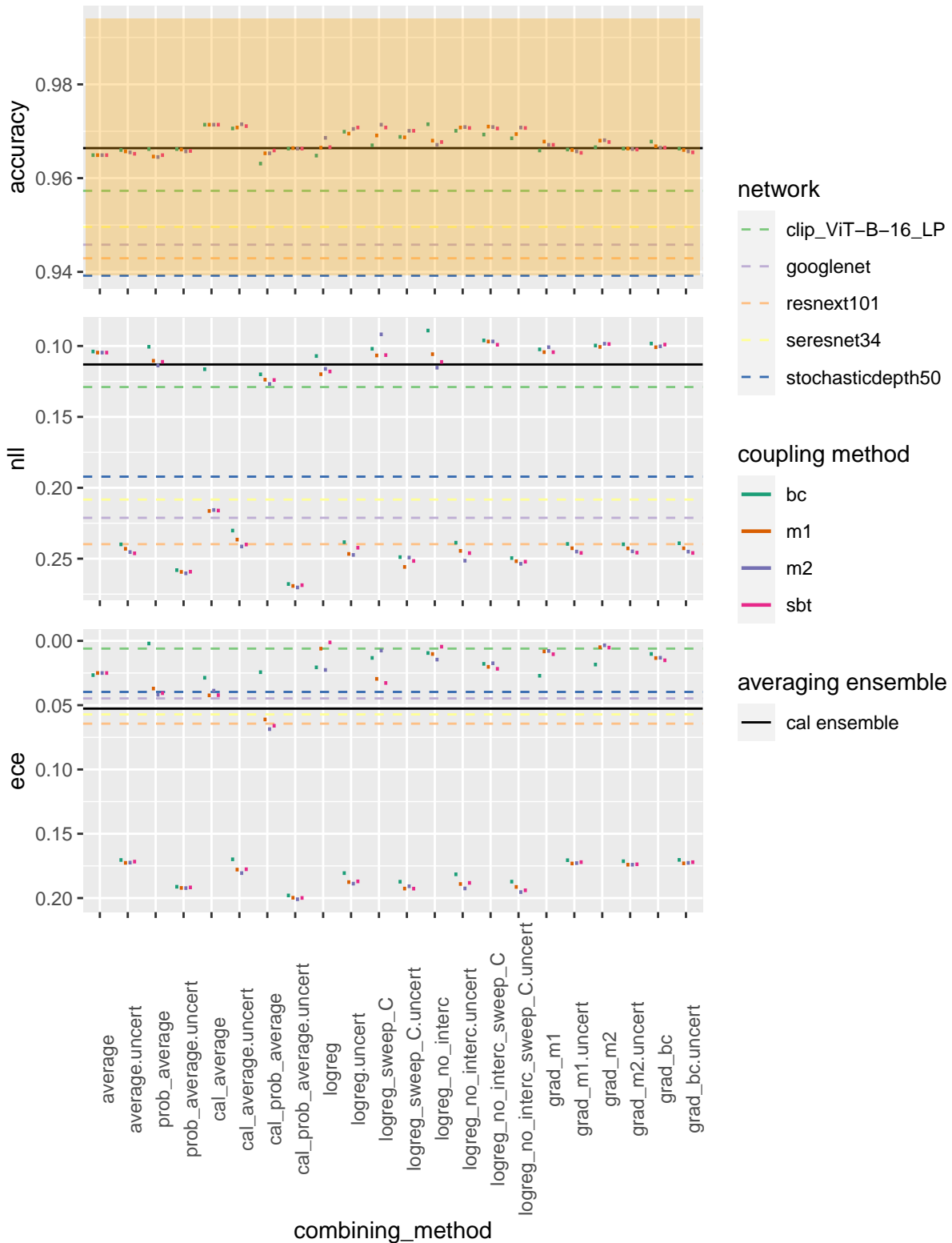
Average pairwise accuracy variance 5.25626228409237e-06



Ensemble metrics

Error inconsistency 0.123499996960163

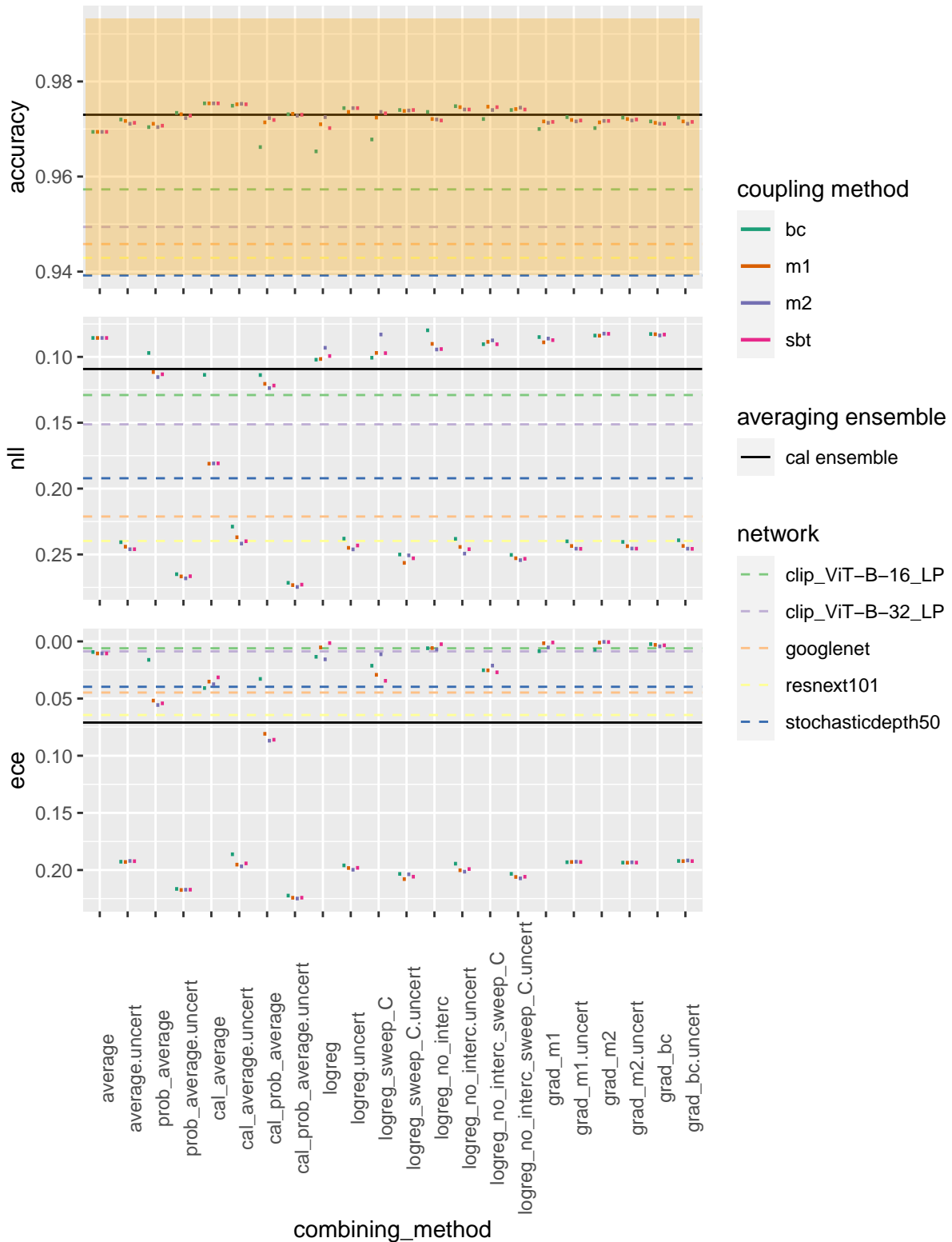
Average pairwise accuracy variance 5.54838925381773e-06



Ensemble metrics

Error inconsistency 0.131099998950958

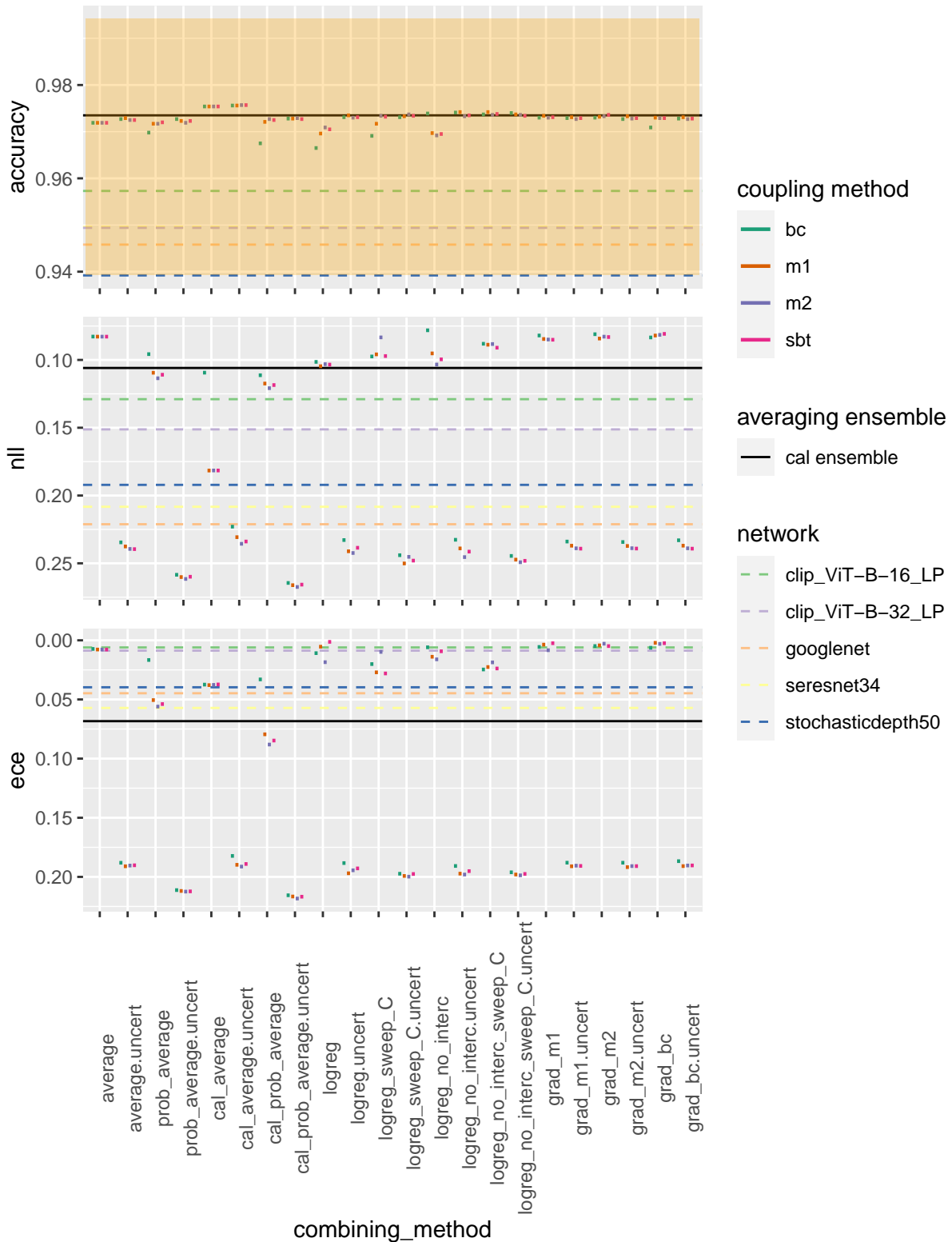
Average pairwise accuracy variance 7.01279577697278e-06



Ensemble metrics

Error inconsistency 0.12950000166893

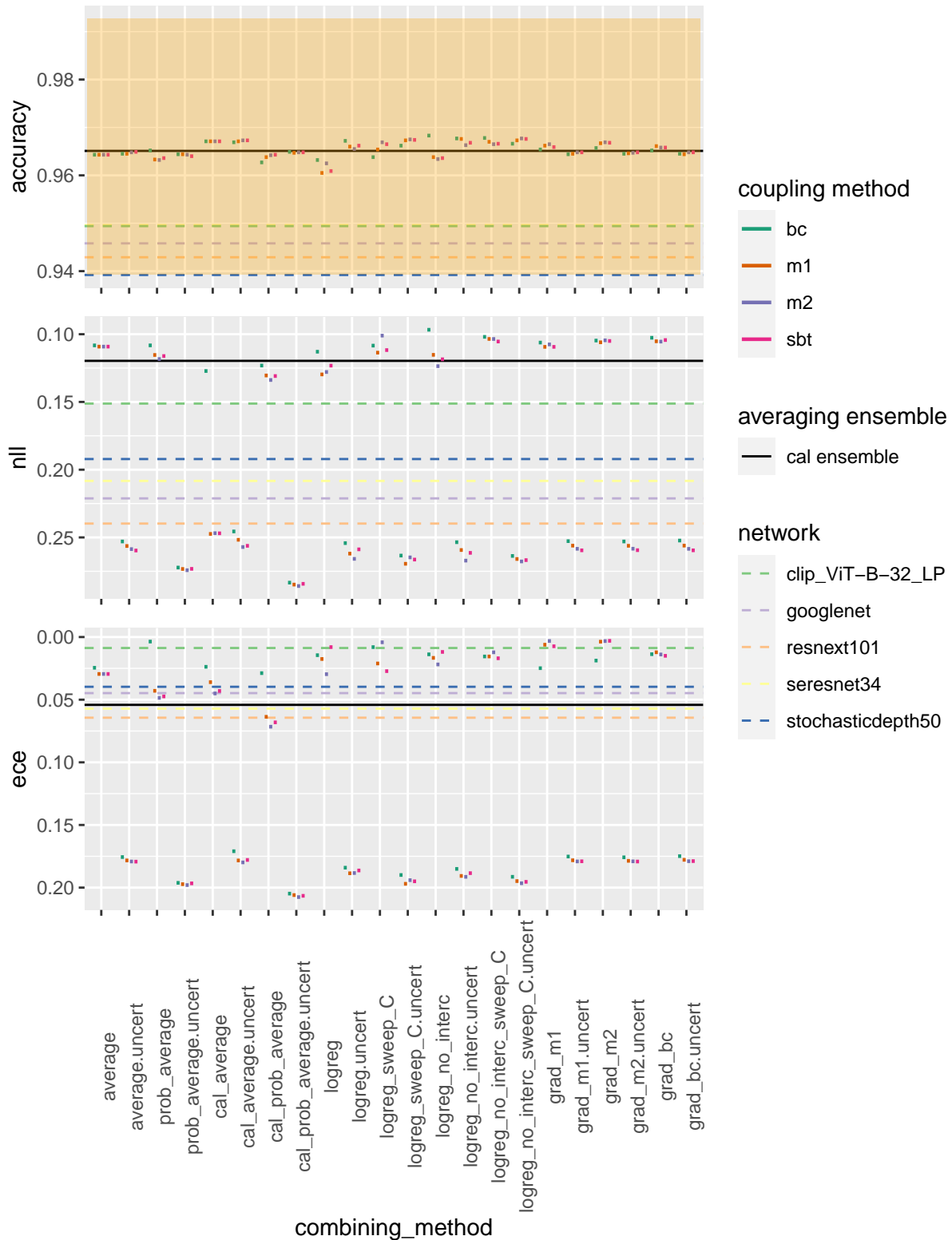
Average pairwise accuracy variance 6.44159354123985e-06



Ensemble metrics

Error inconsistency 0.12610000371933

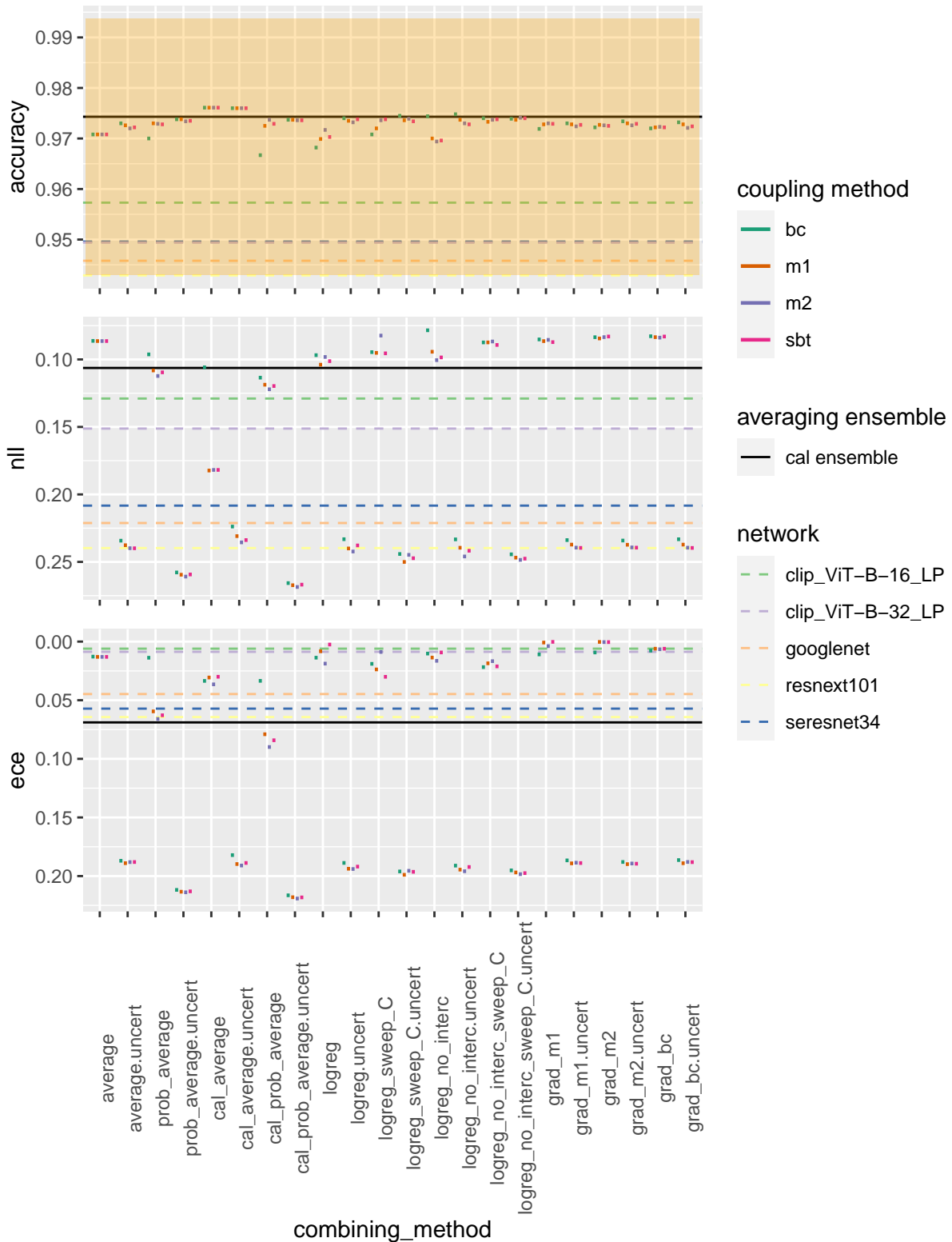
Average pairwise accuracy variance 4.57599617220694e-06



Ensemble metrics

Error inconsistency 0.128299996256828

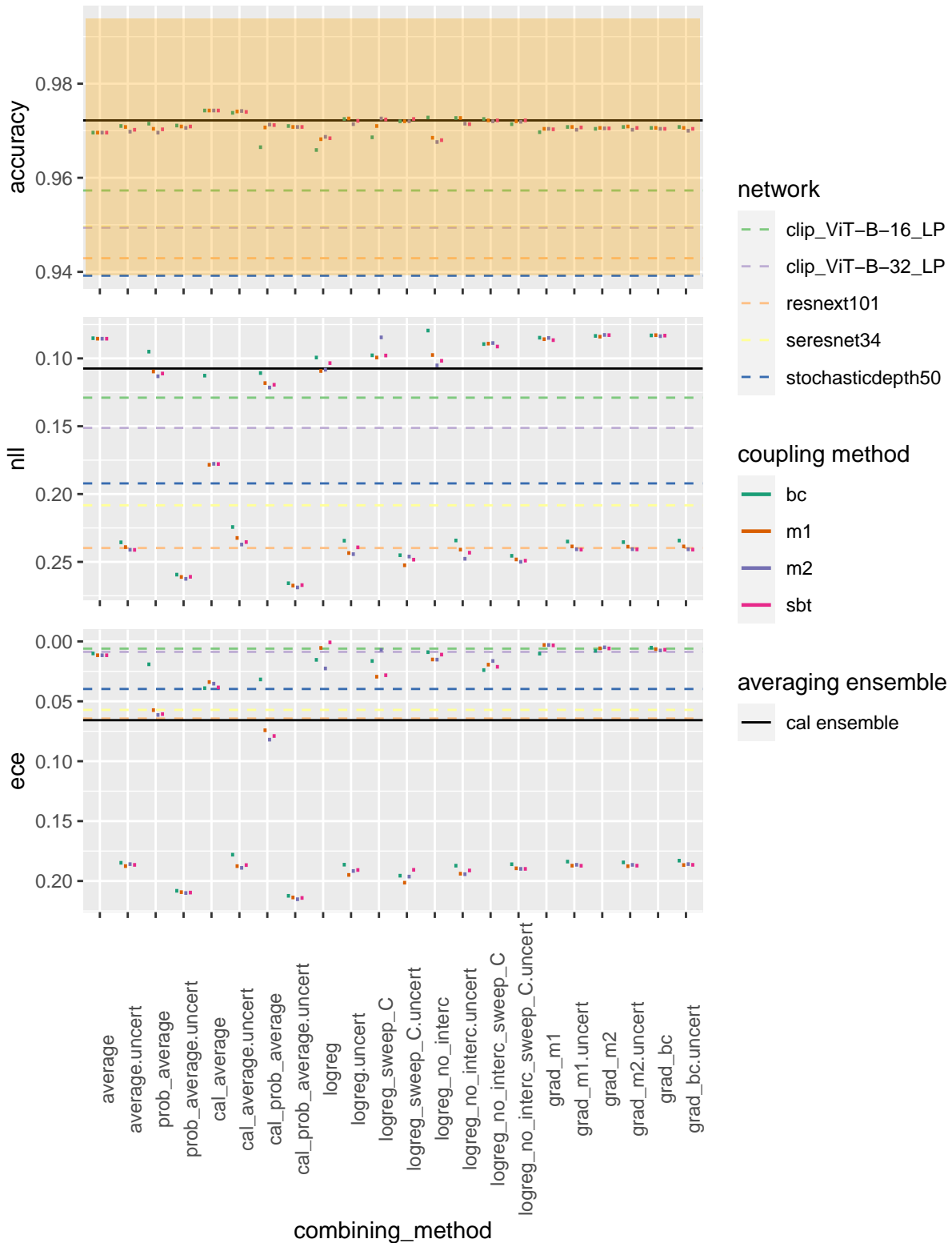
Average pairwise accuracy variance 5.77278888158617e-06



Ensemble metrics

Error inconsistency 0.130699992179871

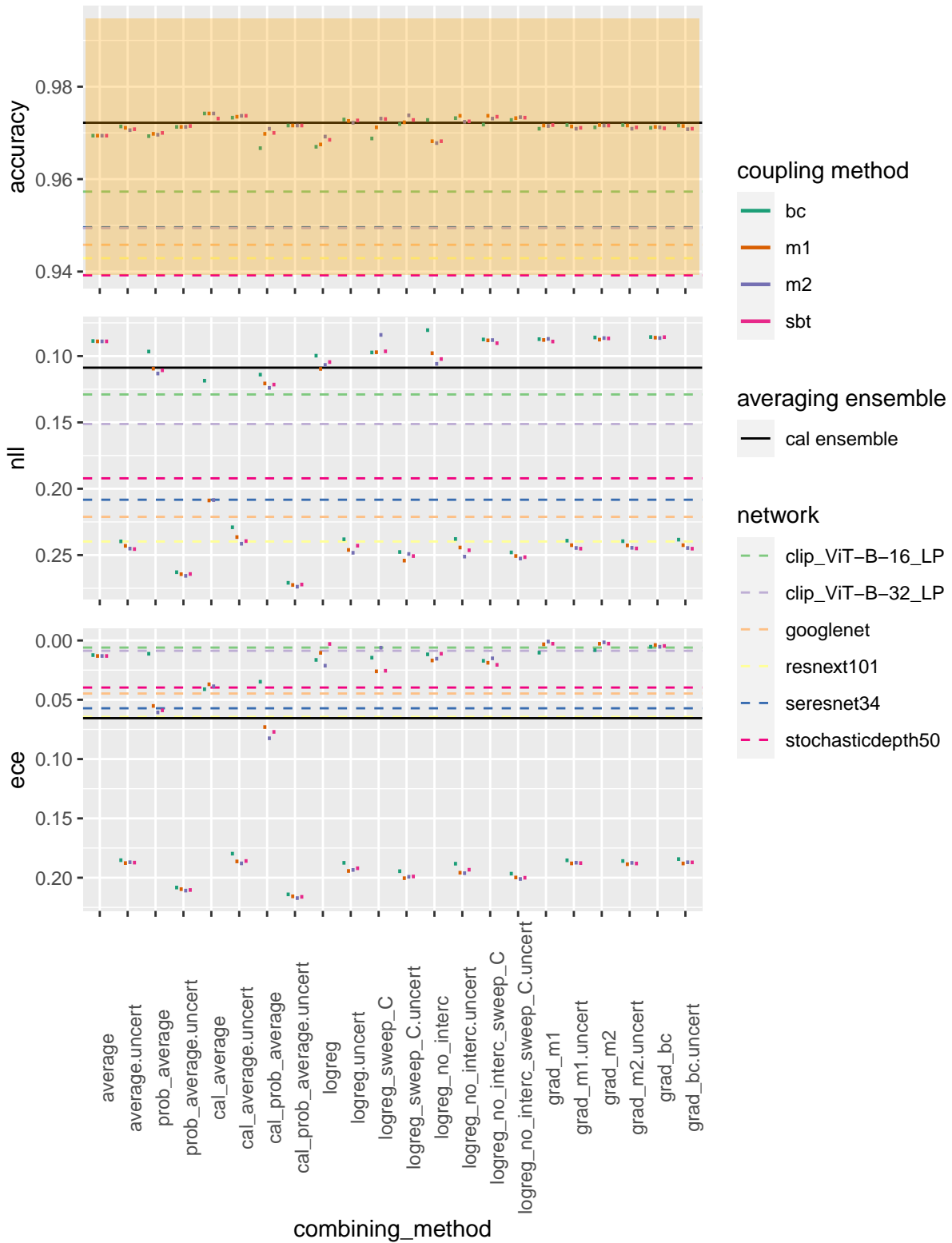
Average pairwise accuracy variance 6.75881392453448e-06



Ensemble metrics

Error inconsistency 0.140599995851517

Average pairwise accuracy variance 6.26915971224662e-06



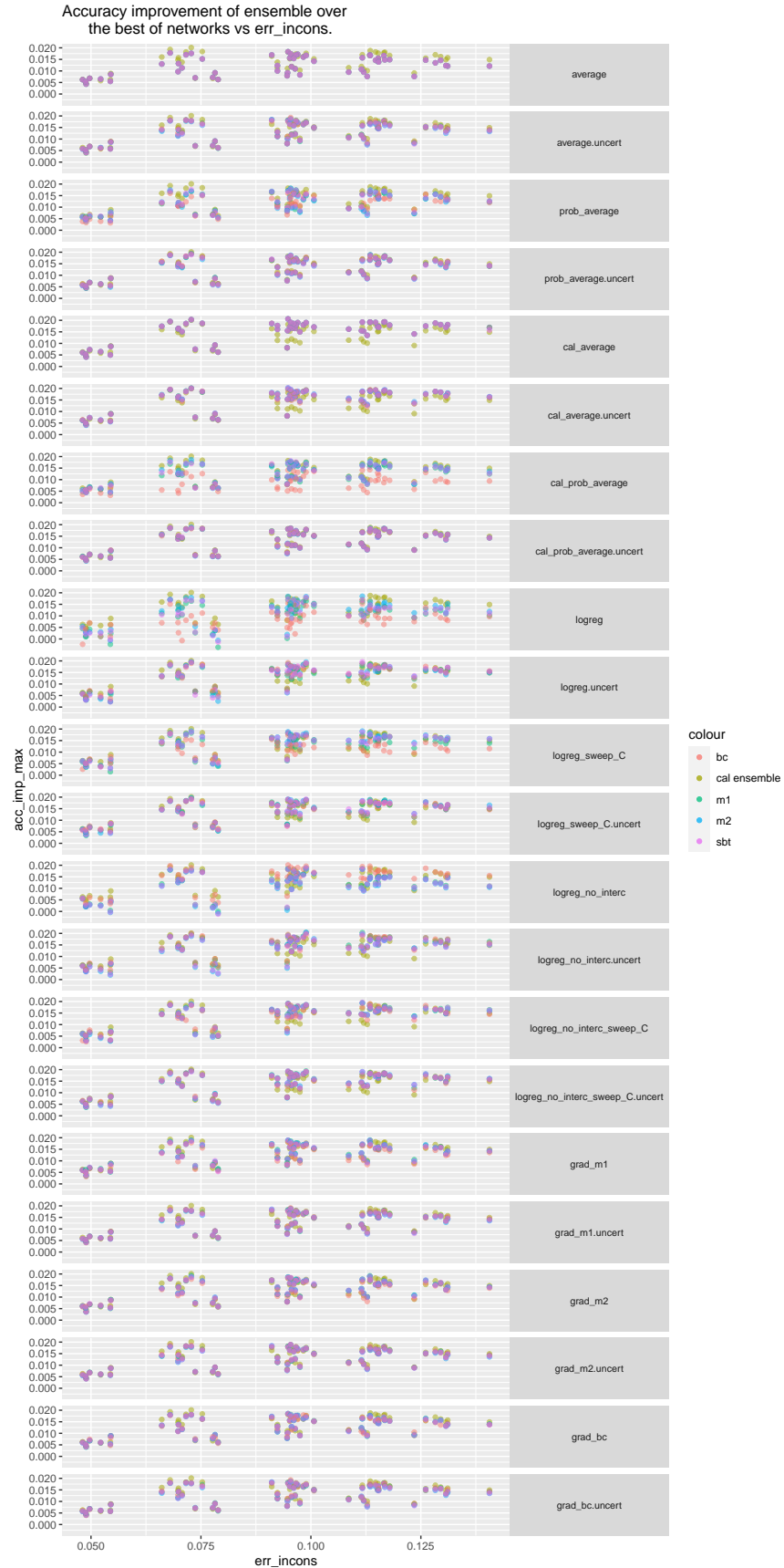

```

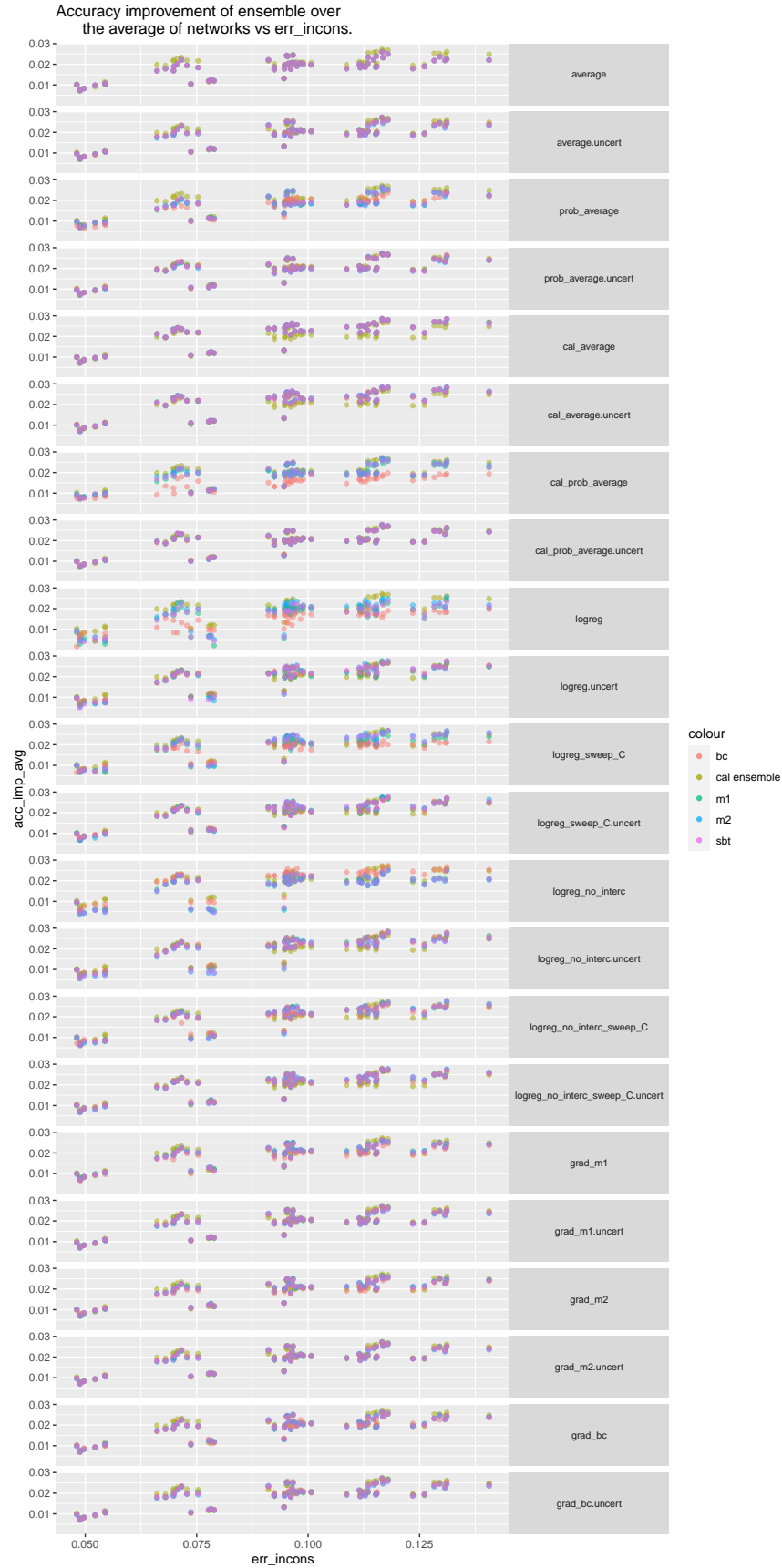
xax <- c(
  "err_incons", "mean_pwa_var"
)
yax <- c("acc_imp_max", "acc_imp_avg")

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

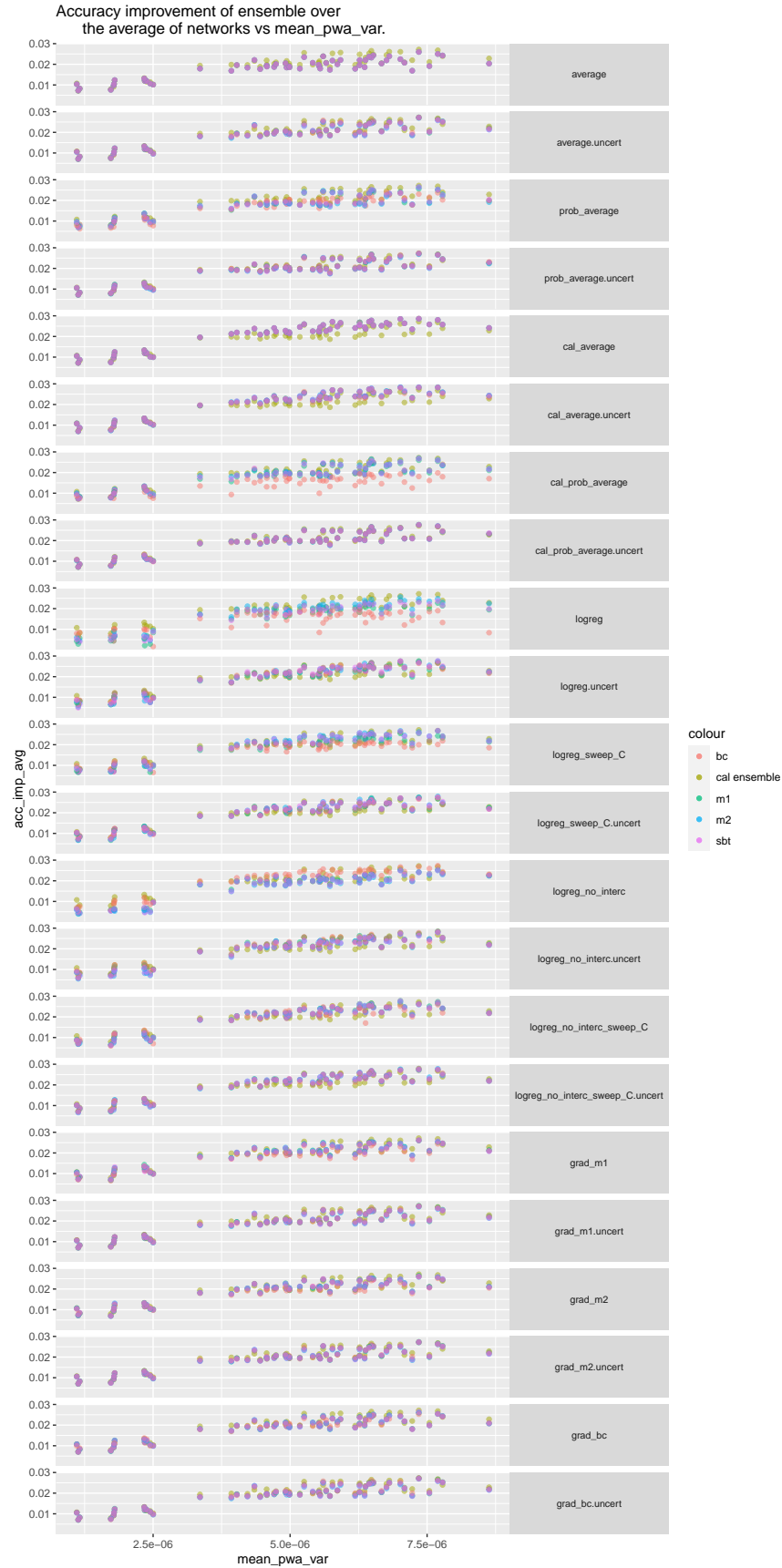
    print(cur_plot)
  }
}

```









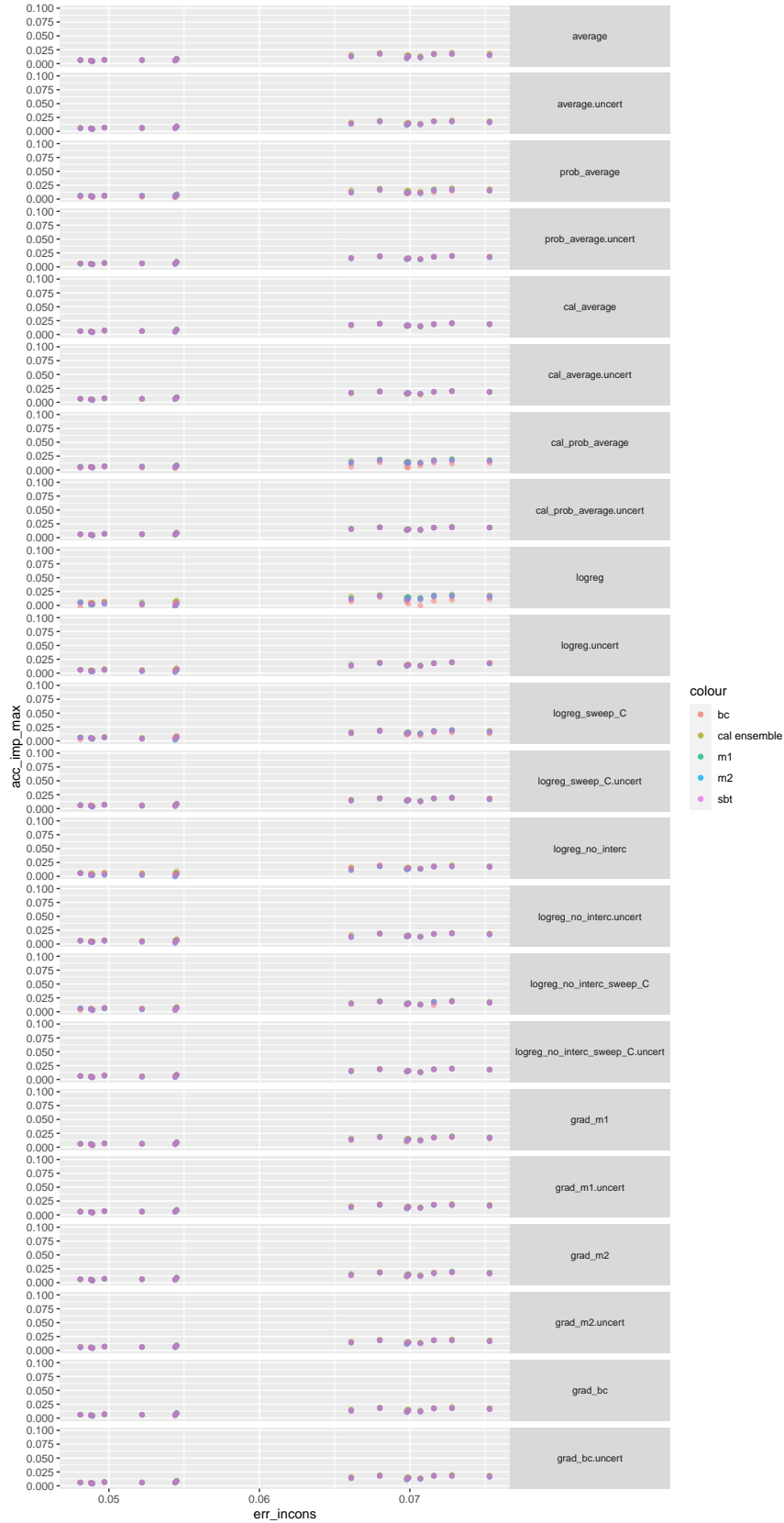
```

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))

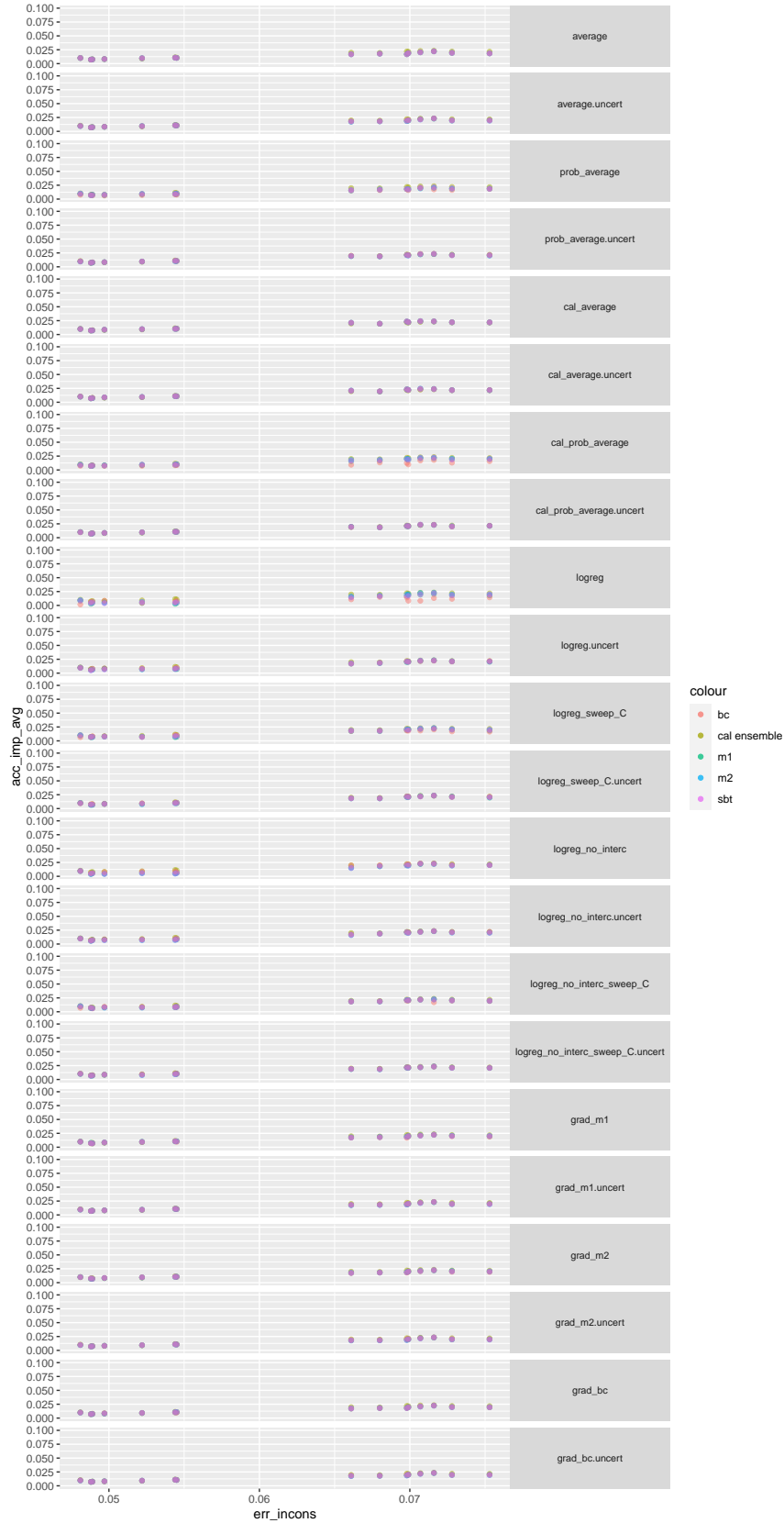
      print(cur_plot)
    }
  }
}

```

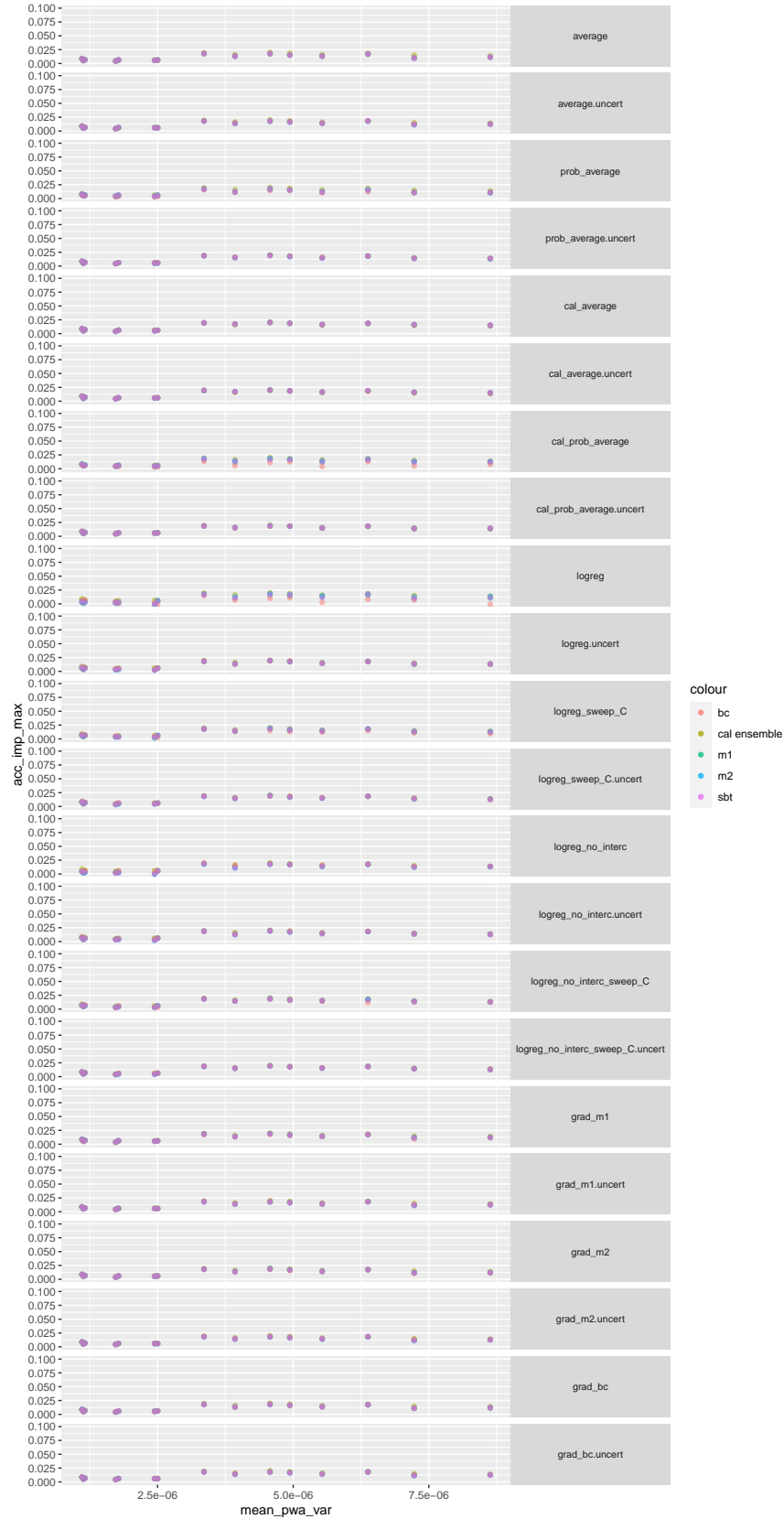
Accuracy improvement of ensemble over
the best of networks vs err_incons.
Ensemble size 2



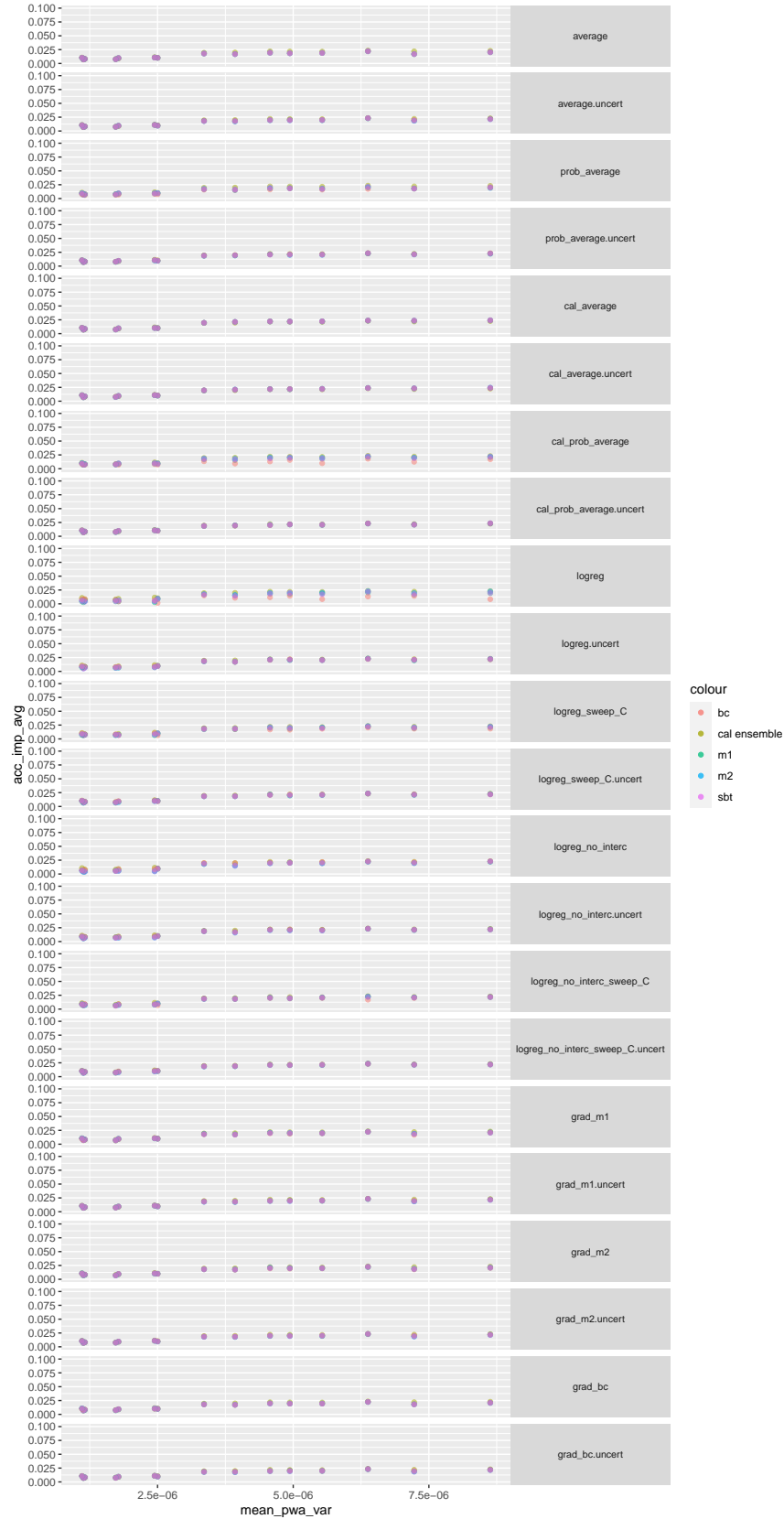
Accuracy improvement of ensemble over
the average of networks vs err_incons.
Ensemble size 2



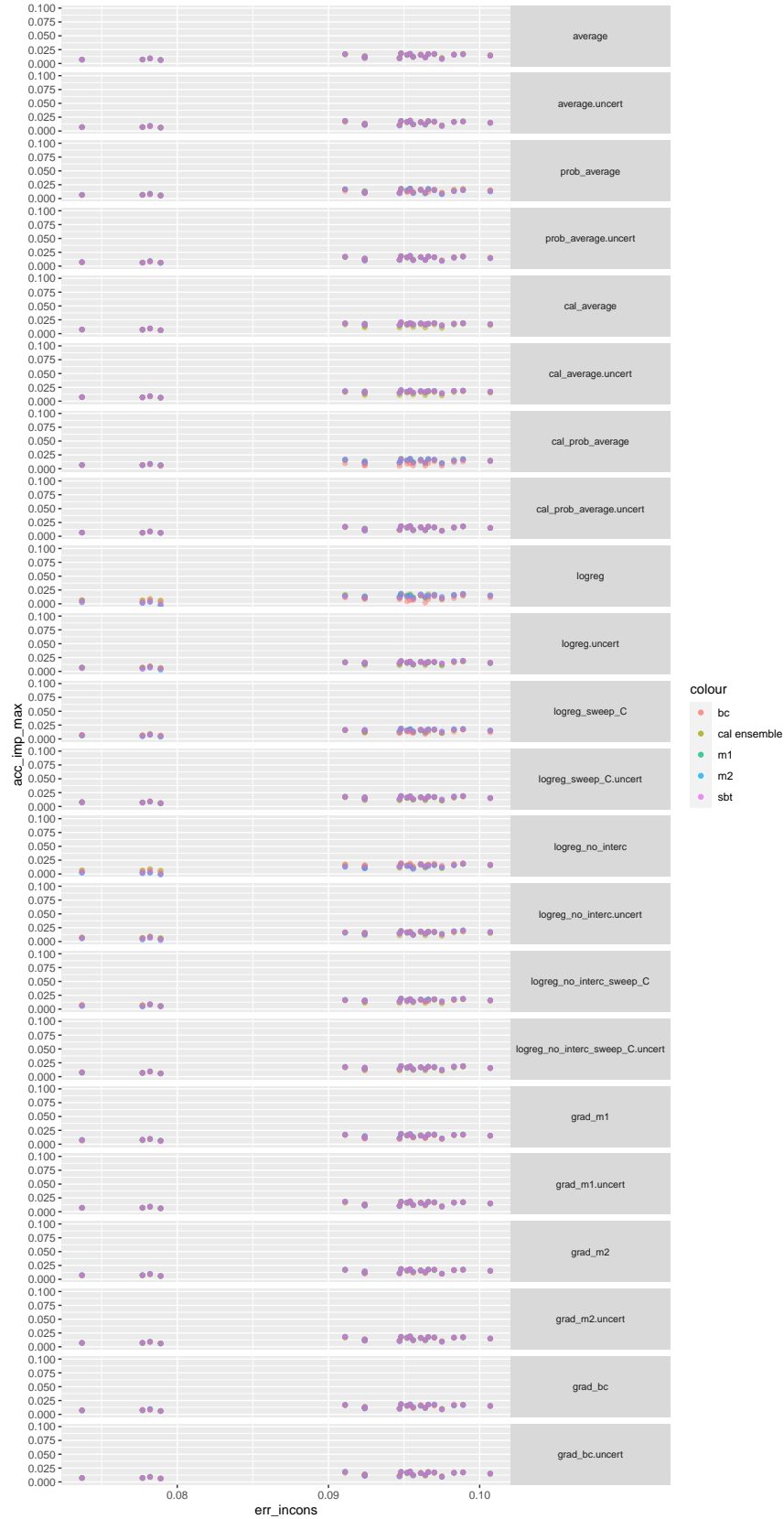
Accuracy improvement of ensemble over
the best of networks vs mean_pwa_var.
Ensemble size 2



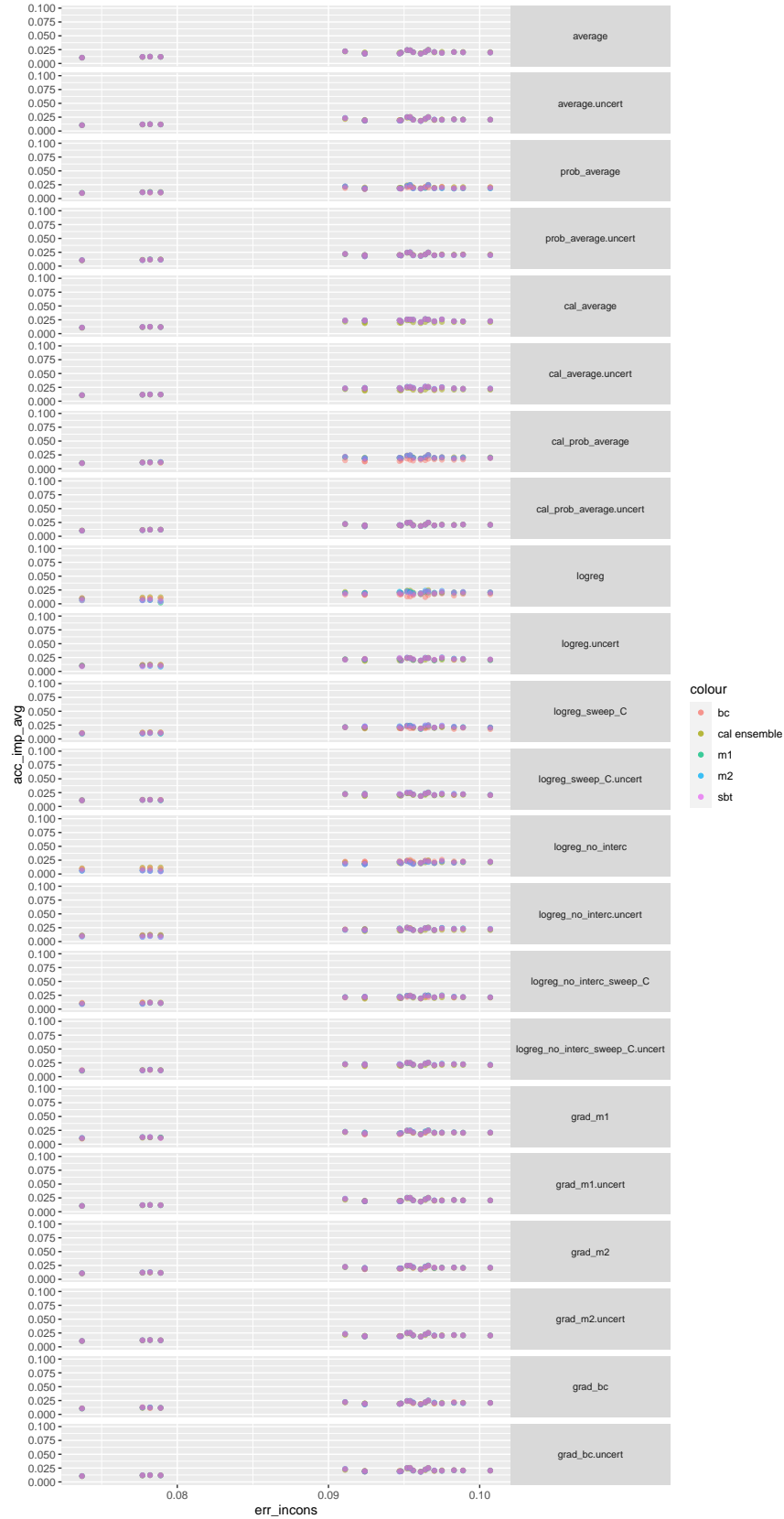
Accuracy improvement of ensemble over
the average of networks vs mean_pwa_var.
Ensemble size 2



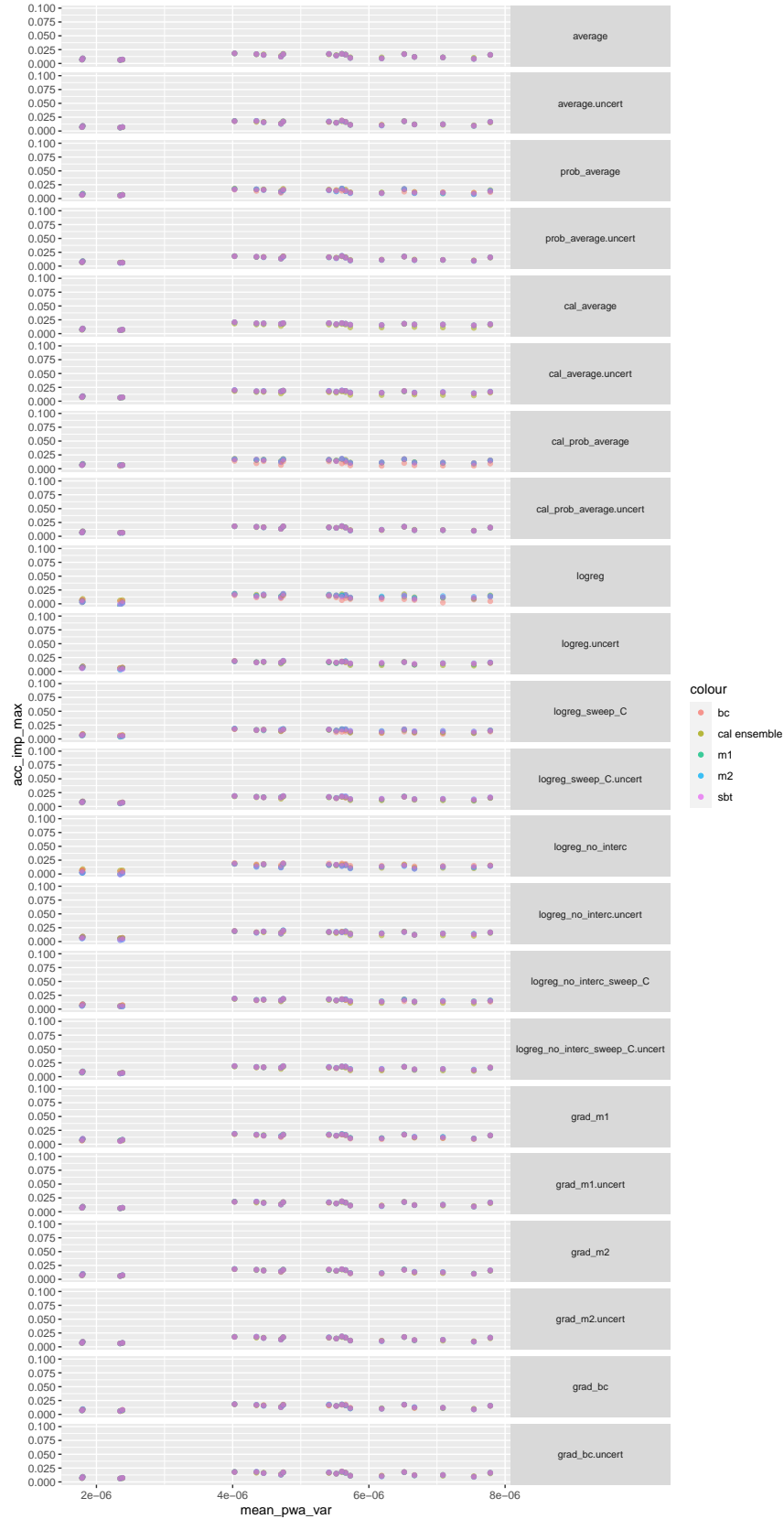
Accuracy improvement of ensemble over
the best of networks vs err_incons.
Ensemble size 3



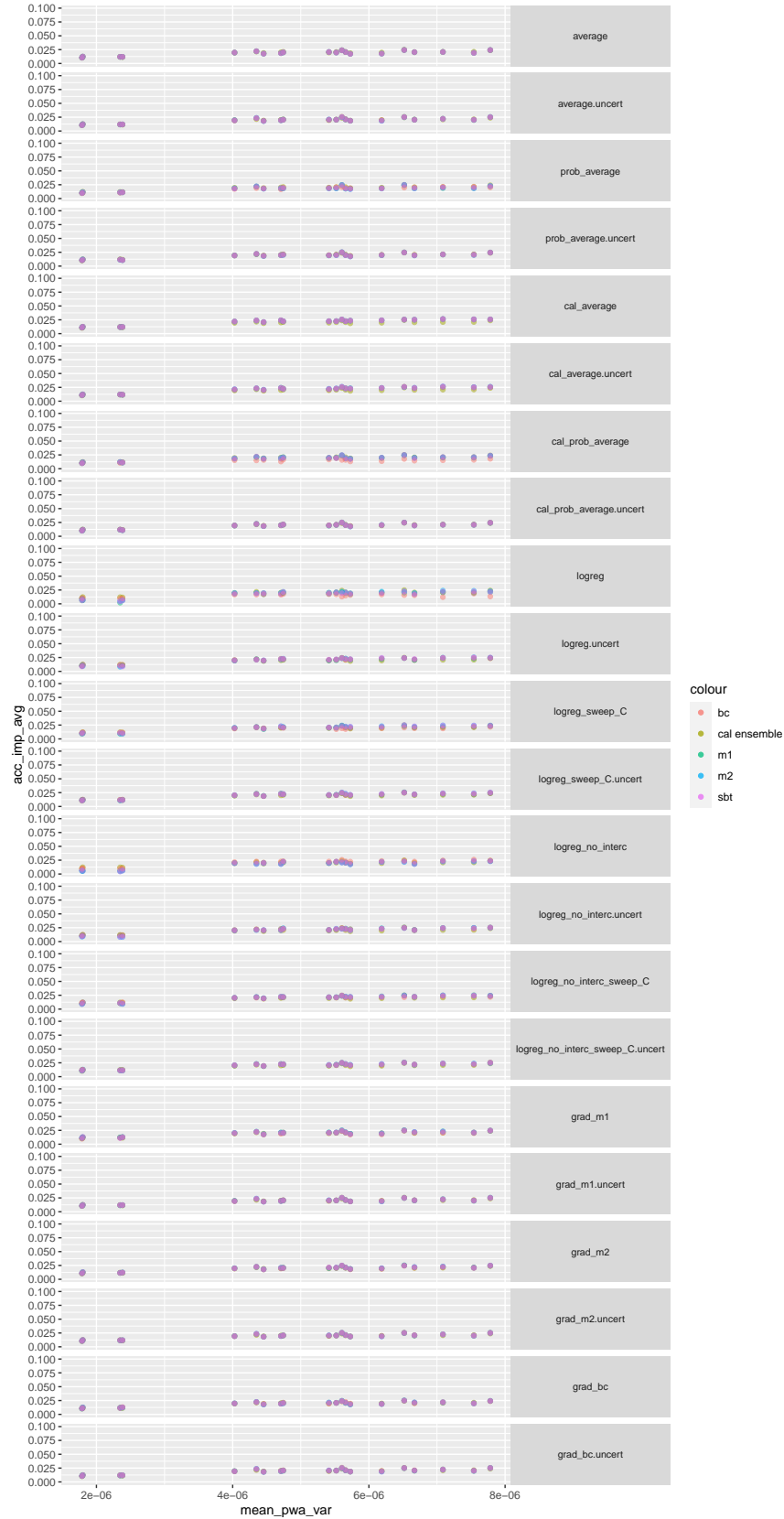
Accuracy improvement of ensemble over
the average of networks vs err_incons.
Ensemble size 3



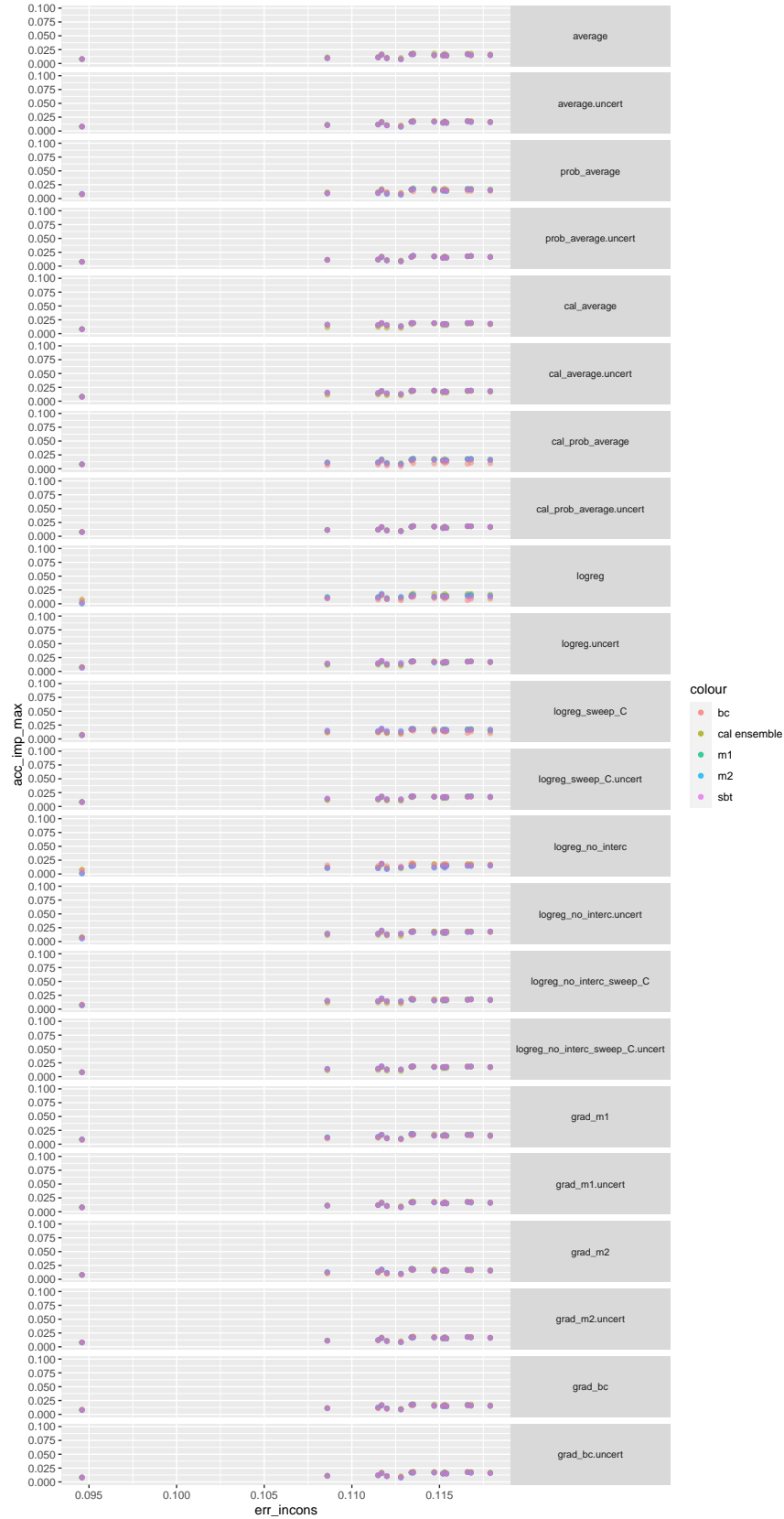
Accuracy improvement of ensemble over
the best of networks vs mean_pwa_var.
Ensemble size 3



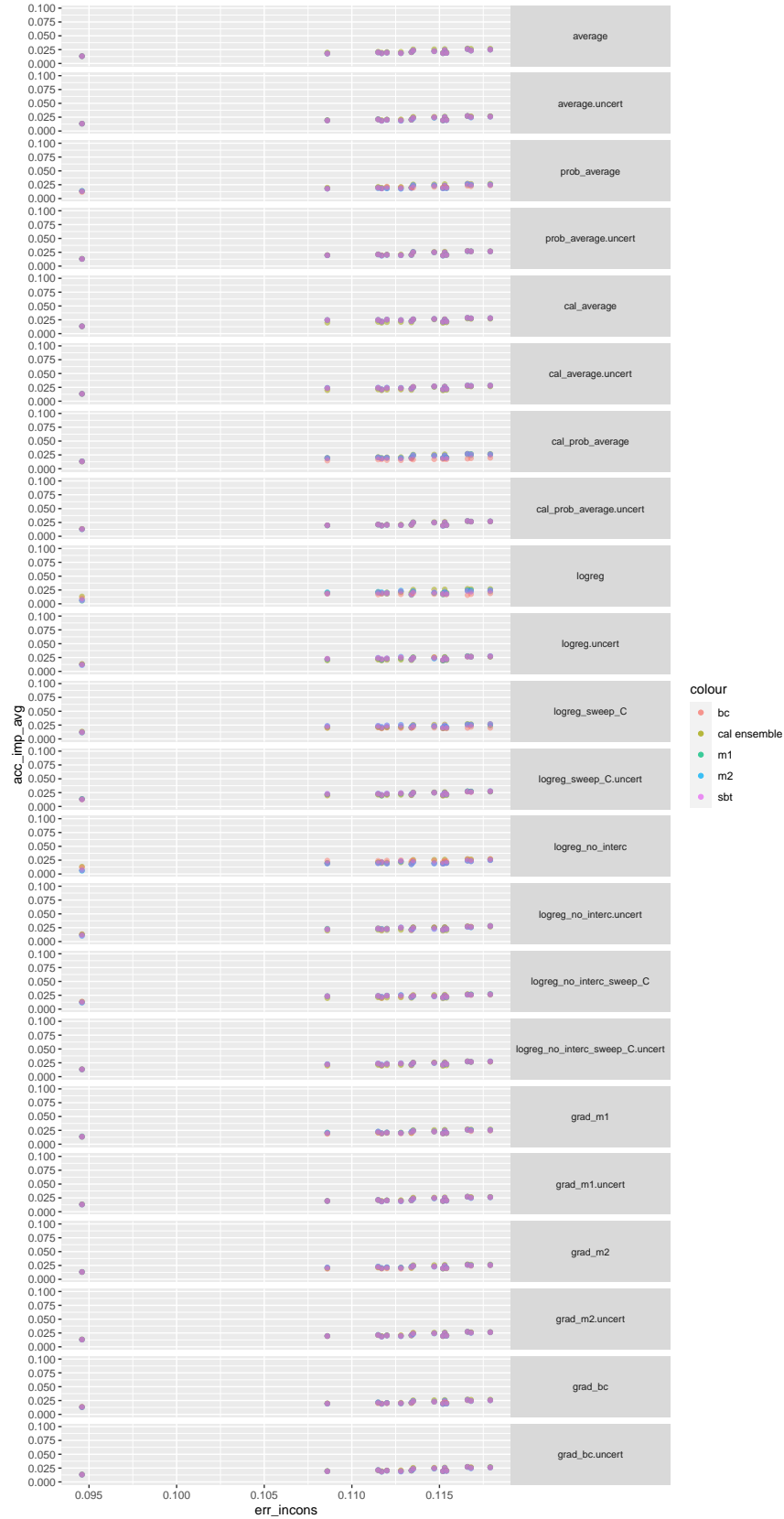
Accuracy improvement of ensemble over
the average of networks vs mean_pwa_var.
Ensemble size 3



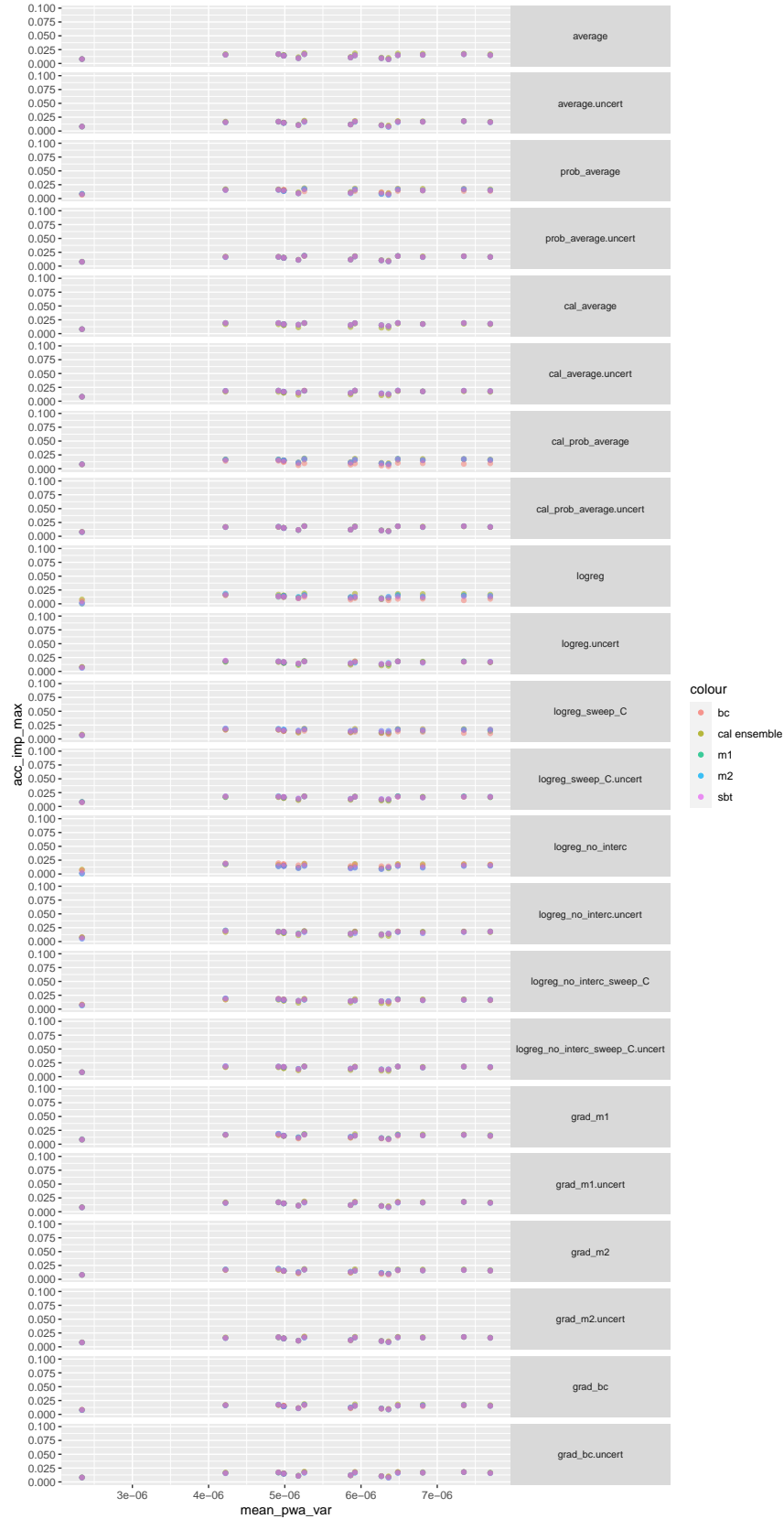
Accuracy improvement of ensemble over
the best of networks vs err_incons.
Ensemble size 4



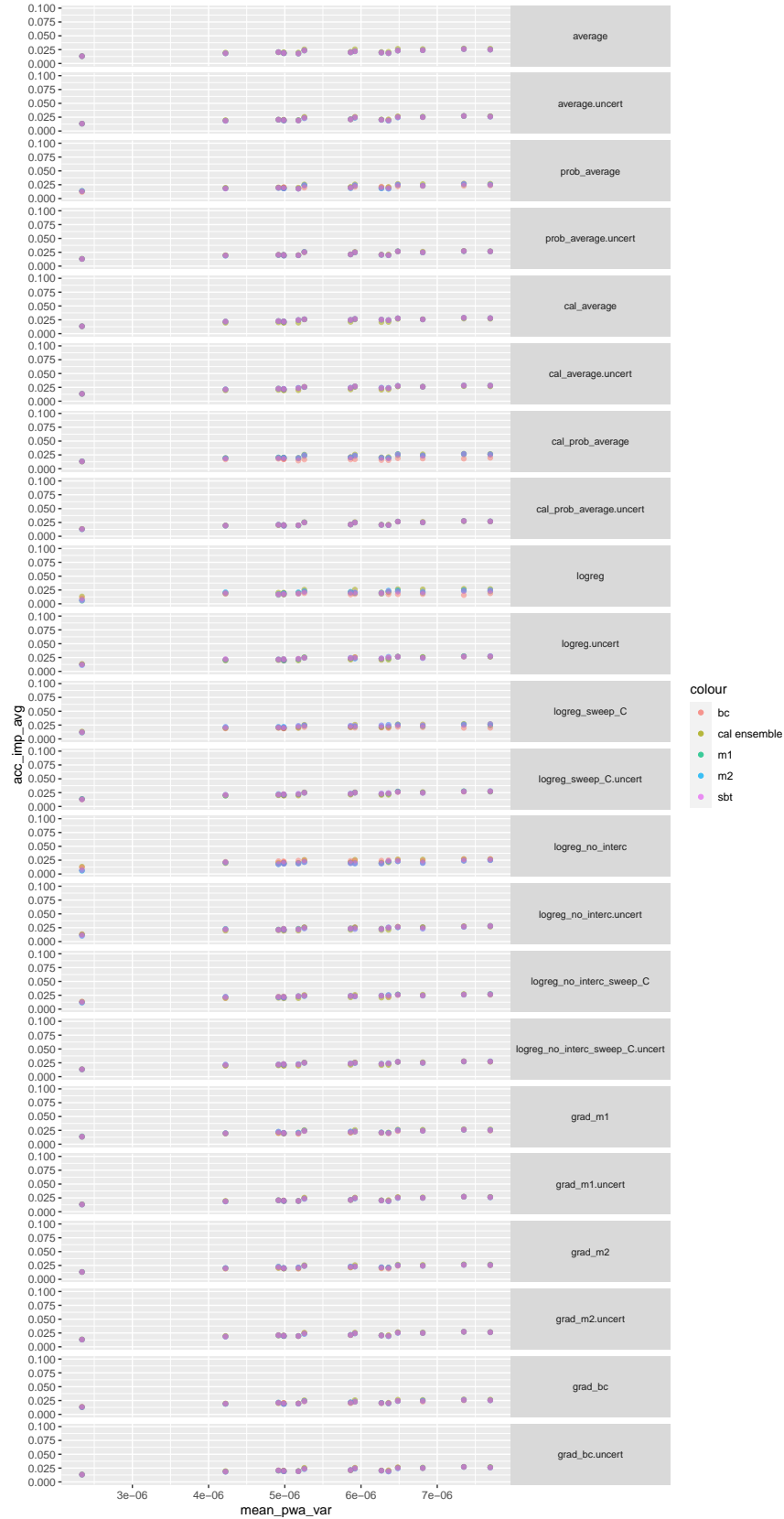
Accuracy improvement of ensemble over
the average of networks vs err_incons.
Ensemble size 4

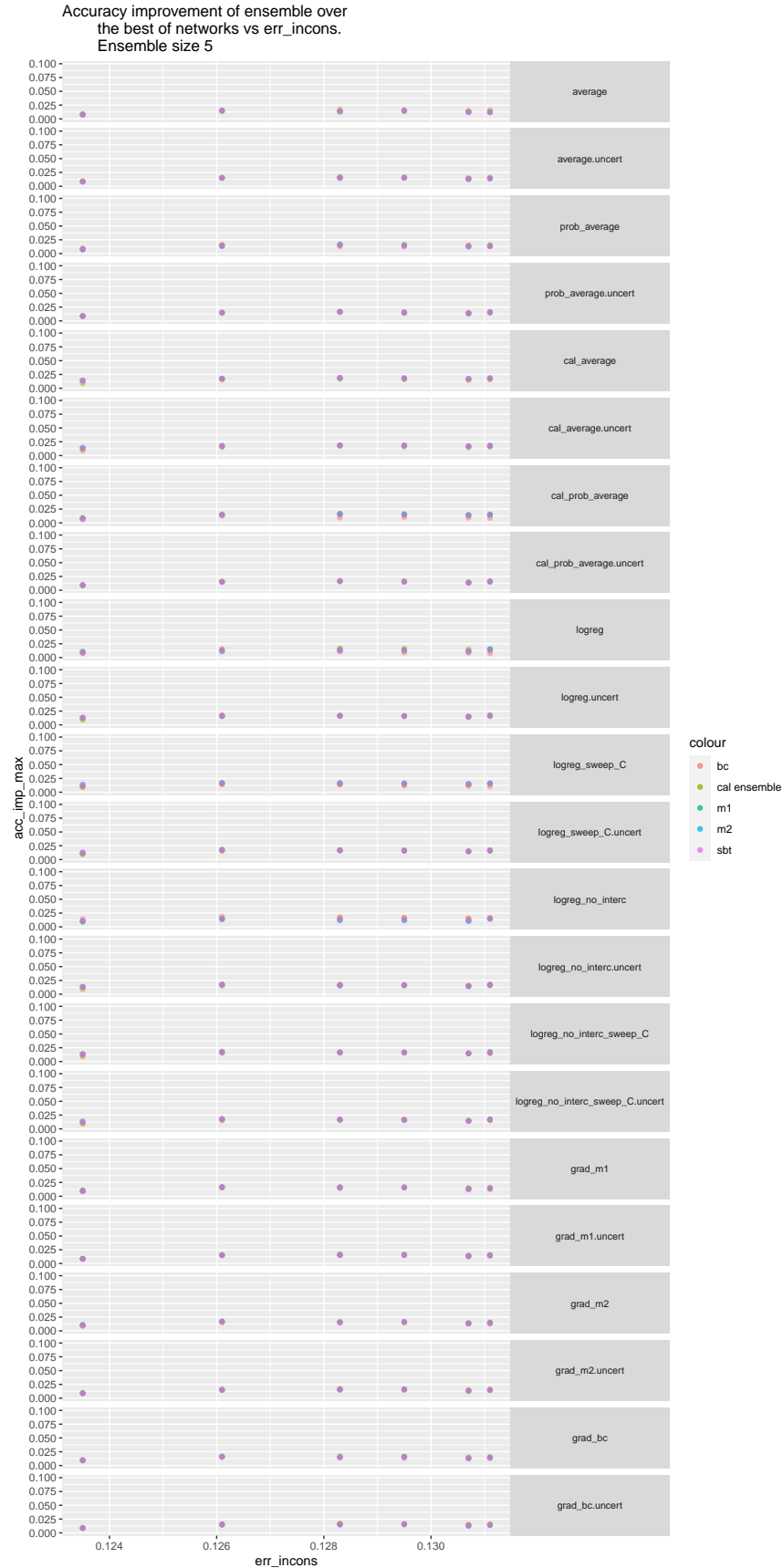


Accuracy improvement of ensemble over
the best of networks vs mean_pwa_var.
Ensemble size 4



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

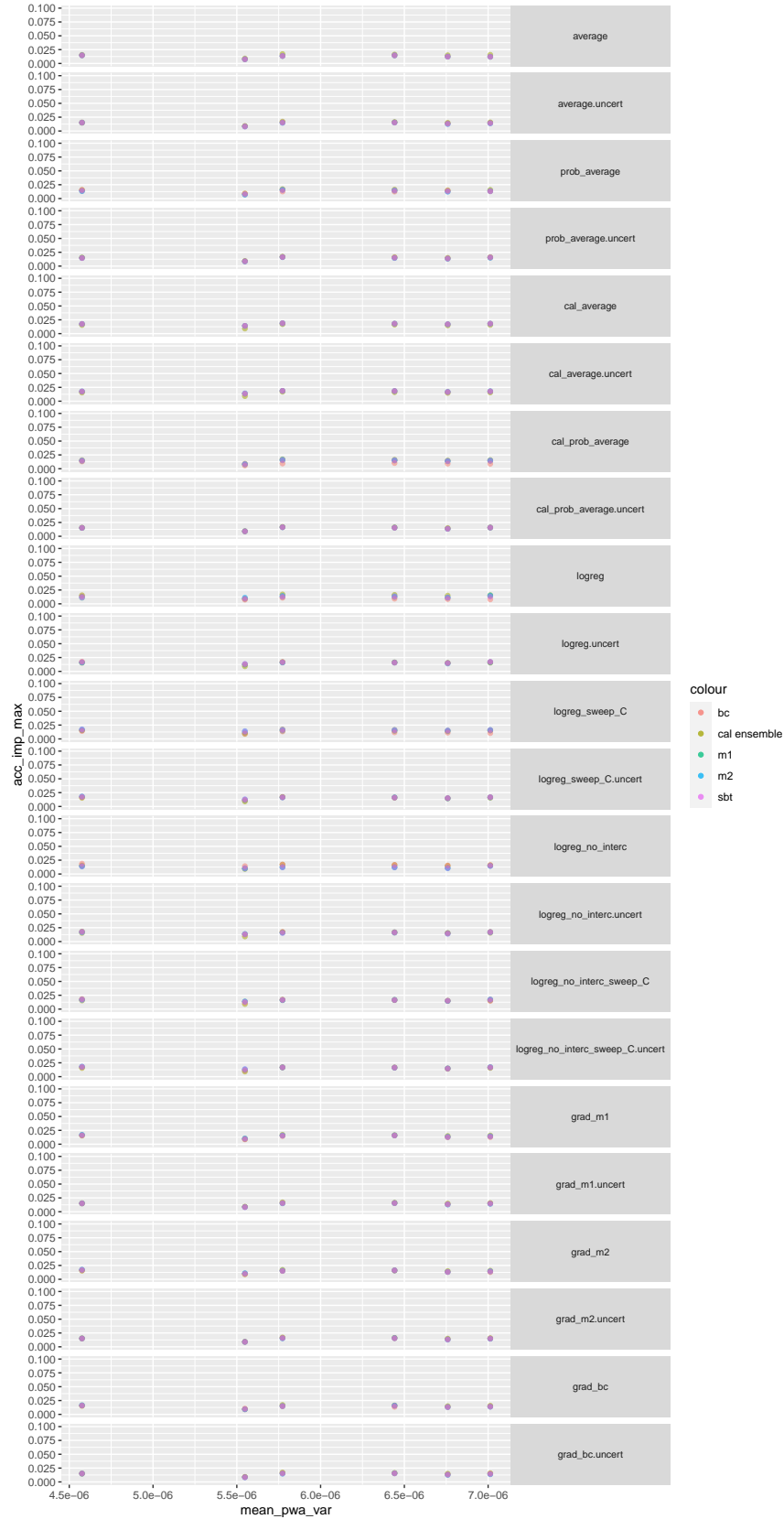




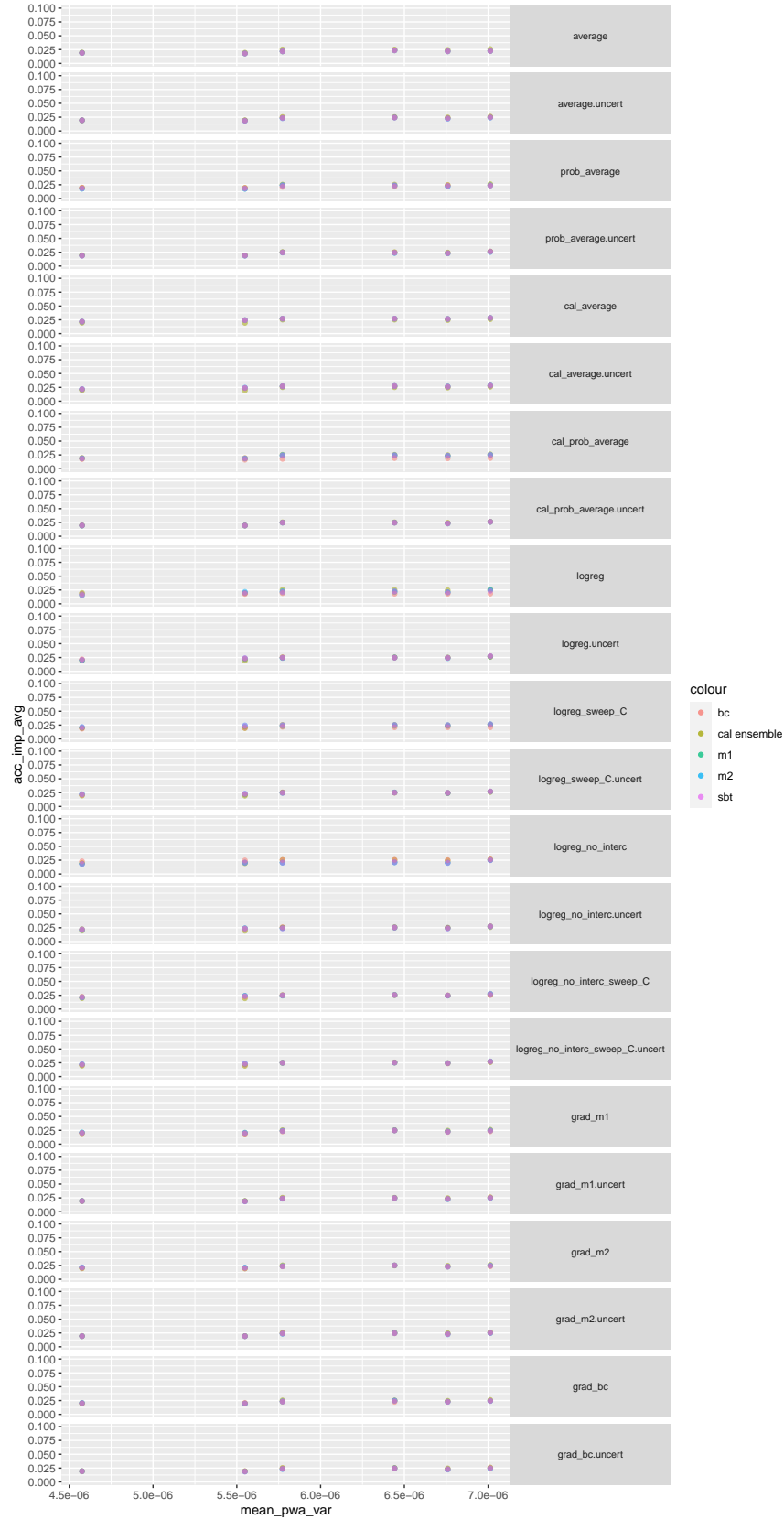
Accuracy improvement of ensemble over
the average of networks vs err_incons.
Ensemble size 5



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

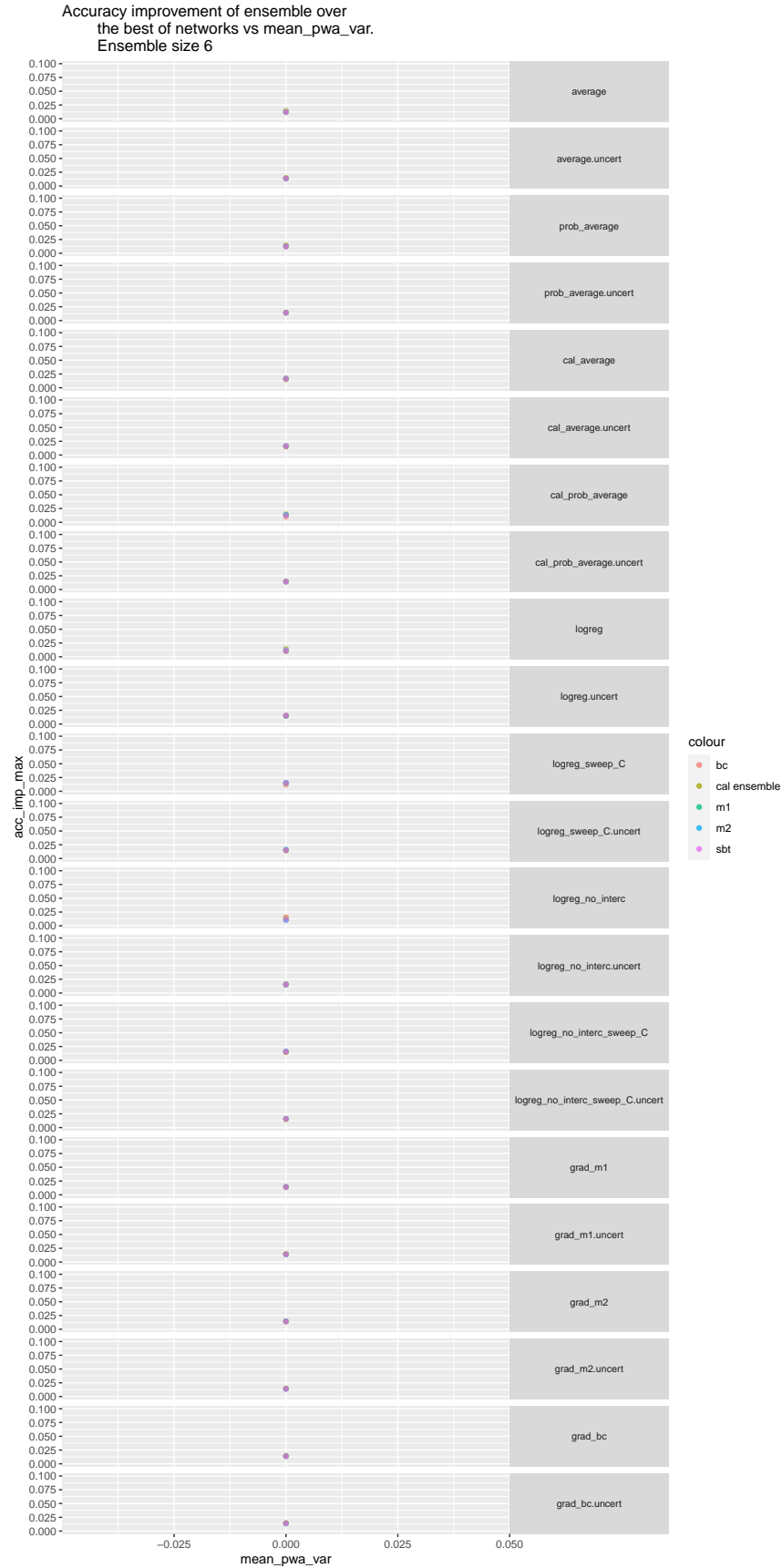


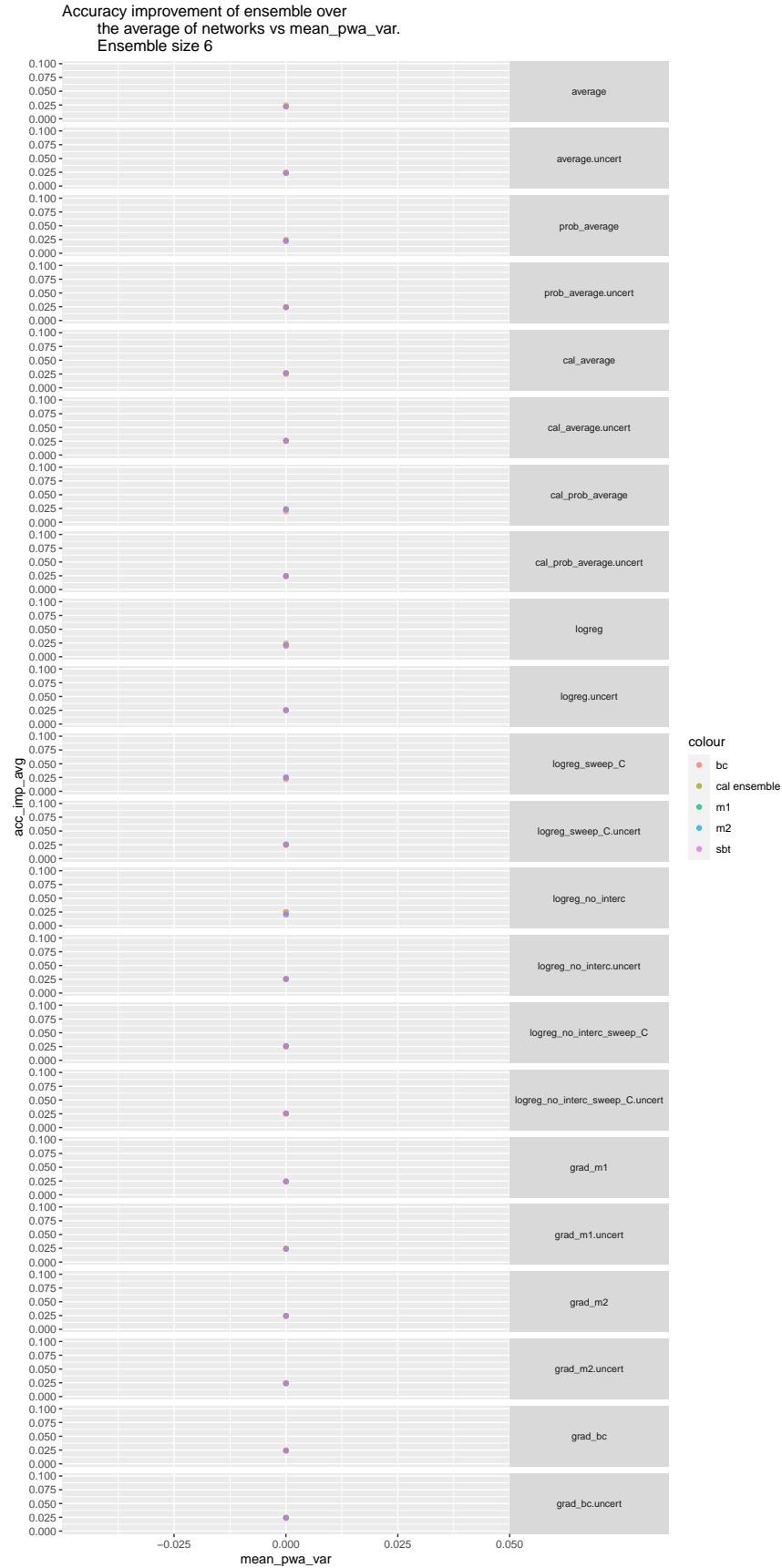
Accuracy improvement of ensemble over
the average of networks vs mean_pwa_var.
Ensemble size 5











```

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

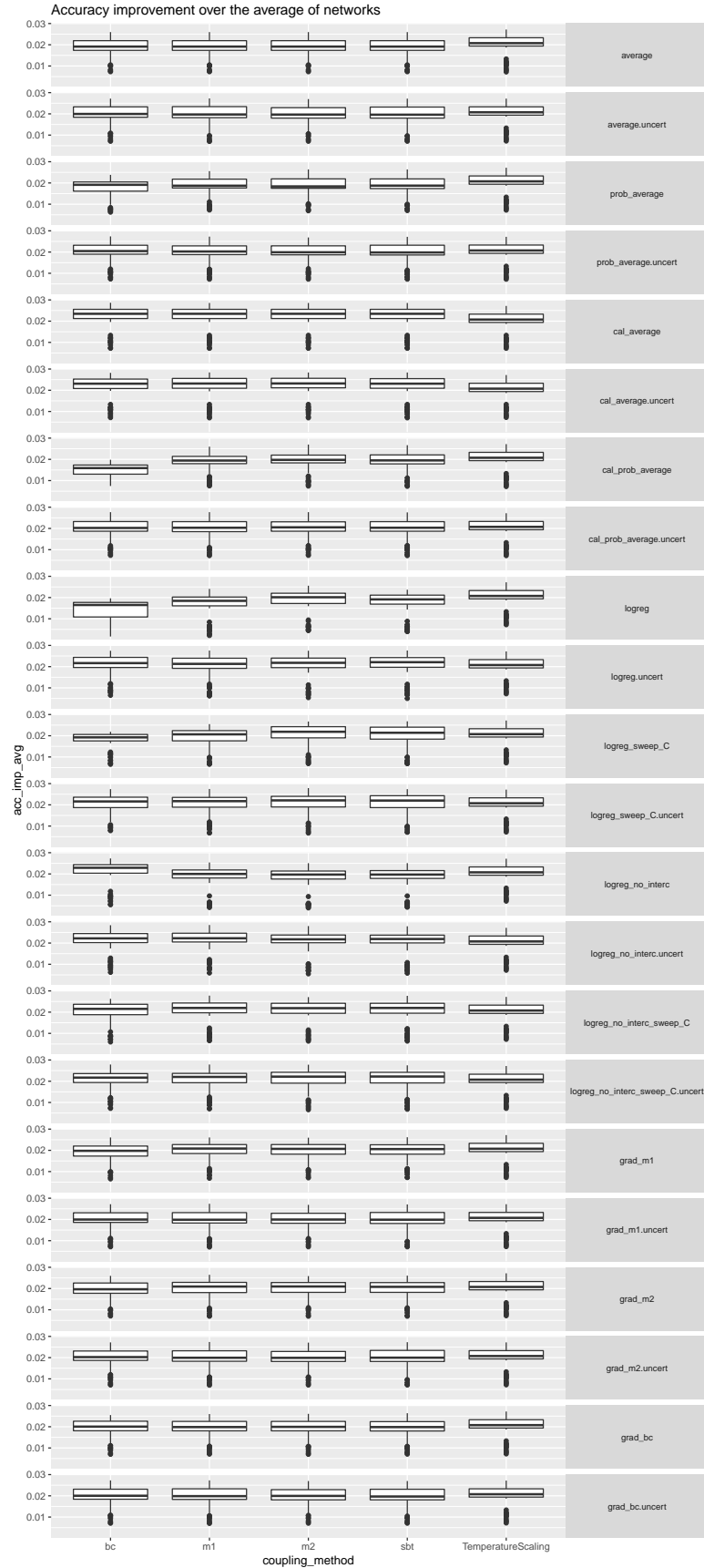
```

```

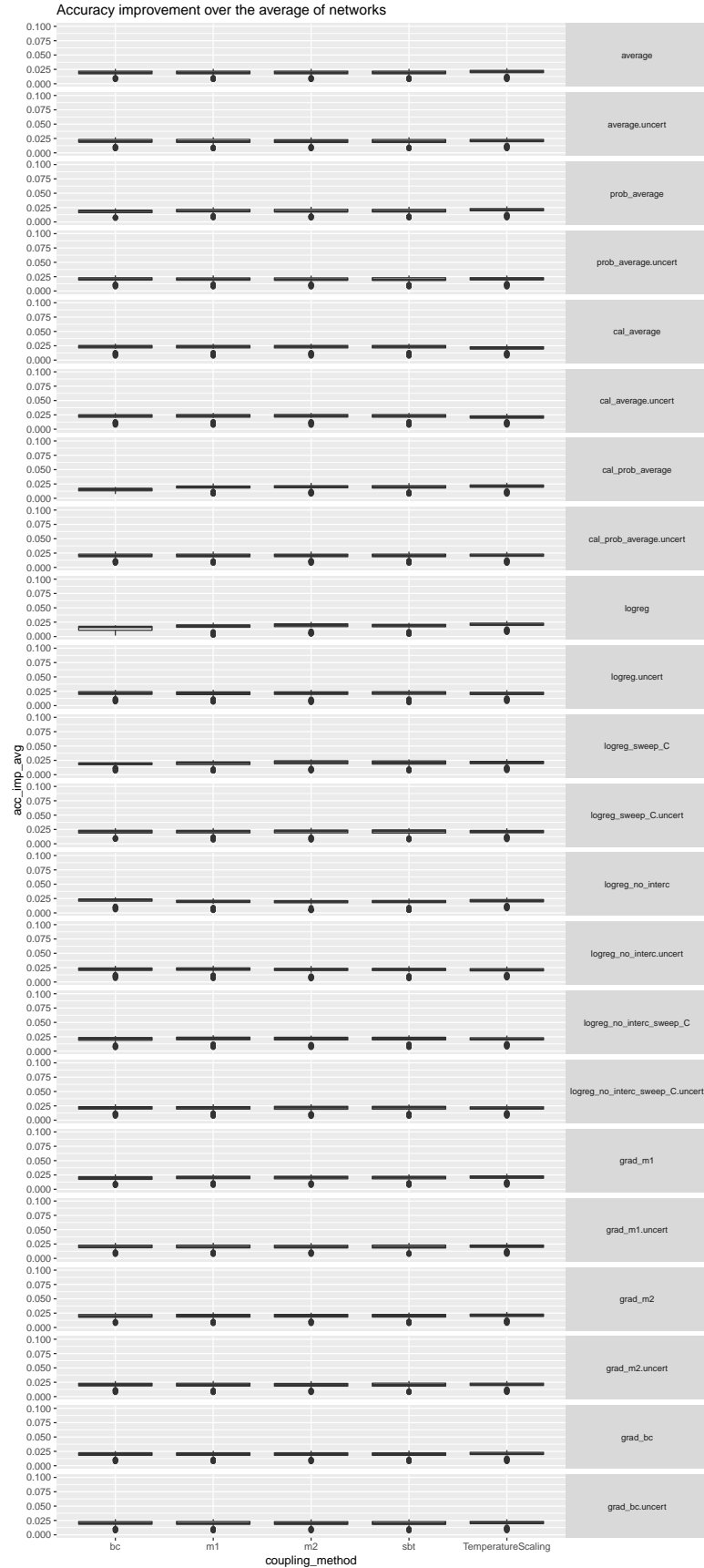
imp_avg_plot <- ggplot() +
  geom_boxplot(data = ens_pwc_plt_df, mapping = aes(x = coupling_method, y = acc_imp_avg)) +
  geom_boxplot(data = ens_cal_plt_df, mapping = aes(x = calibrating_method, y = acc_imp_avg)) +
  facet_grid(rows = vars(combining_method)) +
  ggtitle("Accuracy improvement over the average of networks") +
  theme(strip.text.y = element_text(size = 8, angle = 0))

print(imp_avg_plot)

```



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



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 =
```

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Thu Feb 10 10:49:09 2022

	method	imp_o_avg	imp_o_max
1	cal_average bc	0.0214	0.0153
2	cal_average m1	0.0214	0.0153
3	cal_average m2	0.0214	0.0153
4	cal_average sbt	0.0214	0.0153
5	cal_average.uncert m2	0.0213	0.0152
6	cal_average.uncert sbt	0.0213	0.0152
7	cal_average.uncert m1	0.0212	0.0151
8	cal_average.uncert bc	0.0210	0.0150
9	logreg_no_interc.uncert bc	0.0205	0.0144
10	logreg_no_interc.uncert m1	0.0204	0.0143
11	logreg_no_interc_sweep_C.uncert m1	0.0203	0.0143
12	logreg_no_interc bc	0.0203	0.0143
13	logreg_no_interc_sweep_C.uncert sbt	0.0203	0.0143
14	logreg_no_interc_sweep_C.uncert m2	0.0203	0.0142
15	logreg_sweep_C.uncert m2	0.0203	0.0142
16	logreg_no_interc_sweep_C.uncert bc	0.0203	0.0142
17	logreg_sweep_C.uncert sbt	0.0202	0.0141
18	logreg_no_interc_sweep_C m1	0.0201	0.0140
19	logreg.uncert bc	0.0201	0.0140
20	logreg_sweep_C.uncert bc	0.0200	0.0140
21	logreg_no_interc_sweep_C sbt	0.0200	0.0139
22	logreg_sweep_C m2	0.0200	0.0139
23	logreg_sweep_C.uncert m1	0.0200	0.0139
24	logreg.uncert sbt	0.0199	0.0138
25	logreg_no_interc_sweep_C m2	0.0199	0.0138
26	logreg.uncert m2	0.0198	0.0137
27	logreg.uncert m1	0.0197	0.0136
28	logreg_sweep_C sbt	0.0197	0.0136
29	logreg_no_interc.uncert sbt	0.0197	0.0136
30	logreg_no_interc.uncert m2	0.0196	0.0136
31	average of TemperatureScaling	0.0196	0.0135
32	logreg_no_interc_sweep_C bc	0.0196	0.0135
33	grad_m1 m1	0.0195	0.0134
34	grad_m2 sbt	0.0194	0.0134
35	grad_m2 m2	0.0194	0.0134
36	cal_prob_average.uncert bc	0.0194	0.0133
37	grad_m2 m1	0.0194	0.0133
38	cal_prob_average.uncert m1	0.0194	0.0133
39	prob_average.uncert bc	0.0194	0.0133
40	cal_prob_average.uncert sbt	0.0194	0.0133
41	cal_prob_average.uncert m2	0.0194	0.0133
42	grad_m1 m2	0.0194	0.0133
43	grad_m2.uncert bc	0.0193	0.0133
44	grad_m1 sbt	0.0193	0.0133
45	grad_bc.uncert bc	0.0193	0.0132
46	grad_m1.uncert bc	0.0193	0.0132

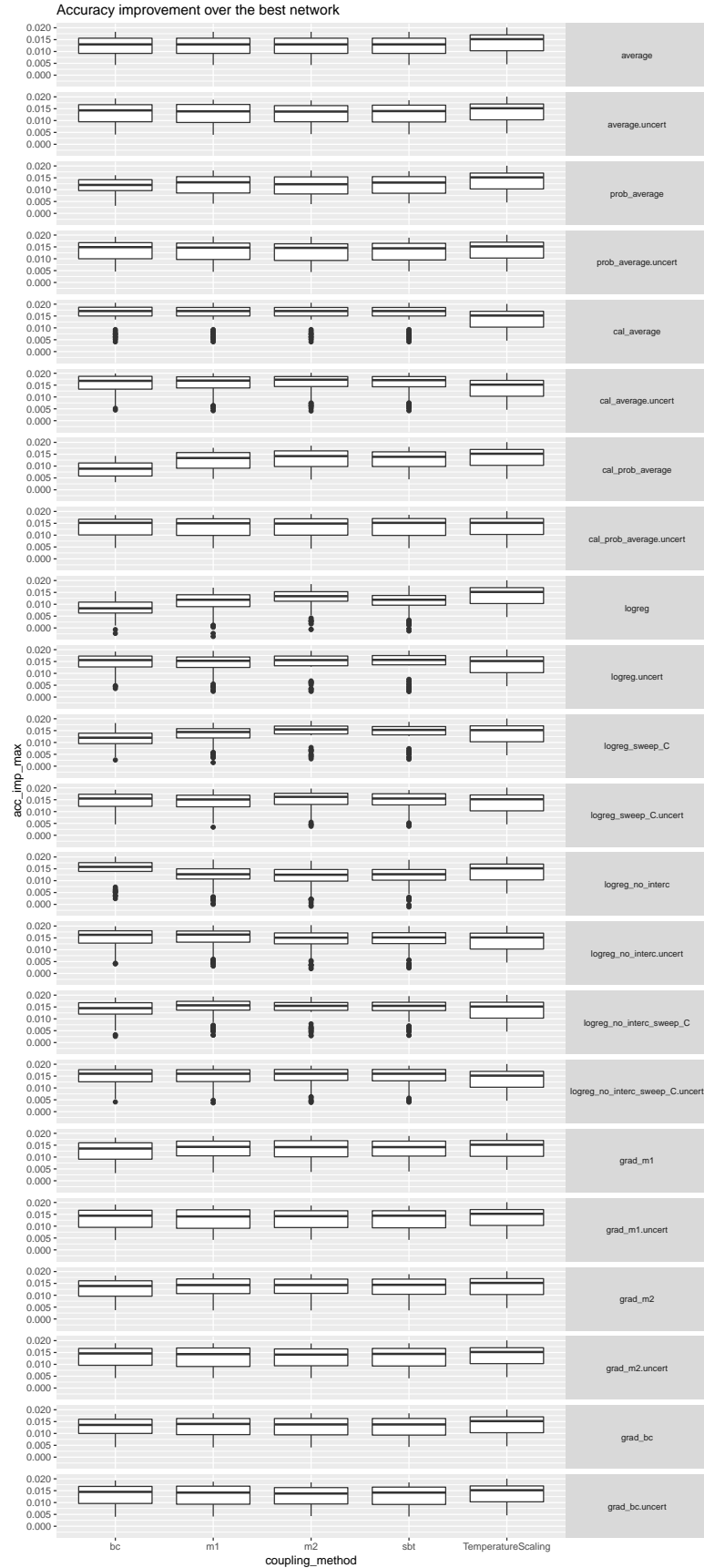
47	prob_average.uncert m1	0.0193	0.0132
48	grad_m2.uncert m1	0.0193	0.0132
49	grad_m1.uncert m1	0.0192	0.0131
50	average.uncert bc	0.0192	0.0131
51	prob_average.uncert sbt	0.0192	0.0131
52	grad_bc.uncert m1	0.0192	0.0131
53	grad_m2.uncert sbt	0.0192	0.0131
54	average.uncert m1	0.0191	0.0131
55	grad_m2.uncert m2	0.0191	0.0130
56	grad_m1.uncert sbt	0.0191	0.0130
57	prob_average.uncert m2	0.0191	0.0130
58	grad_bc.uncert sbt	0.0190	0.0130
59	grad_bc m1	0.0190	0.0129
60	grad_m1.uncert m2	0.0190	0.0129
61	average.uncert sbt	0.0190	0.0129
62	grad_bc.uncert m2	0.0190	0.0129
63	grad_bc sbt	0.0190	0.0129
64	grad_bc m2	0.0190	0.0129
65	cal_prob_average m2	0.0189	0.0129
66	grad_bc bc	0.0189	0.0128
67	grad_m2 bc	0.0189	0.0128
68	average.uncert m2	0.0189	0.0128
69	logreg_sweep_C m1	0.0188	0.0127
70	cal_prob_average sbt	0.0187	0.0126
71	grad_m1 bc	0.0186	0.0125
72	cal_prob_average m1	0.0184	0.0123
73	average bc	0.0184	0.0123
74	average m1	0.0184	0.0123
75	average m2	0.0184	0.0123
76	average sbt	0.0184	0.0123
77	prob_average sbt	0.0181	0.0120
78	logreg m2	0.0180	0.0119
79	prob_average m1	0.0180	0.0119
80	prob_average m2	0.0179	0.0119
81	logreg_no_interc m1	0.0178	0.0117
82	logreg_sweep_C bc	0.0176	0.0115
83	logreg_no_interc sbt	0.0175	0.0114
84	prob_average bc	0.0175	0.0114
85	logreg_no_interc m2	0.0173	0.0112
86	logreg sbt	0.0169	0.0108
87	logreg m1	0.0164	0.0103
88	cal_prob_average bc	0.0148	0.0087
89	logreg bc	0.0144	0.0083

```

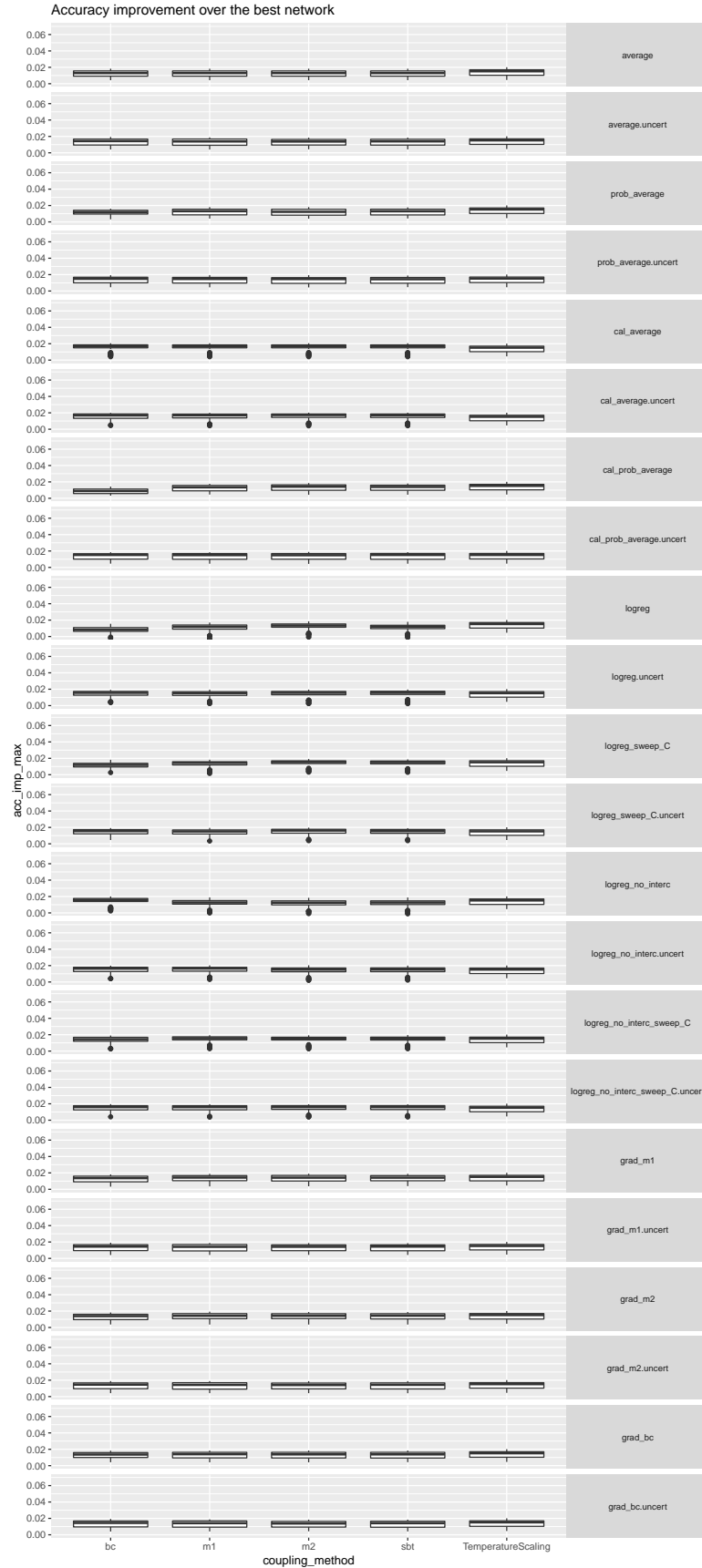
imp_max_plot <- ggplot() +
  geom_boxplot(data = ens_pwc_plt_df, mapping = aes(x = coupling_method, y = acc_imp_max)) +
  geom_boxplot(data = ens_cal_plt_df, mapping = aes(x = calibrating_method, y = acc_imp_max)) +
  facet_grid(rows = vars(combining_method)) +
  ggtitle("Accuracy improvement over the best network") +
  theme(strip.text.y = element_text(size = 8, angle = 0))

print(imp_max_plot)

```

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



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 =
```

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Thu Feb 10 10:49:14 2022

	method	imp_o_avg	imp_o_max
1	cal_average bc	0.0214	0.0153
2	cal_average m1	0.0214	0.0153
3	cal_average m2	0.0214	0.0153
4	cal_average sbt	0.0214	0.0153
5	cal_average.uncert m2	0.0213	0.0152
6	cal_average.uncert sbt	0.0213	0.0152
7	cal_average.uncert m1	0.0212	0.0151
8	cal_average.uncert bc	0.0210	0.0150
9	logreg_no_interc.uncert bc	0.0205	0.0144
10	logreg_no_interc.uncert m1	0.0204	0.0143
11	logreg_no_interc_sweep_C.uncert m1	0.0203	0.0143
12	logreg_no_interc bc	0.0203	0.0143
13	logreg_no_interc_sweep_C.uncert sbt	0.0203	0.0143
14	logreg_no_interc_sweep_C.uncert m2	0.0203	0.0142
15	logreg_sweep_C.uncert m2	0.0203	0.0142
16	logreg_no_interc_sweep_C.uncert bc	0.0203	0.0142
17	logreg_sweep_C.uncert sbt	0.0202	0.0141
18	logreg_no_interc_sweep_C m1	0.0201	0.0140
19	logreg.uncert bc	0.0201	0.0140
20	logreg_sweep_C.uncert bc	0.0200	0.0140
21	logreg_no_interc_sweep_C sbt	0.0200	0.0139
22	logreg_sweep_C m2	0.0200	0.0139
23	logreg_sweep_C.uncert m1	0.0200	0.0139
24	logreg.uncert sbt	0.0199	0.0138
25	logreg_no_interc_sweep_C m2	0.0199	0.0138
26	logreg.uncert m2	0.0198	0.0137
27	logreg.uncert m1	0.0197	0.0136
28	logreg_sweep_C sbt	0.0197	0.0136
29	logreg_no_interc.uncert sbt	0.0197	0.0136
30	logreg_no_interc.uncert m2	0.0196	0.0136
31	average of TemperatureScaling	0.0196	0.0135
32	logreg_no_interc_sweep_C bc	0.0196	0.0135
33	grad_m1 m1	0.0195	0.0134
34	grad_m2 sbt	0.0194	0.0134
35	grad_m2 m2	0.0194	0.0134
36	cal_prob_average.uncert bc	0.0194	0.0133
37	grad_m2 m1	0.0194	0.0133
38	cal_prob_average.uncert m1	0.0194	0.0133
39	prob_average.uncert bc	0.0194	0.0133
40	cal_prob_average.uncert sbt	0.0194	0.0133
41	cal_prob_average.uncert m2	0.0194	0.0133
42	grad_m1 m2	0.0194	0.0133
43	grad_m2.uncert bc	0.0193	0.0133
44	grad_m1 sbt	0.0193	0.0133
45	grad_bc.uncert bc	0.0193	0.0132
46	grad_m1.uncert bc	0.0193	0.0132

47	prob_average.uncert m1	0.0193	0.0132
48	grad_m2.uncert m1	0.0193	0.0132
49	grad_m1.uncert m1	0.0192	0.0131
50	average.uncert bc	0.0192	0.0131
51	prob_average.uncert sbt	0.0192	0.0131
52	grad_bc.uncert m1	0.0192	0.0131
53	grad_m2.uncert sbt	0.0192	0.0131
54	average.uncert m1	0.0191	0.0131
55	grad_m2.uncert m2	0.0191	0.0130
56	grad_m1.uncert sbt	0.0191	0.0130
57	prob_average.uncert m2	0.0191	0.0130
58	grad_bc.uncert sbt	0.0190	0.0130
59	grad_bc m1	0.0190	0.0129
60	grad_m1.uncert m2	0.0190	0.0129
61	average.uncert sbt	0.0190	0.0129
62	grad_bc.uncert m2	0.0190	0.0129
63	grad_bc sbt	0.0190	0.0129
64	grad_bc m2	0.0190	0.0129
65	cal_prob_average m2	0.0189	0.0129
66	grad_bc bc	0.0189	0.0128
67	grad_m2 bc	0.0189	0.0128
68	average.uncert m2	0.0189	0.0128
69	logreg_sweep_C m1	0.0188	0.0127
70	cal_prob_average sbt	0.0187	0.0126
71	grad_m1 bc	0.0186	0.0125
72	cal_prob_average m1	0.0184	0.0123
73	average bc	0.0184	0.0123
74	average m1	0.0184	0.0123
75	average m2	0.0184	0.0123
76	average sbt	0.0184	0.0123
77	prob_average sbt	0.0181	0.0120
78	logreg m2	0.0180	0.0119
79	prob_average m1	0.0180	0.0119
80	prob_average m2	0.0179	0.0119
81	logreg_no_interc m1	0.0178	0.0117
82	logreg_sweep_C bc	0.0176	0.0115
83	logreg_no_interc sbt	0.0175	0.0114
84	prob_average bc	0.0175	0.0114
85	logreg_no_interc m2	0.0173	0.0112
86	logreg sbt	0.0169	0.0108
87	logreg m1	0.0164	0.0103
88	cal_prob_average bc	0.0148	0.0087
89	logreg bc	0.0144	0.0083

```

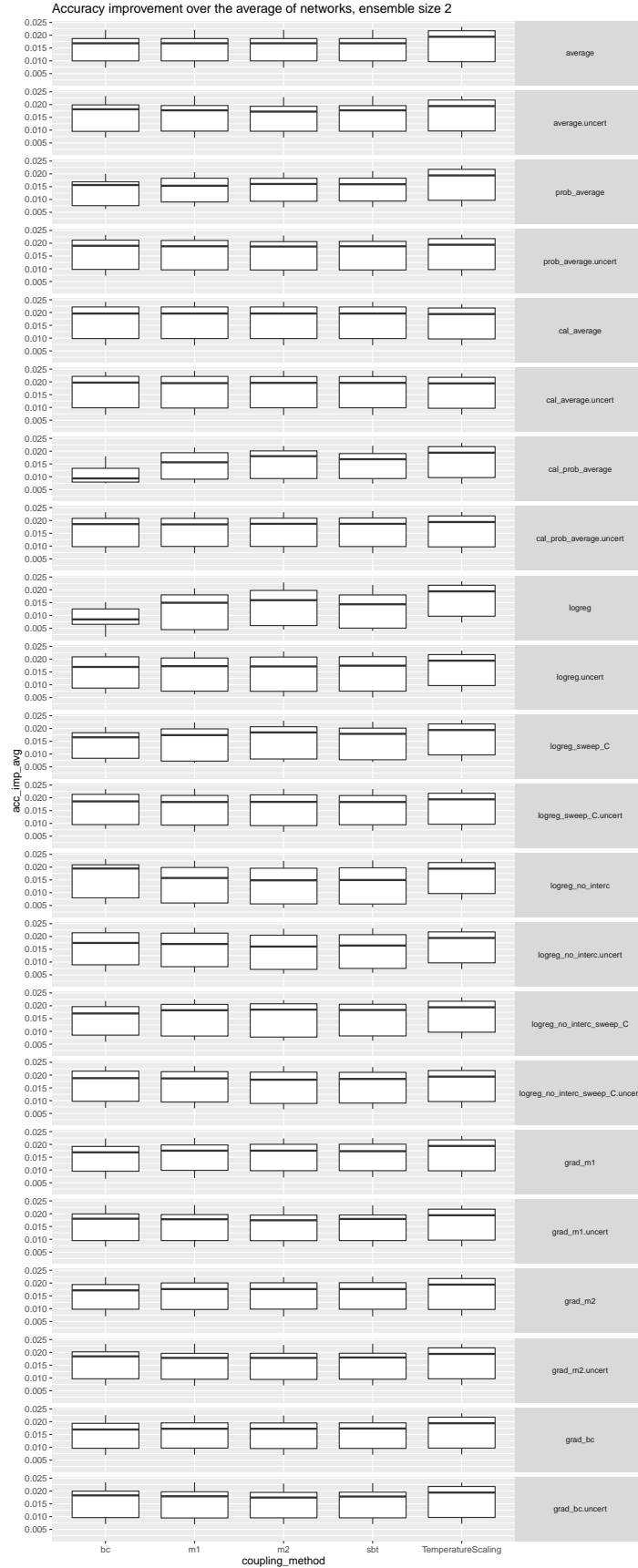
avg_imp_table_cs <- rbind(
  ens_pwc_plt_df %>%
    mutate(method = paste(combining_method, coupling_method, sep = " ")) %>%
    group_by(method, combination_size) %>%
    summarise(imp_o_avg = mean(acc_imp_avg), imp_o_max = mean(acc_imp_max)),
  ens_cal_plt_df %>%
    mutate(method = paste0("average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

```

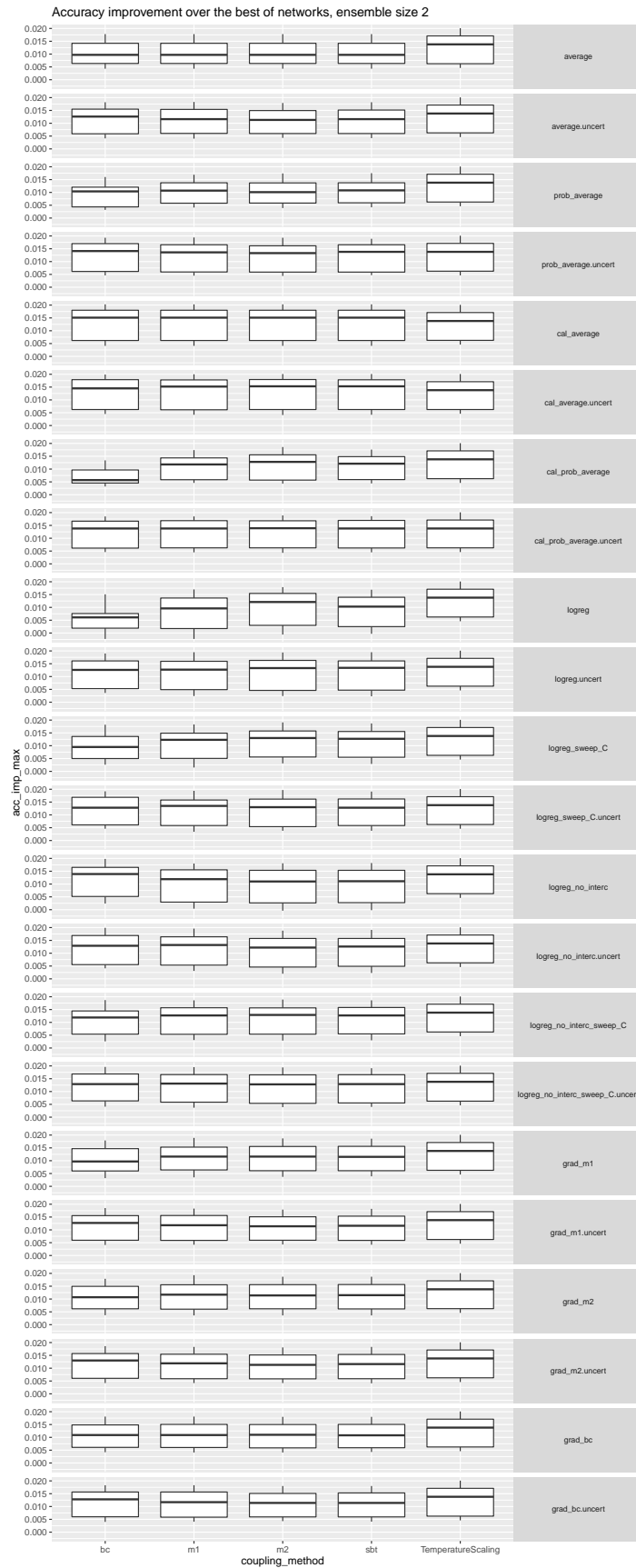
```
)
```

```
## '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.
```

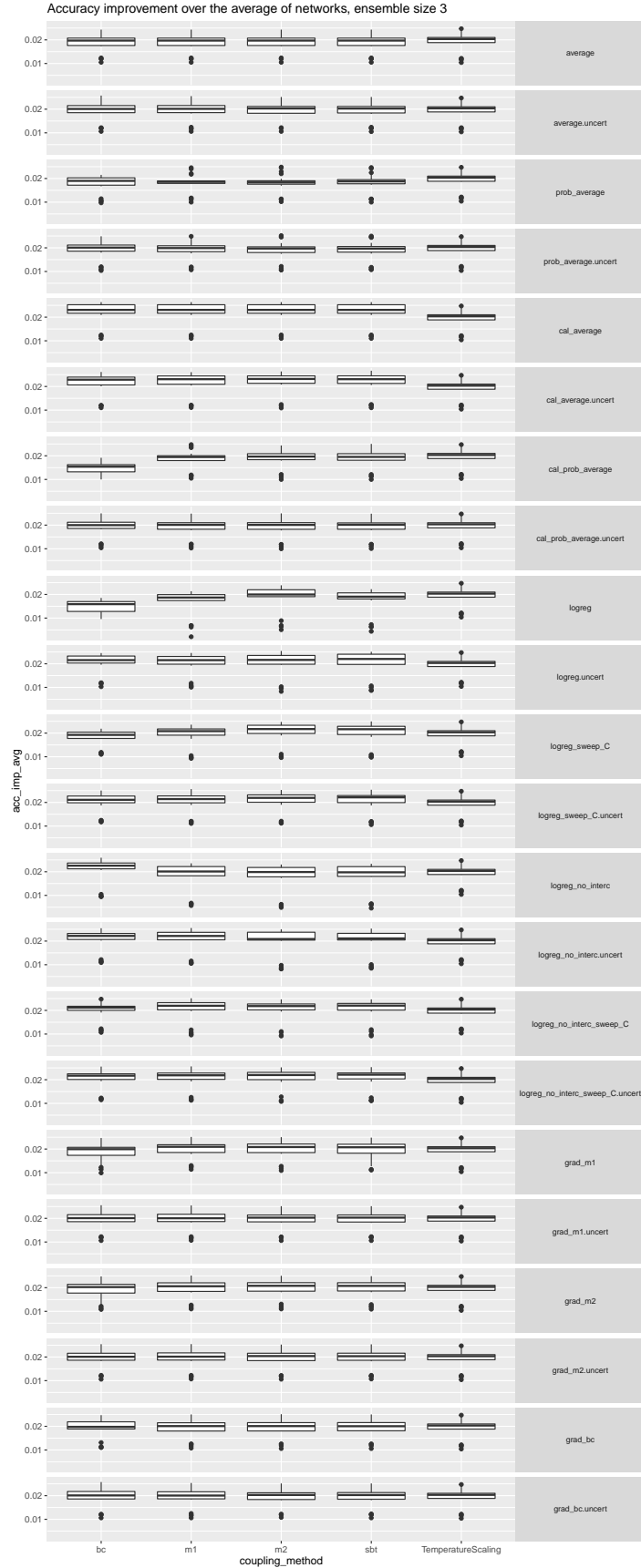
```
for (sss in unique(ens_cal_plt_df$combination_size))  
{  
  imp_avg_plot <- ggplot() +  
    geom_boxplot(data = ens_pwc_plt_df %>% filter(combination_size == sss), mapping = aes(x = coupling_r, y = accuracy),  
                 data2 = ens_cal_plt_df %>% filter(combination_size == sss), mapping = aes(x = calibration_r, y = accuracy)) +  
    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_r, y = accuracy),  
                 data2 = ens_cal_plt_df %>% filter(combination_size == sss), mapping = aes(x = calibration_r, y = accuracy)) +  
    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)))  
}
```

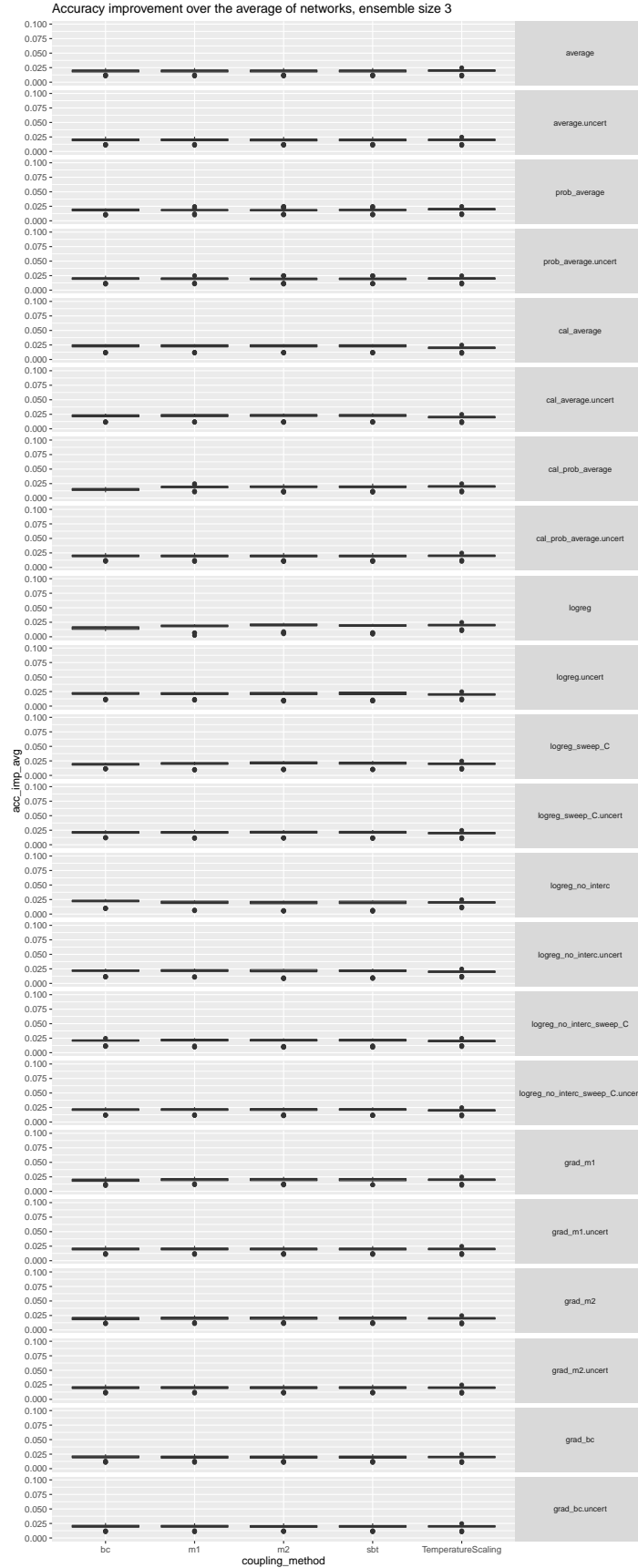


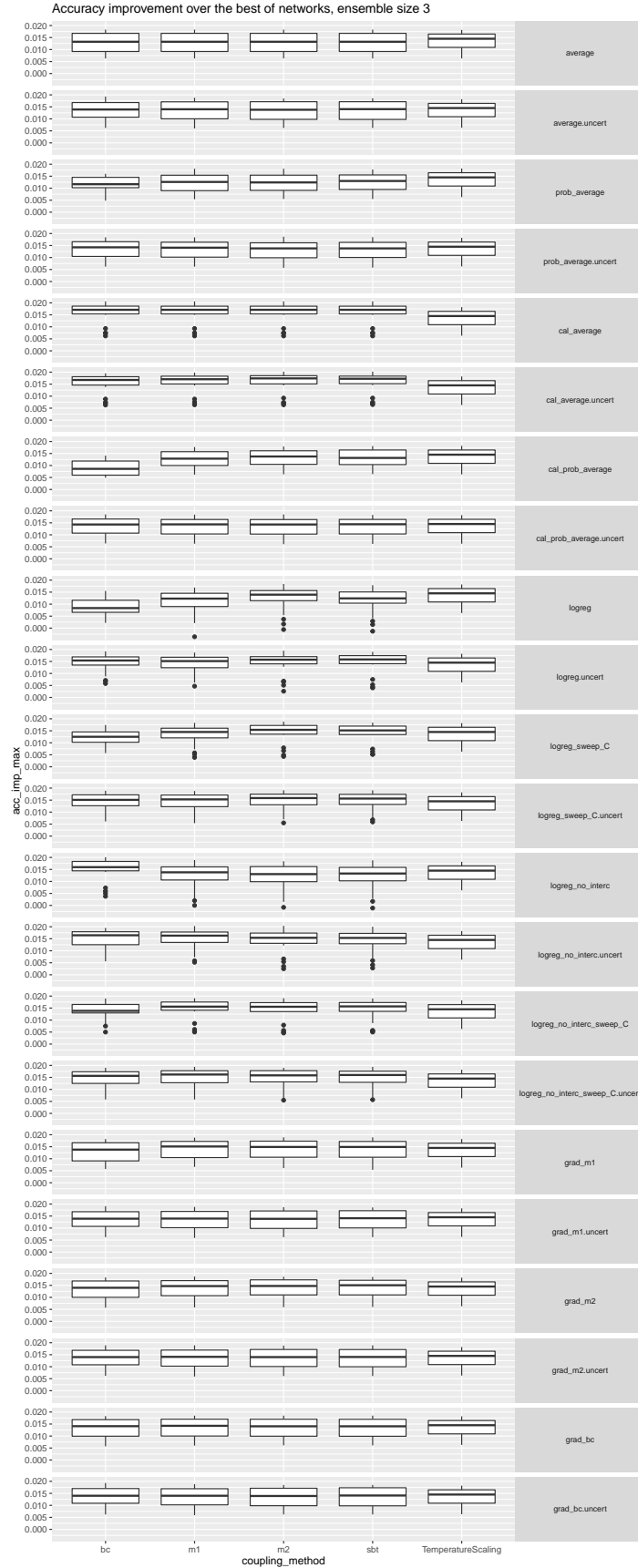


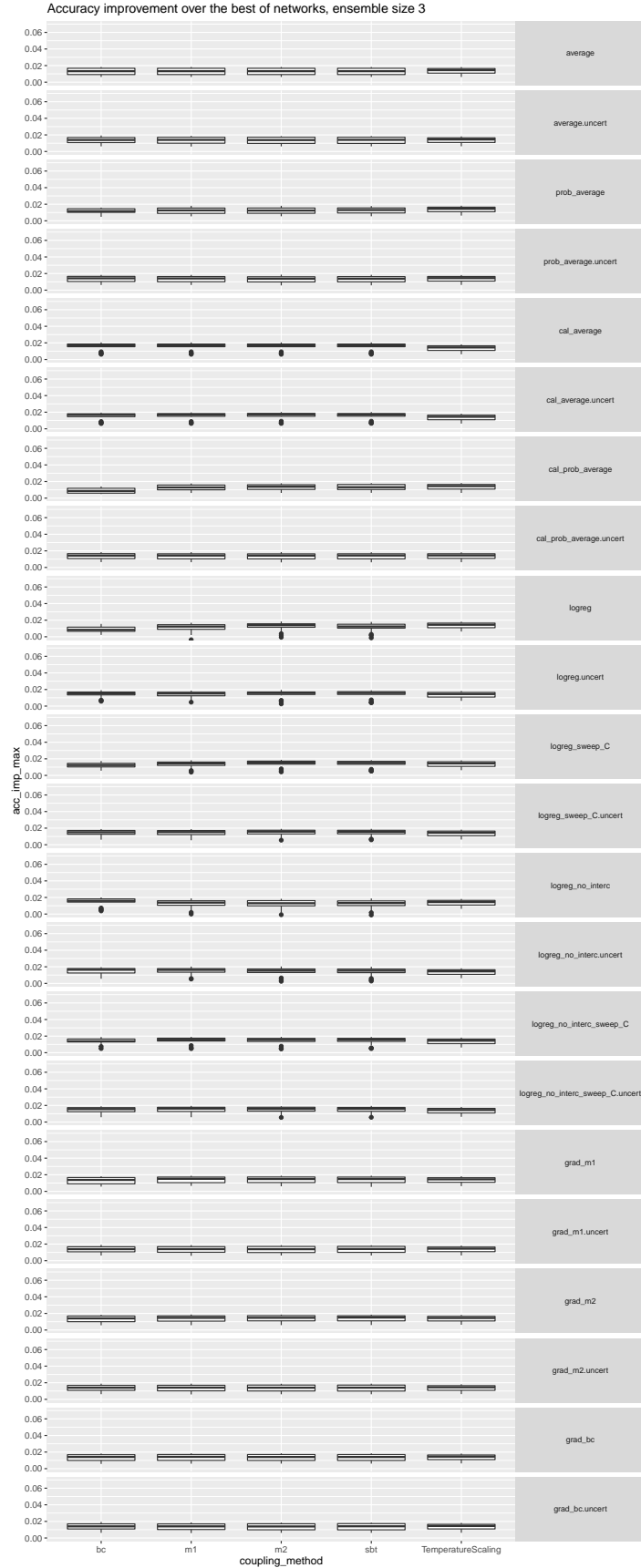


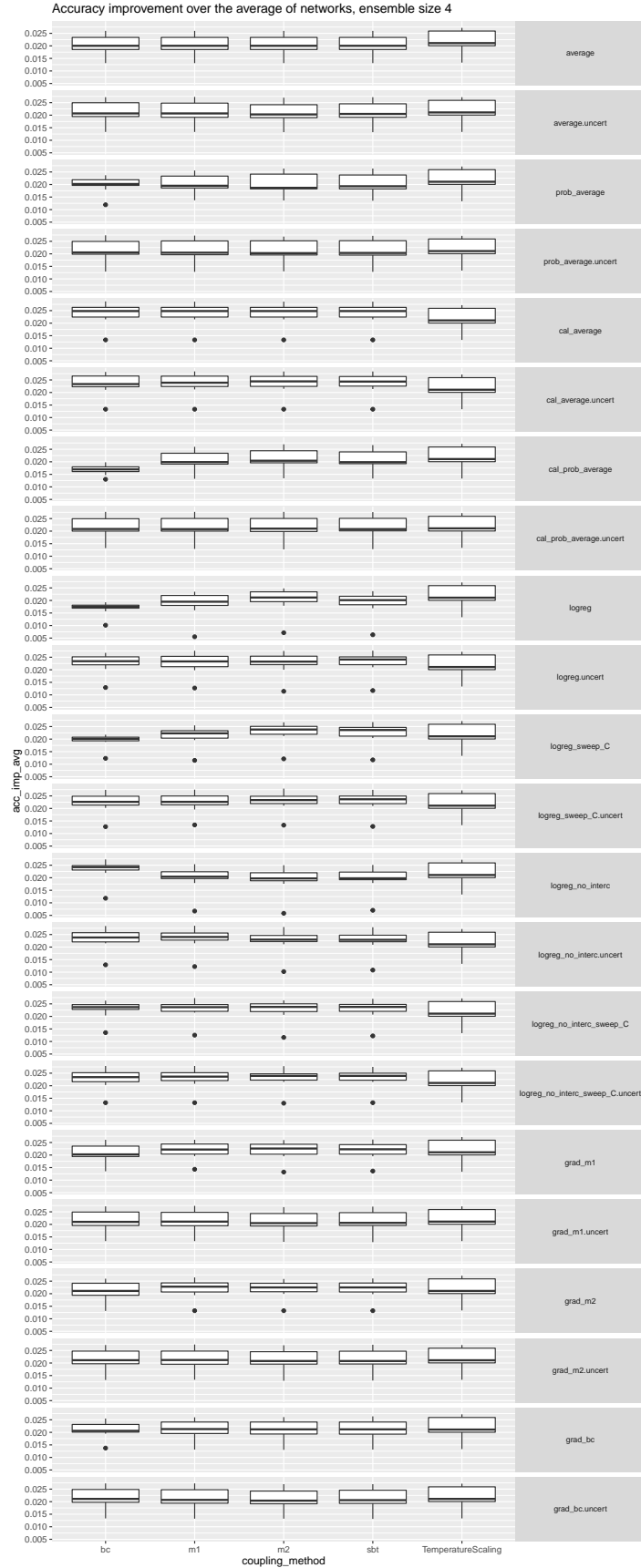


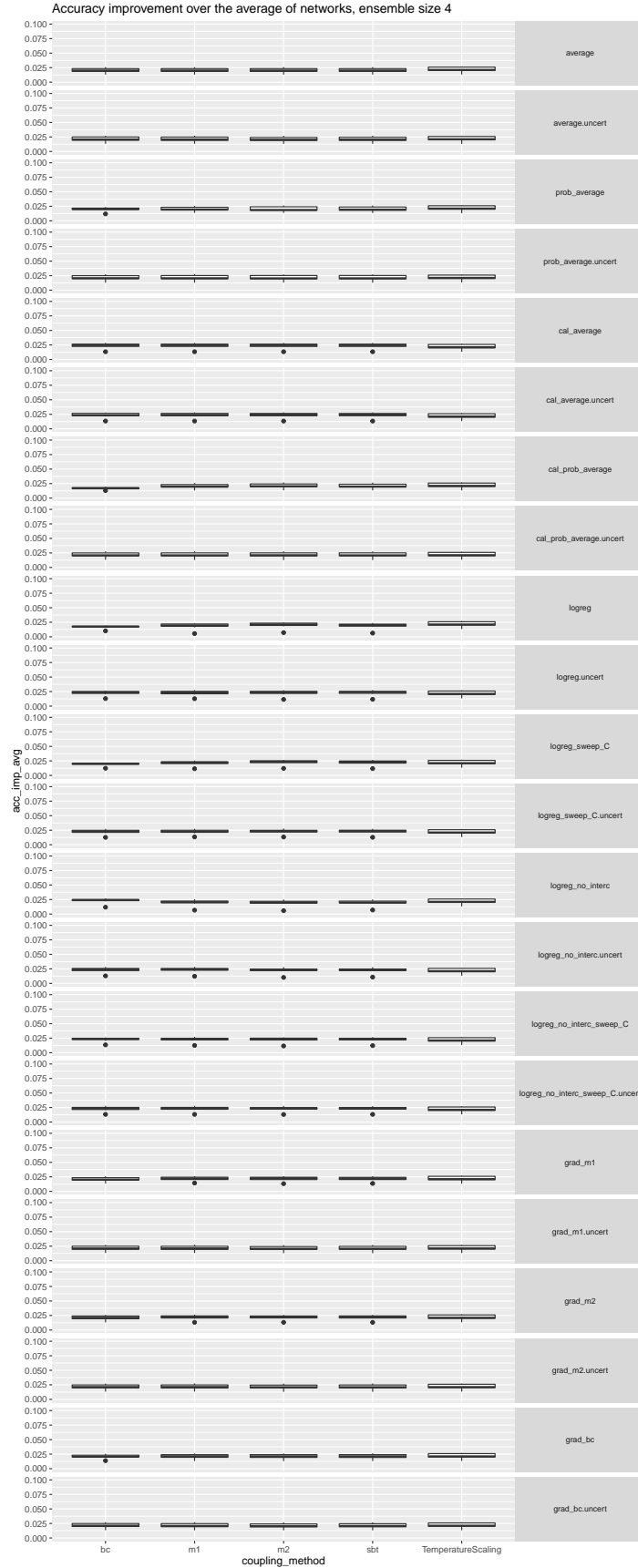


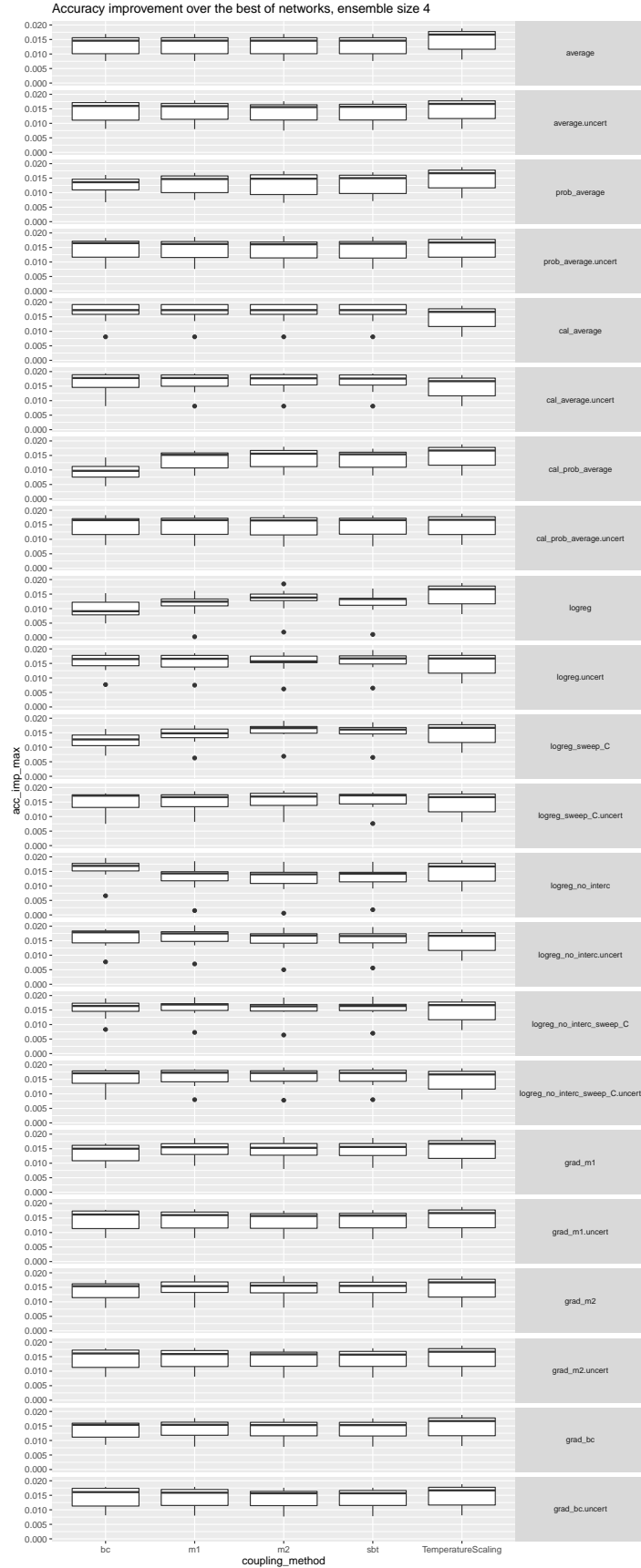


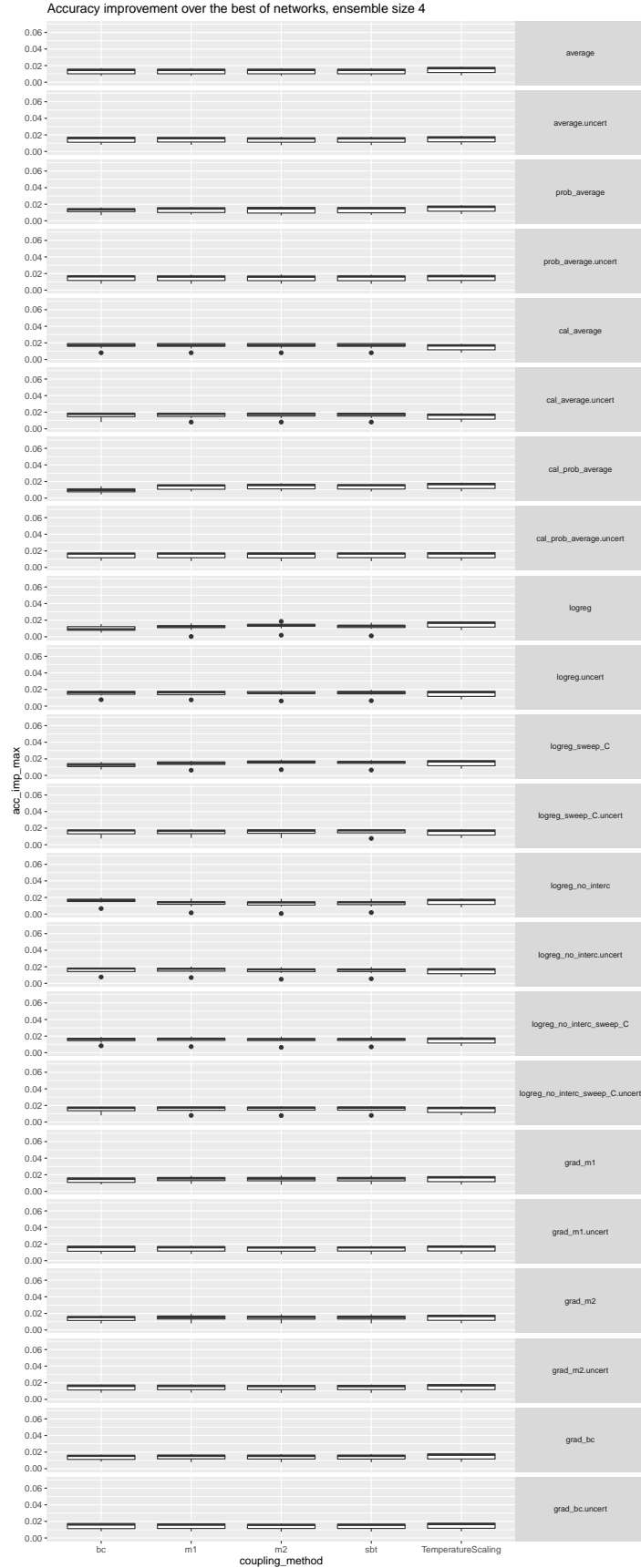


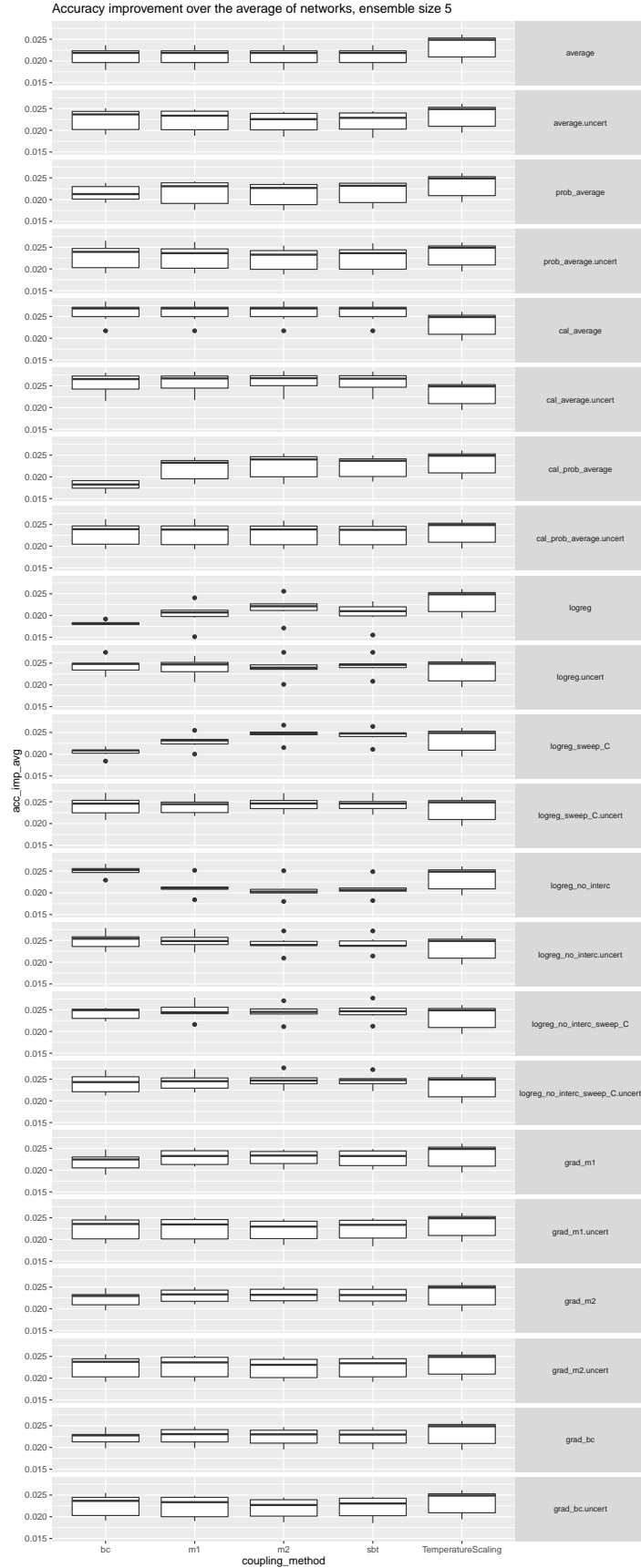


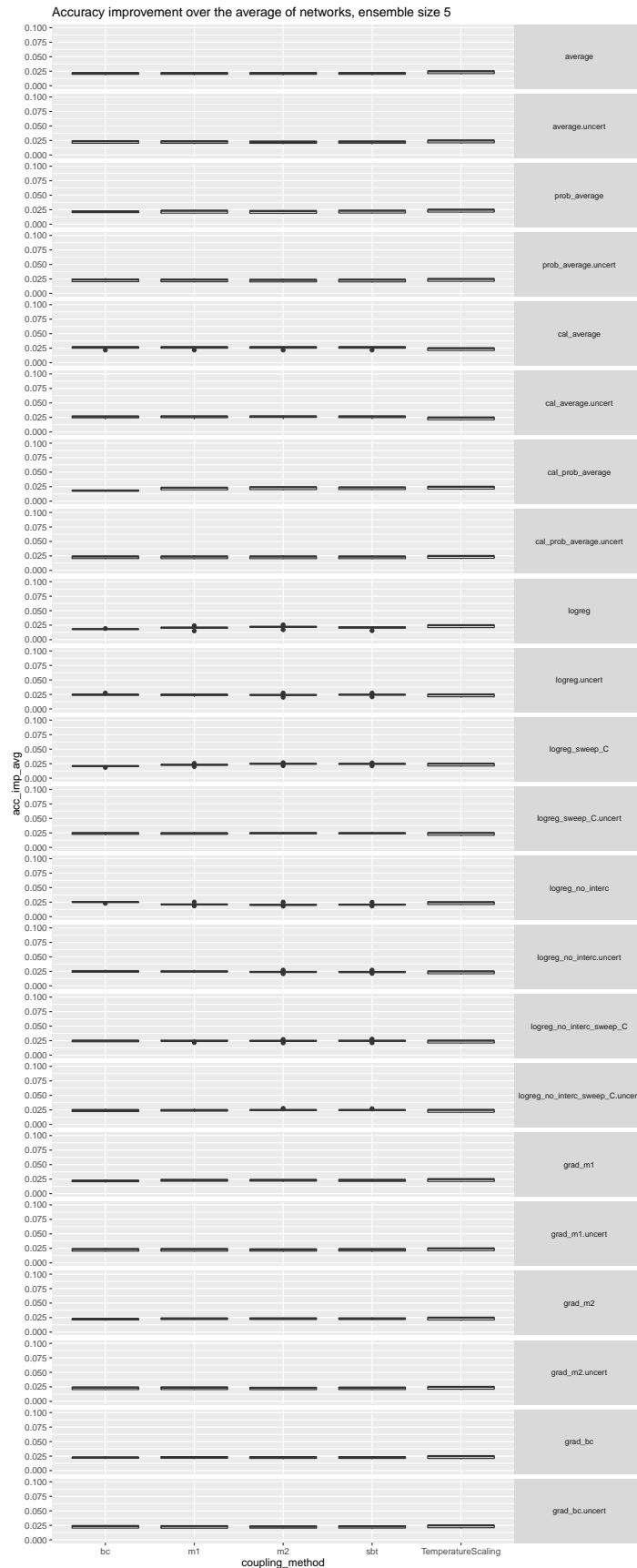




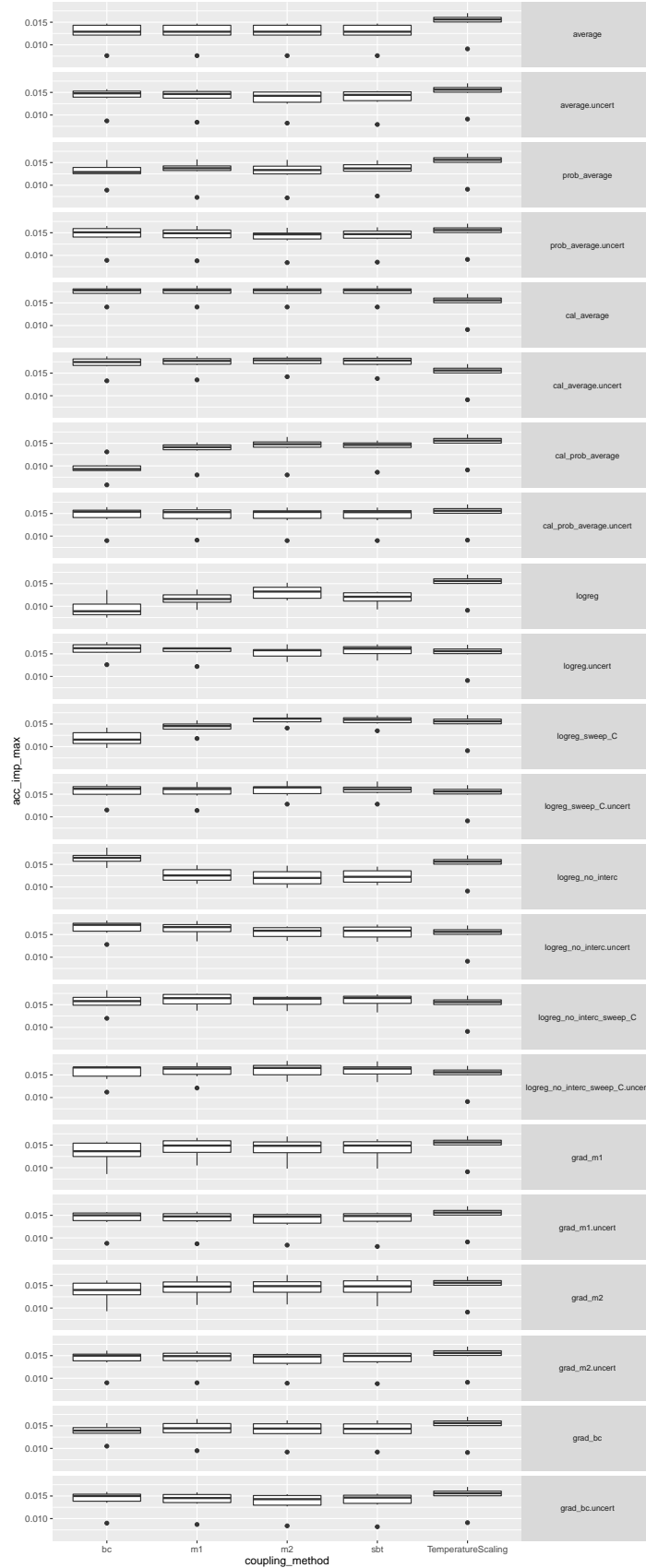




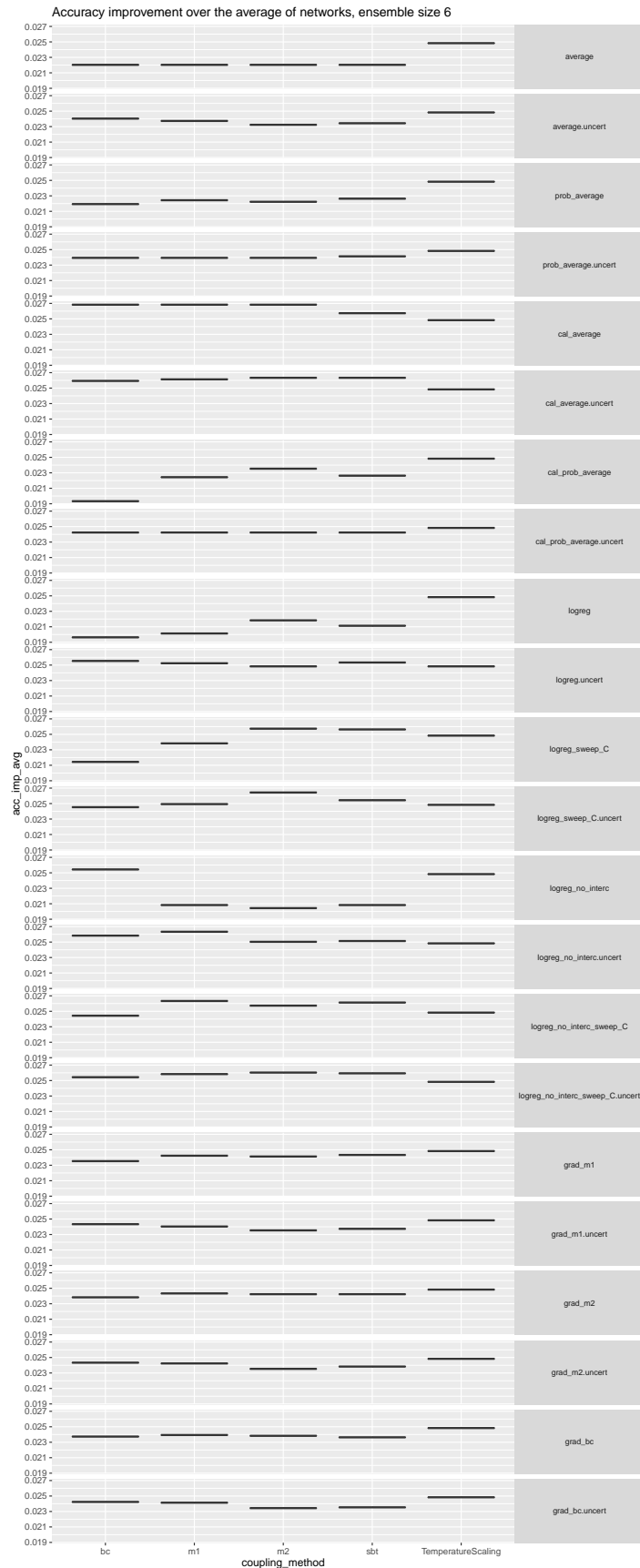




Accuracy improvement over the best of networks, ensemble size 5













```

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=2))
}

```

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Thu Feb 10 10:49:55 2022

	method	combination_size	imp_o_avg	imp_o_max
1	cal_average.uncert m2	2	0.0162	0.0124
2	cal_average.uncert m1	2	0.0162	0.0124
3	cal_average.uncert bc	2	0.0162	0.0124
4	cal_average.uncert sbt	2	0.0162	0.0124
5	cal_average bc	2	0.0162	0.0123
6	cal_average m1	2	0.0162	0.0123
7	cal_average m2	2	0.0162	0.0123
8	cal_average sbt	2	0.0162	0.0123
9	average of TemperatureScaling	2	0.0158	0.0120
10	prob_average.uncert bc	2	0.0156	0.0118
11	cal_prob_average.uncert sbt	2	0.0156	0.0118
12	logreg_no_interc_sweep_C.uncert bc	2	0.0156	0.0117
13	logreg_sweep_C.uncert bc	2	0.0155	0.0117
14	cal_prob_average.uncert m1	2	0.0155	0.0117
15	cal_prob_average.uncert m2	2	0.0155	0.0117
16	cal_prob_average.uncert bc	2	0.0155	0.0117
17	prob_average.uncert m1	2	0.0155	0.0116
18	prob_average.uncert sbt	2	0.0154	0.0116
19	logreg_no_interc_sweep_C.uncert m1	2	0.0154	0.0116
20	prob_average.uncert m2	2	0.0152	0.0114
21	logreg_no_interc_sweep_C.uncert sbt	2	0.0152	0.0114
22	grad_m2.uncert bc	2	0.0151	0.0113
23	logreg_sweep_C.uncert sbt	2	0.0151	0.0113
24	logreg_no_interc.uncert bc	2	0.0151	0.0113
25	logreg_sweep_C.uncert m2	2	0.0151	0.0113
26	logreg_no_interc_sweep_C.uncert m2	2	0.0151	0.0113
27	logreg_sweep_C.uncert m1	2	0.0151	0.0113
28	grad_m1.uncert bc	2	0.0150	0.0112
29	grad_bc.uncert bc	2	0.0150	0.0112
30	average.uncert bc	2	0.0150	0.0112
31	grad_m1.uncert m1	2	0.0149	0.0111
32	grad_m2.uncert m1	2	0.0149	0.0111
33	grad_m1 m1	2	0.0149	0.0111
34	grad_bc.uncert m1	2	0.0149	0.0111
35	average.uncert m1	2	0.0149	0.0111
36	grad_m2.uncert sbt	2	0.0148	0.0110
37	cal_prob_average m2	2	0.0148	0.0110
38	grad_m2 sbt	2	0.0148	0.0110
39	logreg_no_interc.uncert m1	2	0.0148	0.0110
40	logreg.uncert bc	2	0.0148	0.0110
41	grad_m1.uncert sbt	2	0.0148	0.0110
42	grad_m2 m1	2	0.0148	0.0110
43	grad_m1 m2	2	0.0148	0.0110
44	grad_m2 m2	2	0.0148	0.0110
45	grad_m1 sbt	2	0.0148	0.0110

46	average.uncert sbt	2	0.0148	0.0110
47	grad_bc.uncert sbt	2	0.0148	0.0110
48	logreg_no_interc bc	2	0.0148	0.0110
49	grad_m2.uncert m2	2	0.0148	0.0109
50	grad_bc.uncert m2	2	0.0147	0.0109
51	grad_m1.uncert m2	2	0.0147	0.0109
52	logreg_sweep_C m2	2	0.0147	0.0109
53	average.uncert m2	2	0.0147	0.0108
54	logreg_no_interc_sweep_C m1	2	0.0146	0.0108
55	grad_bc m1	2	0.0146	0.0108
56	grad_bc sbt	2	0.0146	0.0108
57	logreg_no_interc_sweep_C sbt	2	0.0146	0.0108
58	grad_bc m2	2	0.0146	0.0108
59	grad_m2 bc	2	0.0145	0.0107
60	grad_bc bc	2	0.0145	0.0107
61	logreg_no_interc_sweep_C m2	2	0.0145	0.0107
62	logreg_sweep_C sbt	2	0.0144	0.0106
63	logreg.uncert m2	2	0.0144	0.0106
64	cal_prob_average sbt	2	0.0144	0.0106
65	logreg_no_interc.uncert sbt	2	0.0144	0.0106
66	logreg.uncert sbt	2	0.0144	0.0106
67	logreg.uncert m1	2	0.0144	0.0105
68	average bc	2	0.0143	0.0105
69	average m1	2	0.0143	0.0105
70	average m2	2	0.0143	0.0105
71	average sbt	2	0.0143	0.0105
72	grad_m1 bc	2	0.0143	0.0105
73	logreg_no_interc.uncert m2	2	0.0142	0.0104
74	cal_prob_average m1	2	0.0142	0.0104
75	logreg_no_interc_sweep_C bc	2	0.0140	0.0102
76	logreg_sweep_C m1	2	0.0140	0.0101
77	prob_average sbt	2	0.0139	0.0101
78	prob_average m2	2	0.0137	0.0099
79	prob_average m1	2	0.0137	0.0099
80	logreg_sweep_C bc	2	0.0135	0.0097
81	logreg_no_interc m1	2	0.0134	0.0096
82	logreg m2	2	0.0133	0.0095
83	logreg_no_interc sbt	2	0.0132	0.0094
84	logreg_no_interc m2	2	0.0131	0.0093
85	prob_average bc	2	0.0126	0.0088
86	logreg sbt	2	0.0122	0.0084
87	logreg m1	2	0.0118	0.0080
88	cal_prob_average bc	2	0.0110	0.0072
89	logreg bc	2	0.0092	0.0054

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Thu Feb 10 10:49:55 2022

	method	combination_size	imp_o_avg	imp_o_max
1	cal_average bc	3	0.0215	0.0156
2	cal_average m1	3	0.0215	0.0156
3	cal_average m2	3	0.0215	0.0156
4	cal_average sbt	3	0.0215	0.0156
5	cal_average.uncert sbt	3	0.0214	0.0156

6	cal_average.uncert m2	3	0.0214	0.0156
7	cal_average.uncert m1	3	0.0213	0.0154
8	cal_average.uncert bc	3	0.0210	0.0152
9	logreg_no_interc bc	3	0.0205	0.0147
10	logreg_no_interc.uncert m1	3	0.0205	0.0147
11	logreg_no_interc.uncert bc	3	0.0205	0.0146
12	logreg_no_interc_sweep_C.uncert m1	3	0.0204	0.0145
13	logreg_no_interc_sweep_C.uncert m2	3	0.0203	0.0145
14	logreg_no_interc_sweep_C.uncert sbt	3	0.0203	0.0145
15	logreg_sweep_C.uncert m2	3	0.0203	0.0144
16	logreg_sweep_C.uncert sbt	3	0.0202	0.0144
17	logreg_no_interc_sweep_C m1	3	0.0202	0.0143
18	logreg_no_interc_sweep_C.uncert bc	3	0.0202	0.0143
19	logreg.uncert bc	3	0.0202	0.0143
20	logreg_no_interc_sweep_C sbt	3	0.0201	0.0142
21	logreg.uncert sbt	3	0.0200	0.0141
22	logreg_sweep_C.uncert bc	3	0.0199	0.0141
23	logreg_sweep_C m2	3	0.0199	0.0141
24	logreg_sweep_C.uncert m1	3	0.0199	0.0141
25	logreg_no_interc_sweep_C m2	3	0.0199	0.0141
26	logreg.uncert m2	3	0.0199	0.0140
27	logreg.uncert m1	3	0.0198	0.0139
28	logreg_no_interc.uncert sbt	3	0.0198	0.0139
29	logreg_no_interc.uncert m2	3	0.0198	0.0139
30	logreg_sweep_C sbt	3	0.0197	0.0139
31	grad_m1 m1	3	0.0197	0.0138
32	logreg_no_interc_sweep_C bc	3	0.0196	0.0138
33	grad_m1 m2	3	0.0196	0.0137
34	grad_m2 m2	3	0.0196	0.0137
35	grad_m2 sbt	3	0.0196	0.0137
36	grad_m1 sbt	3	0.0195	0.0137
37	grad_m2 m1	3	0.0195	0.0136
38	grad_bc.uncert bc	3	0.0194	0.0136
39	grad_m2.uncert bc	3	0.0194	0.0135
40	grad_m2.uncert m1	3	0.0194	0.0135
41	grad_m2.uncert sbt	3	0.0193	0.0135
42	grad_m1.uncert m1	3	0.0193	0.0135
43	grad_m2.uncert m2	3	0.0193	0.0135
44	grad_m1.uncert bc	3	0.0193	0.0135
45	average.uncert m1	3	0.0193	0.0135
46	grad_bc.uncert m1	3	0.0193	0.0135
47	grad_m1.uncert sbt	3	0.0193	0.0134
48	average.uncert bc	3	0.0193	0.0134
49	grad_bc.uncert sbt	3	0.0193	0.0134
50	grad_m1.uncert m2	3	0.0193	0.0134
51	grad_bc.uncert m2	3	0.0192	0.0134
52	average.uncert sbt	3	0.0192	0.0134
53	grad_bc m1	3	0.0192	0.0134
54	grad_bc m2	3	0.0192	0.0133
55	average.uncert m2	3	0.0192	0.0133
56	grad_bc sbt	3	0.0192	0.0133
57	grad_bc bc	3	0.0192	0.0133
58	grad_m2 bc	3	0.0191	0.0133
59	cal_prob_average.uncert bc	3	0.0191	0.0132

60	average of TemperatureScaling	3	0.0191	0.0132
61	cal_prob_average.uncert m1	3	0.0190	0.0132
62	prob_average.uncert bc	3	0.0190	0.0132
63	cal_prob_average.uncert sbt	3	0.0190	0.0132
64	cal_prob_average.uncert m2	3	0.0190	0.0131
65	prob_average.uncert m1	3	0.0189	0.0130
66	logreg_sweep_C m1	3	0.0189	0.0130
67	prob_average.uncert sbt	3	0.0188	0.0129
68	prob_average.uncert m2	3	0.0188	0.0129
69	grad_m1 bc	3	0.0187	0.0129
70	cal_prob_average m2	3	0.0187	0.0129
71	average bc	3	0.0187	0.0128
72	average m1	3	0.0187	0.0128
73	average m2	3	0.0187	0.0128
74	average sbt	3	0.0187	0.0128
75	cal_prob_average sbt	3	0.0187	0.0128
76	cal_prob_average m1	3	0.0184	0.0125
77	logreg m2	3	0.0182	0.0123
78	logreg_sweep_C bc	3	0.0181	0.0122
79	prob_average sbt	3	0.0180	0.0122
80	logreg_no_interc m1	3	0.0180	0.0122
81	prob_average m1	3	0.0180	0.0121
82	prob_average m2	3	0.0179	0.0120
83	logreg_no_interc sbt	3	0.0177	0.0119
84	prob_average bc	3	0.0176	0.0117
85	logreg_no_interc m2	3	0.0175	0.0117
86	logreg sbt	3	0.0172	0.0113
87	logreg m1	3	0.0165	0.0107
88	cal_prob_average bc	3	0.0148	0.0089
89	logreg bc	3	0.0148	0.0089

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Thu Feb 10 10:49:55 2022

	method	combination_size	imp_o_avg	imp_o_max
1	cal_average bc	4	0.0242	0.0169
2	cal_average m1	4	0.0242	0.0169
3	cal_average m2	4	0.0242	0.0169
4	cal_average sbt	4	0.0242	0.0169
5	cal_average.uncert m2	4	0.0241	0.0167
6	cal_average.uncert sbt	4	0.0240	0.0167
7	cal_average.uncert m1	4	0.0239	0.0166
8	cal_average.uncert bc	4	0.0238	0.0164
9	logreg_no_interc.uncert bc	4	0.0237	0.0164
10	logreg_no_interc.uncert m1	4	0.0237	0.0163
11	logreg_no_interc bc	4	0.0235	0.0161
12	logreg_no_interc_sweep_C.uncert sbt	4	0.0234	0.0161
13	logreg_no_interc_sweep_C.uncert m2	4	0.0234	0.0160
14	logreg_no_interc_sweep_C.uncert m1	4	0.0233	0.0160
15	logreg_sweep_C.uncert m2	4	0.0233	0.0159
16	logreg.uncert sbt	4	0.0232	0.0159
17	logreg_sweep_C.uncert sbt	4	0.0232	0.0159
18	logreg_no_interc_sweep_C m1	4	0.0232	0.0159
19	logreg_no_interc_sweep_C.uncert bc	4	0.0232	0.0159

20	logreg_sweep_C m2	4	0.0231	0.0158
21	logreg.uncert bc	4	0.0231	0.0158
22	logreg.uncert m2	4	0.0231	0.0157
23	logreg_no_interc_sweep_C sbt	4	0.0231	0.0157
24	logreg_no_interc_sweep_C bc	4	0.0230	0.0157
25	logreg_no_interc_sweep_C m2	4	0.0230	0.0156
26	logreg.uncert m1	4	0.0229	0.0155
27	logreg_sweep_C.uncert m1	4	0.0228	0.0155
28	logreg_sweep_C.uncert bc	4	0.0228	0.0154
29	logreg_sweep_C sbt	4	0.0228	0.0154
30	logreg_no_interc.uncert sbt	4	0.0228	0.0154
31	logreg_no_interc.uncert m2	4	0.0227	0.0154
32	average of TemperatureScaling	4	0.0223	0.0149
33	grad_m2 m1	4	0.0221	0.0148
34	grad_m2 sbt	4	0.0221	0.0147
35	grad_m2 m2	4	0.0221	0.0147
36	cal_prob_average.uncert bc	4	0.0221	0.0147
37	cal_prob_average.uncert m1	4	0.0221	0.0147
38	cal_prob_average.uncert sbt	4	0.0220	0.0147
39	grad_m1 m1	4	0.0220	0.0146
40	cal_prob_average.uncert m2	4	0.0220	0.0146
41	prob_average.uncert bc	4	0.0220	0.0146
42	grad_m1 sbt	4	0.0219	0.0146
43	prob_average.uncert m1	4	0.0219	0.0145
44	grad_m1 m2	4	0.0219	0.0145
45	grad_bc.uncert bc	4	0.0219	0.0145
46	grad_m2.uncert bc	4	0.0218	0.0145
47	grad_m1.uncert bc	4	0.0218	0.0145
48	grad_m2.uncert m1	4	0.0218	0.0144
49	prob_average.uncert sbt	4	0.0218	0.0144
50	prob_average.uncert m2	4	0.0218	0.0144
51	grad_m1.uncert m1	4	0.0218	0.0144
52	average.uncert bc	4	0.0217	0.0144
53	grad_bc.uncert m1	4	0.0217	0.0143
54	logreg_sweep_C m1	4	0.0216	0.0143
55	grad_m2.uncert sbt	4	0.0216	0.0143
56	cal_prob_average m2	4	0.0216	0.0142
57	average.uncert m1	4	0.0216	0.0142
58	grad_m2.uncert m2	4	0.0216	0.0142
59	grad_m1.uncert sbt	4	0.0215	0.0141
60	grad_bc.uncert sbt	4	0.0214	0.0141
61	grad_bc m1	4	0.0214	0.0140
62	grad_m1.uncert m2	4	0.0214	0.0140
63	grad_bc sbt	4	0.0214	0.0140
64	grad_bc m2	4	0.0213	0.0140
65	grad_bc.uncert m2	4	0.0213	0.0140
66	average.uncert sbt	4	0.0213	0.0140
67	grad_bc bc	4	0.0213	0.0139
68	grad_m2 bc	4	0.0213	0.0139
69	cal_prob_average sbt	4	0.0213	0.0139
70	average.uncert m2	4	0.0212	0.0139
71	grad_m1 bc	4	0.0210	0.0137
72	cal_prob_average m1	4	0.0209	0.0136
73	logreg m2	4	0.0207	0.0133

74	prob_ average sbt	4	0.0207	0.0133
75	average bc	4	0.0206	0.0133
76	average m1	4	0.0206	0.0133
77	average m2	4	0.0206	0.0133
78	average sbt	4	0.0206	0.0133
79	prob_ average m1	4	0.0206	0.0132
80	prob_ average m2	4	0.0206	0.0132
81	logreg_ no_ interc m1	4	0.0202	0.0129
82	prob_ average bc	4	0.0202	0.0128
83	logreg_ no_ interc sbt	4	0.0199	0.0126
84	logreg_ sweep_ C bc	4	0.0197	0.0123
85	logreg_ no_ interc m2	4	0.0197	0.0123
86	logreg sbt	4	0.0194	0.0121
87	logreg m1	4	0.0189	0.0116
88	logreg bc	4	0.0171	0.0098
89	cal_ prob_ average bc	4	0.0169	0.0095

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Thu Feb 10 10:49:55 2022

	method	combination_size	imp_o_avg	imp_o_max
1	cal_ average.uncert m2	5	0.0260	0.0173
2	cal_ average bc	5	0.0259	0.0173
3	cal_ average m1	5	0.0259	0.0173
4	cal_ average m2	5	0.0259	0.0173
5	cal_ average sbt	5	0.0259	0.0173
6	cal_ average.uncert sbt	5	0.0259	0.0172
7	cal_ average.uncert m1	5	0.0258	0.0171
8	cal_ average.uncert bc	5	0.0256	0.0169
9	logreg_ no_ interc bc	5	0.0251	0.0164
10	logreg_ no_ interc.uncert bc	5	0.0250	0.0163
11	logreg_ no_ interc.uncert m1	5	0.0249	0.0162
12	logreg_ no_ interc_ sweep_ C.uncert m2	5	0.0247	0.0161
13	logreg_ no_ interc_ sweep_ C m1	5	0.0247	0.0160
14	logreg_ no_ interc_ sweep_ C.uncert sbt	5	0.0246	0.0160
15	logreg_ no_ interc_ sweep_ C sbt	5	0.0246	0.0159
16	logreg_ sweep_ C m2	5	0.0246	0.0159
17	logreg.uncert bc	5	0.0245	0.0158
18	logreg_ sweep_ C.uncert m2	5	0.0245	0.0158
19	logreg_ sweep_ C.uncert sbt	5	0.0244	0.0158
20	logreg_ no_ interc_ sweep_ C m2	5	0.0244	0.0158
21	logreg.uncert sbt	5	0.0244	0.0157
22	logreg_ no_ interc_ sweep_ C.uncert m1	5	0.0244	0.0157
23	logreg_ sweep_ C sbt	5	0.0243	0.0156
24	logreg_ no_ interc_ sweep_ C bc	5	0.0242	0.0155
25	logreg_ no_ interc.uncert sbt	5	0.0242	0.0155
26	logreg_ no_ interc.uncert m2	5	0.0242	0.0155
27	logreg.uncert m1	5	0.0241	0.0154
28	logreg_ sweep_ C.uncert m1	5	0.0241	0.0154
29	logreg_ sweep_ C.uncert bc	5	0.0241	0.0154
30	logreg_ no_ interc_ sweep_ C.uncert bc	5	0.0240	0.0154
31	logreg.uncert m2	5	0.0240	0.0153
32	average of TemperatureScaling	5	0.0234	0.0147
33	grad_ m2 m2	5	0.0232	0.0145

34	grad_m2 sbt	5	0.0231	0.0144
35	grad_m2 m1	5	0.0231	0.0144
36	grad_m1 m1	5	0.0230	0.0143
37	cal_prob_average.uncert bc	5	0.0229	0.0143
38	cal_prob_average.uncert m1	5	0.0229	0.0142
39	logreg_sweep_C m1	5	0.0229	0.0142
40	prob_average.uncert bc	5	0.0229	0.0142
41	grad_m1 m2	5	0.0229	0.0142
42	cal_prob_average.uncert m2	5	0.0228	0.0142
43	cal_prob_average.uncert sbt	5	0.0228	0.0142
44	grad_m1 sbt	5	0.0228	0.0141
45	prob_average.uncert m1	5	0.0227	0.0141
46	grad_m2.uncert m1	5	0.0227	0.0140
47	grad_m2.uncert bc	5	0.0227	0.0140
48	grad_bc.uncert bc	5	0.0226	0.0140
49	grad_bc m1	5	0.0226	0.0140
50	grad_m1.uncert bc	5	0.0226	0.0139
51	cal_prob_average m2	5	0.0226	0.0139
52	prob_average.uncert sbt	5	0.0225	0.0139
53	grad_m2.uncert sbt	5	0.0225	0.0139
54	average.uncert bc	5	0.0225	0.0138
55	grad_m1.uncert m1	5	0.0225	0.0138
56	grad_bc m2	5	0.0225	0.0138
57	grad_bc sbt	5	0.0225	0.0138
58	cal_prob_average sbt	5	0.0224	0.0138
59	grad_bc.uncert m1	5	0.0224	0.0137
60	grad_m1.uncert sbt	5	0.0224	0.0137
61	grad_m2.uncert m2	5	0.0223	0.0137
62	prob_average.uncert m2	5	0.0223	0.0137
63	average.uncert m1	5	0.0223	0.0137
64	grad_bc bc	5	0.0223	0.0137
65	grad_m2 bc	5	0.0223	0.0137
66	grad_bc.uncert sbt	5	0.0222	0.0135
67	grad_m1.uncert m2	5	0.0222	0.0135
68	grad_bc.uncert m2	5	0.0220	0.0134
69	average.uncert sbt	5	0.0220	0.0134
70	grad_m1 bc	5	0.0220	0.0133
71	cal_prob_average m1	5	0.0219	0.0133
72	average.uncert m2	5	0.0219	0.0133
73	logreg m2	5	0.0218	0.0131
74	prob_average sbt	5	0.0217	0.0130
75	prob_average m1	5	0.0217	0.0130
76	prob_average bc	5	0.0215	0.0128
77	prob_average m2	5	0.0214	0.0127
78	logreg_no_interc m1	5	0.0213	0.0127
79	average bc	5	0.0211	0.0125
80	average m1	5	0.0211	0.0125
81	average m2	5	0.0211	0.0125
82	average sbt	5	0.0211	0.0125
83	logreg_no_interc sbt	5	0.0210	0.0123
84	logreg_no_interc m2	5	0.0208	0.0121
85	logreg_sweep_C bc	5	0.0205	0.0118
86	logreg sbt	5	0.0205	0.0118
87	logreg m1	5	0.0203	0.0116

88	logreg bc	5	0.0183	0.0096
89	cal_prob_average bc	5	0.0181	0.0094

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Thu Feb 10 10:49:55 2022

	method	combination_size	imp_o_avg	imp_o_max
1	cal_average bc	6	0.0268	0.0169
2	cal_average m1	6	0.0268	0.0169
3	cal_average m2	6	0.0268	0.0169
4	logreg_sweep_C.uncert m2	6	0.0264	0.0165
5	cal_average.uncert m2	6	0.0263	0.0164
6	cal_average.uncert sbt	6	0.0263	0.0164
7	logreg_no_interc.uncert m1	6	0.0263	0.0164
8	logreg_no_interc_sweep_C m1	6	0.0263	0.0164
9	cal_average.uncert m1	6	0.0261	0.0162
10	logreg_no_interc_sweep_C sbt	6	0.0261	0.0162
11	logreg_no_interc_sweep_C.uncert m2	6	0.0260	0.0161
12	cal_average.uncert bc	6	0.0259	0.0160
13	logreg_no_interc_sweep_C.uncert sbt	6	0.0259	0.0160
14	logreg_no_interc.uncert bc	6	0.0258	0.0159
15	logreg_no_interc_sweep_C.uncert m1	6	0.0258	0.0159
16	cal_average sbt	6	0.0257	0.0158
17	logreg_no_interc_sweep_C m2	6	0.0257	0.0158
18	logreg_sweep_C m2	6	0.0257	0.0158
19	logreg_sweep_C sbt	6	0.0256	0.0157
20	logreg.uncert bc	6	0.0255	0.0156
21	logreg_no_interc bc	6	0.0254	0.0155
22	logreg_no_interc_sweep_C.uncert bc	6	0.0254	0.0155
23	logreg_sweep_C.uncert sbt	6	0.0254	0.0155
24	logreg.uncert sbt	6	0.0253	0.0154
25	logreg.uncert m1	6	0.0252	0.0153
26	logreg_no_interc.uncert sbt	6	0.0251	0.0152
27	logreg_no_interc.uncert m2	6	0.0250	0.0151
28	logreg_sweep_C.uncert m1	6	0.0249	0.0150
29	logreg.uncert m2	6	0.0248	0.0149
30	average of TemperatureScaling	6	0.0248	0.0149
31	logreg_sweep_C.uncert bc	6	0.0245	0.0146
32	logreg_no_interc_sweep_C bc	6	0.0244	0.0145
33	grad_m1 sbt	6	0.0243	0.0144
34	grad_m1.uncert bc	6	0.0243	0.0144
35	grad_m2 m1	6	0.0243	0.0144
36	grad_m2.uncert bc	6	0.0243	0.0144
37	cal_prob_average.uncert bc	6	0.0242	0.0143
38	cal_prob_average.uncert m1	6	0.0242	0.0143
39	cal_prob_average.uncert m2	6	0.0242	0.0143
40	cal_prob_average.uncert sbt	6	0.0242	0.0143
41	grad_bc.uncert bc	6	0.0242	0.0143
42	grad_m1 m1	6	0.0242	0.0143
43	grad_m2 m2	6	0.0242	0.0143
44	grad_m2 sbt	6	0.0242	0.0143
45	grad_m2.uncert m1	6	0.0242	0.0143
46	grad_bc.uncert m1	6	0.0241	0.0142
47	grad_m1 m2	6	0.0241	0.0142

48	prob_ average.uncert sbt	6	0.0241	0.0142
49	average.uncert bc	6	0.0240	0.0141
50	grad_m1.uncert m1	6	0.0240	0.0141
51	grad_bc m1	6	0.0239	0.0140
52	prob_ average.uncert bc	6	0.0239	0.0140
53	prob_ average.uncert m1	6	0.0239	0.0140
54	prob_ average.uncert m2	6	0.0239	0.0140
55	grad_bc m2	6	0.0238	0.0139
56	grad_m2 bc	6	0.0238	0.0139
57	grad_m2.uncert sbt	6	0.0238	0.0139
58	logreg_sweep_C m1	6	0.0238	0.0139
59	average.uncert m1	6	0.0237	0.0138
60	grad_bc bc	6	0.0237	0.0138
61	grad_m1.uncert sbt	6	0.0237	0.0138
62	grad_bc sbt	6	0.0236	0.0137
63	cal_prob_ average m2	6	0.0235	0.0136
64	grad_bc.uncert sbt	6	0.0235	0.0136
65	grad_m1 bc	6	0.0235	0.0136
66	grad_m1.uncert m2	6	0.0235	0.0136
67	grad_m2.uncert m2	6	0.0235	0.0136
68	average.uncert sbt	6	0.0234	0.0135
69	grad_bc.uncert m2	6	0.0234	0.0135
70	average.uncert m2	6	0.0232	0.0133
71	cal_prob_ average sbt	6	0.0226	0.0127
72	prob_ average sbt	6	0.0226	0.0127
73	cal_prob_ average m1	6	0.0224	0.0125
74	prob_ average m1	6	0.0224	0.0125
75	prob_ average m2	6	0.0222	0.0123
76	average bc	6	0.0220	0.0121
77	average m1	6	0.0220	0.0121
78	average m2	6	0.0220	0.0121
79	average sbt	6	0.0220	0.0121
80	prob_ average bc	6	0.0219	0.0120
81	logreg m2	6	0.0218	0.0119
82	logreg_sweep_C bc	6	0.0214	0.0115
83	logreg sbt	6	0.0211	0.0112
84	logreg_no_interc m1	6	0.0208	0.0109
85	logreg_no_interc sbt	6	0.0208	0.0109
86	logreg_no_interc m2	6	0.0204	0.0105
87	logreg m1	6	0.0201	0.0102
88	logreg bc	6	0.0196	0.0097
89	cal_prob_ average bc	6	0.0193	0.0094