

WeightedLinearEnsemble evaluation new

```
library(ggplot2)

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

library(dplyr)

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

## 
## Attaching package: 'dplyr'

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

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

library(tidyr)

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

library(comprehendr)

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

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

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

library(patchwork)

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

```

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

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

networks <- net_df$network

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

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

```

```

    }
}

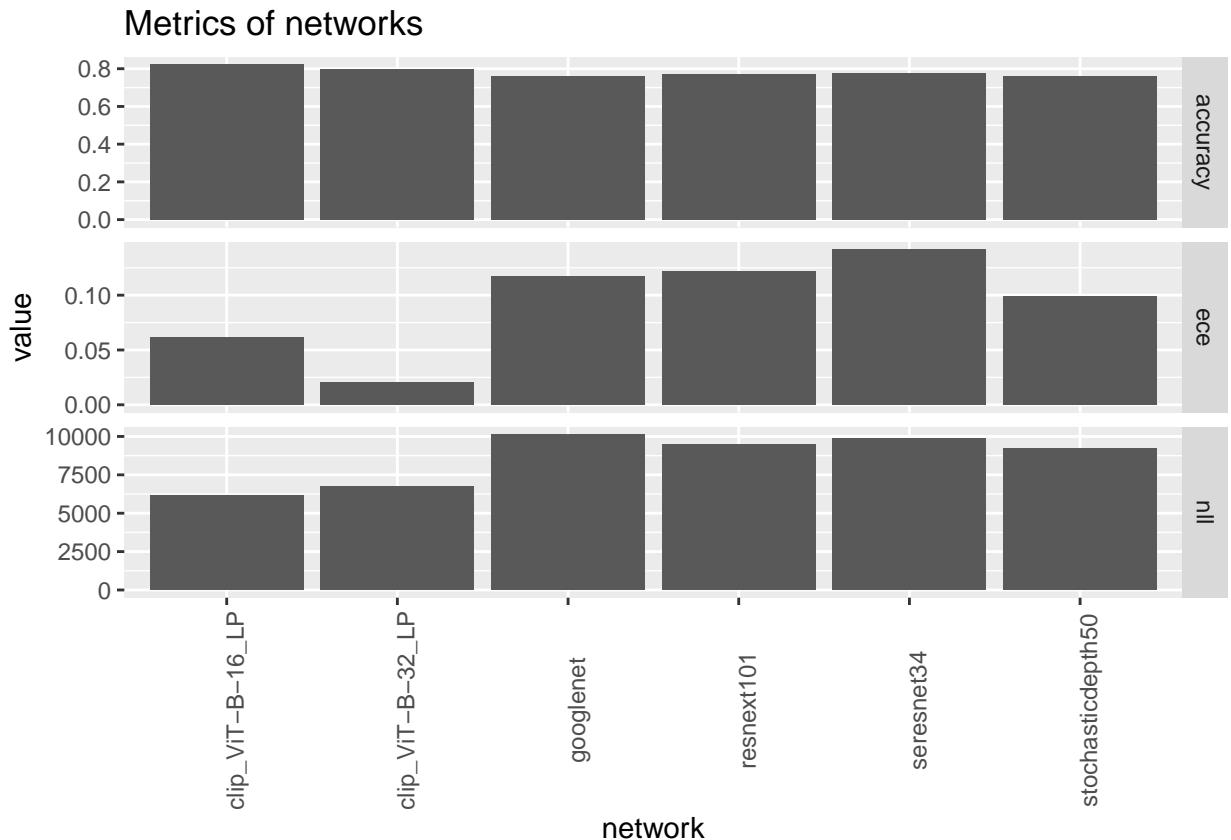
ens_df_cal <- merge(ens_df_cal, comb_stats_df)
ens_df_cal$acc_imp_avg <- ens_df_cal$accuracy - ens_df_cal$acc_avg
ens_df_cal$acc_imp_max <- ens_df_cal$accuracy - ens_df_cal$acc_max

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

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

nets_plot

```



```

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

comb_methods <- c(
  "average", "prob_average",
  "cal_average", "cal_prob_average",

```

```

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

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

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

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

```

```

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

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

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

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

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

```

```

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

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

```

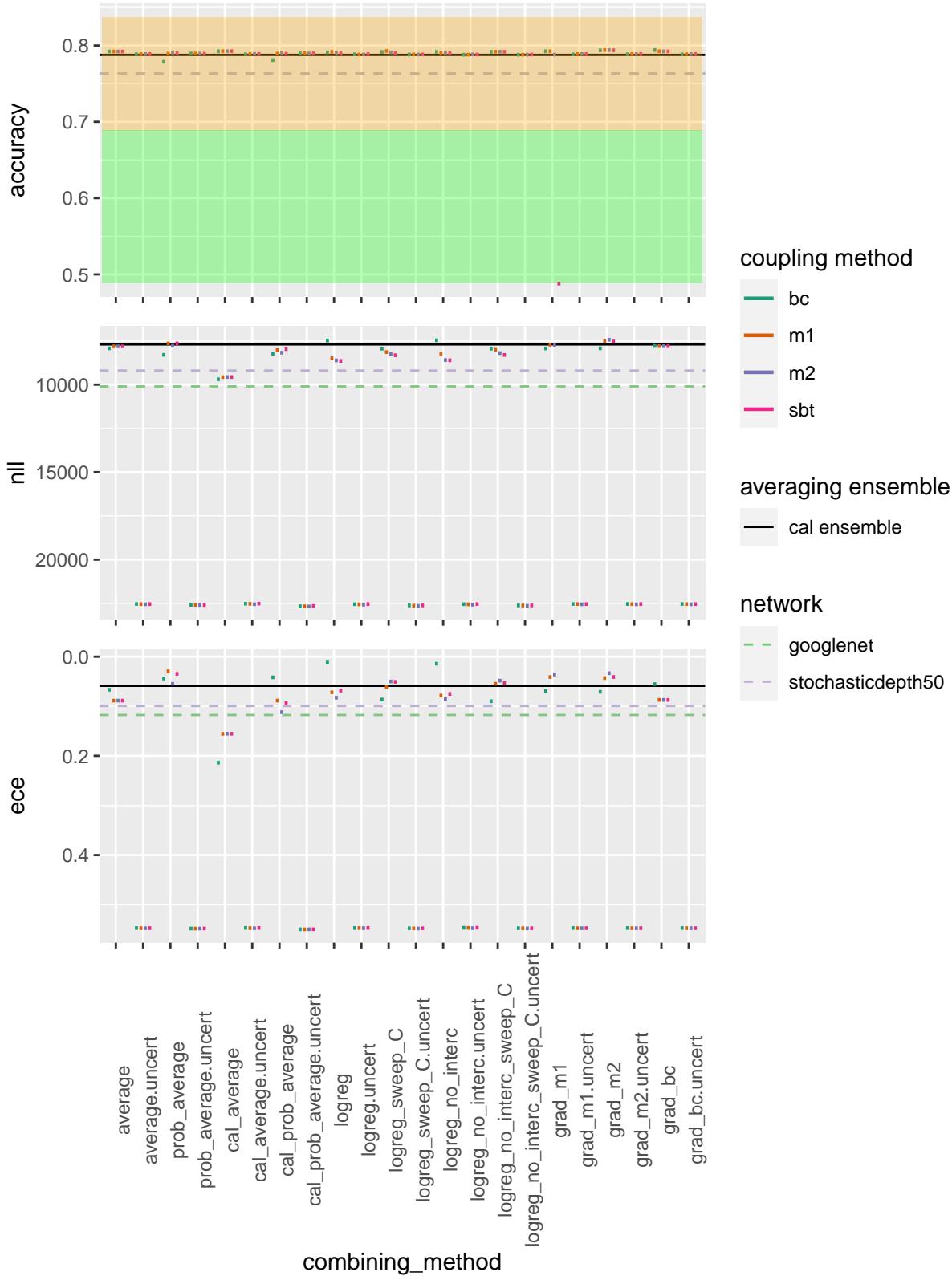
```

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

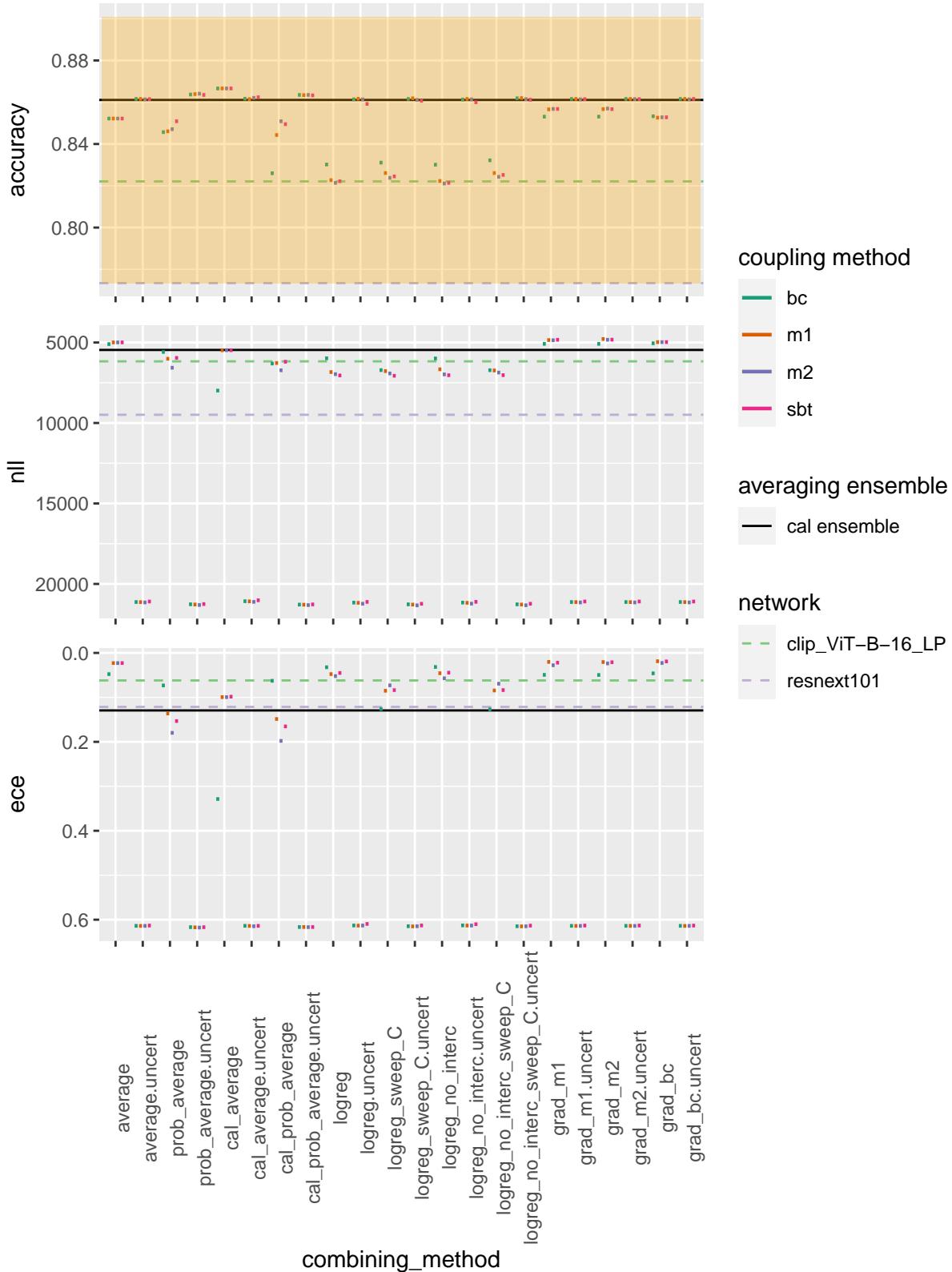
print(acc_plot / nll_plot / ece_plot + plot_layout(guides = "collect") +
  plot_annotation(title = paste(
    "Ensemble metrics",
    paste(
      c("Error inconsistency", cur_ens_cal$err_incons[[1]]),
      collapse = " "
    ),
    sep = "\n"
  )))
}
}
}

```

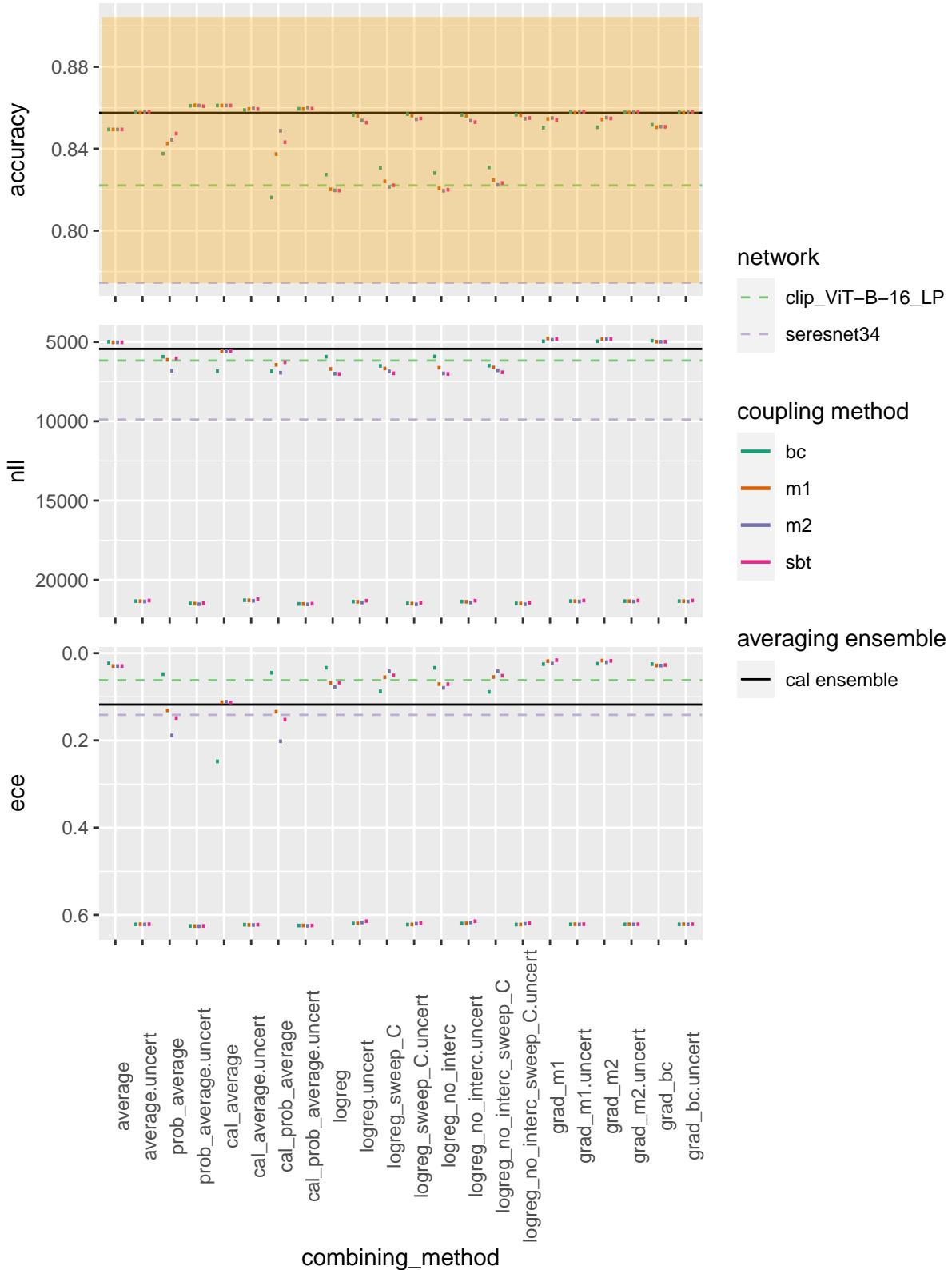
Ensemble metrics
Error inconsistency 0.148800000548363



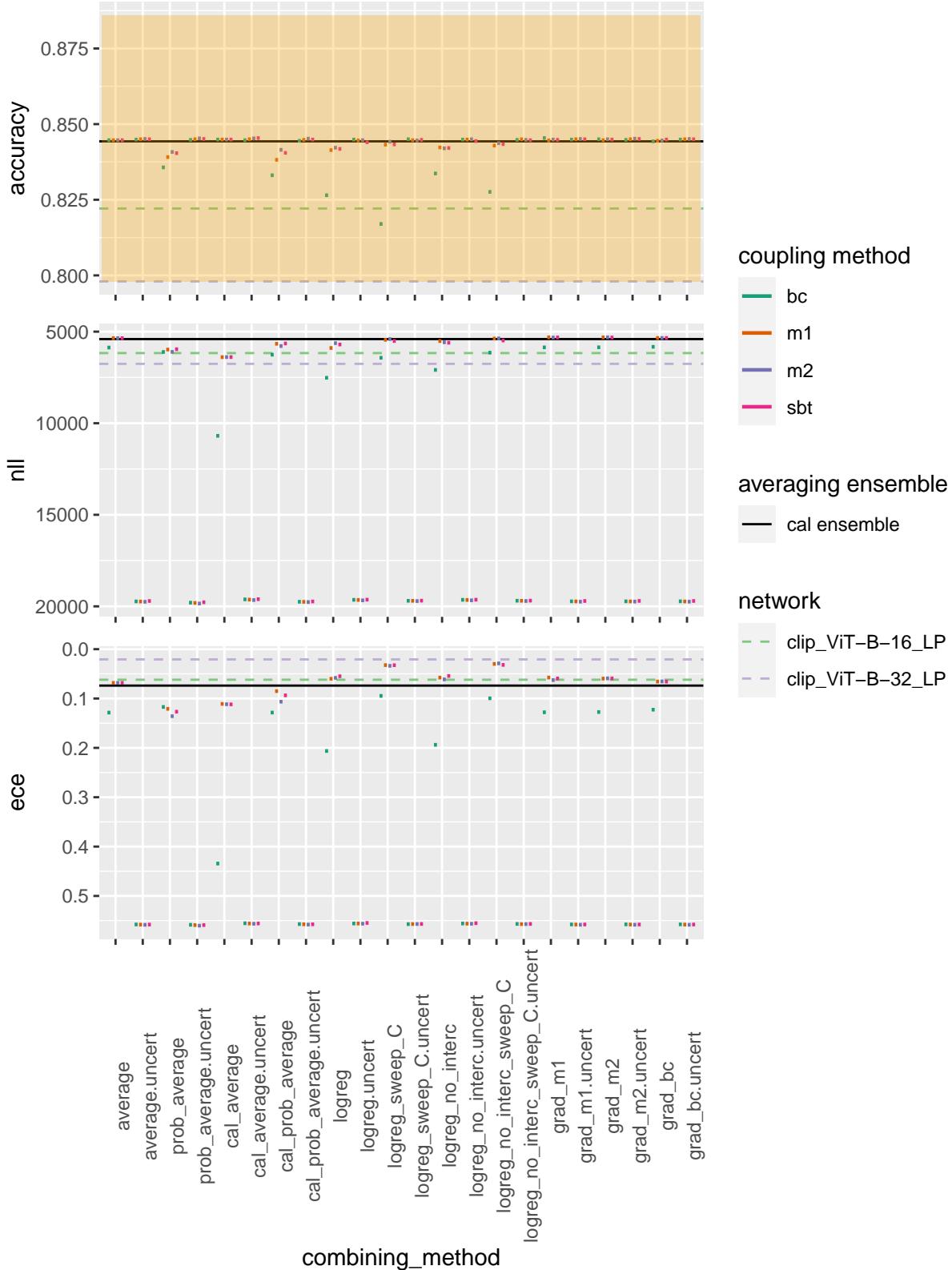
Ensemble metrics
Error inconsistency 0.206299990415573



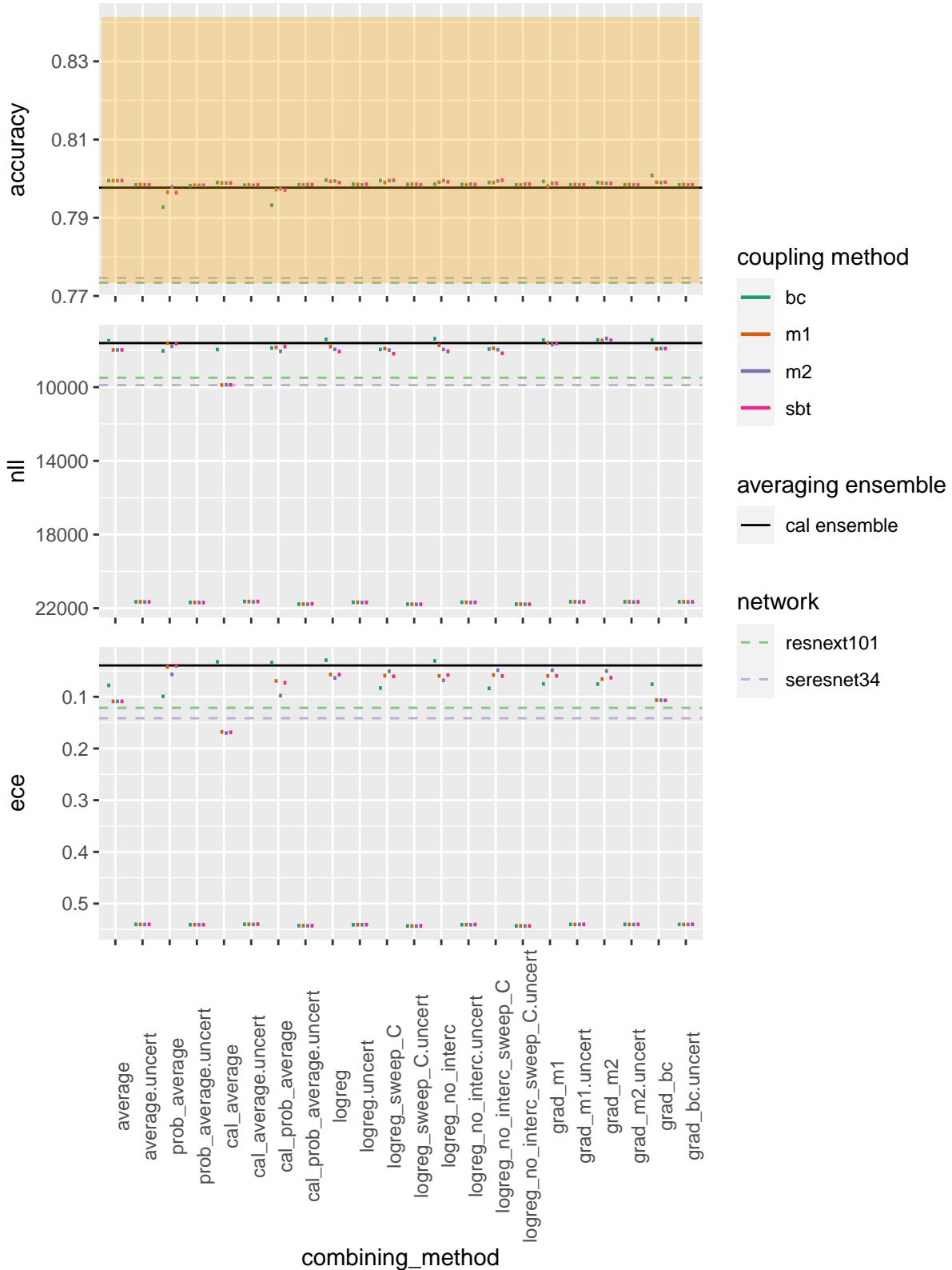
Ensemble metrics
Error inconsistency 0.212099999189377



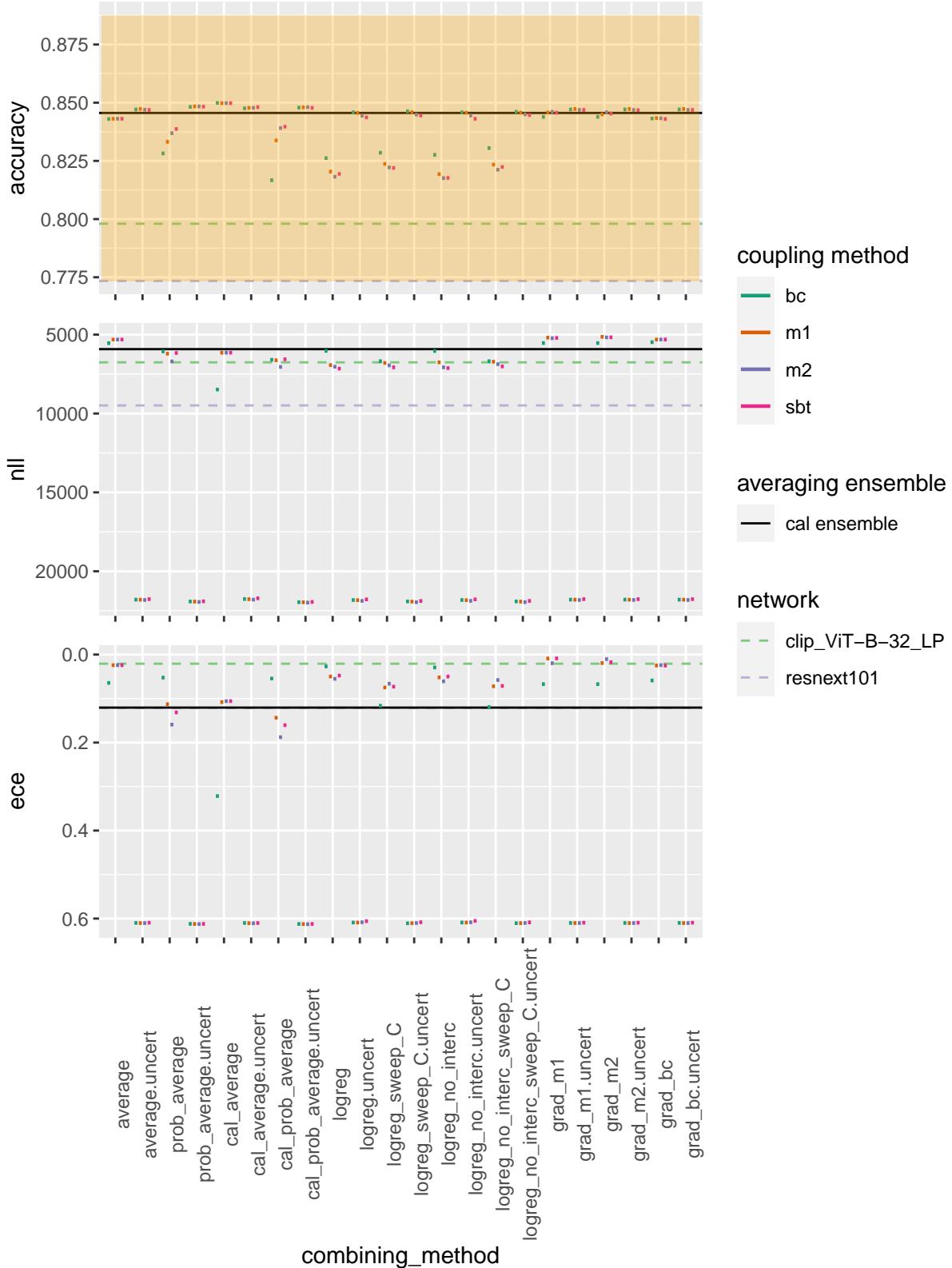
Ensemble metrics
Error inconsistency 0.151899993419647



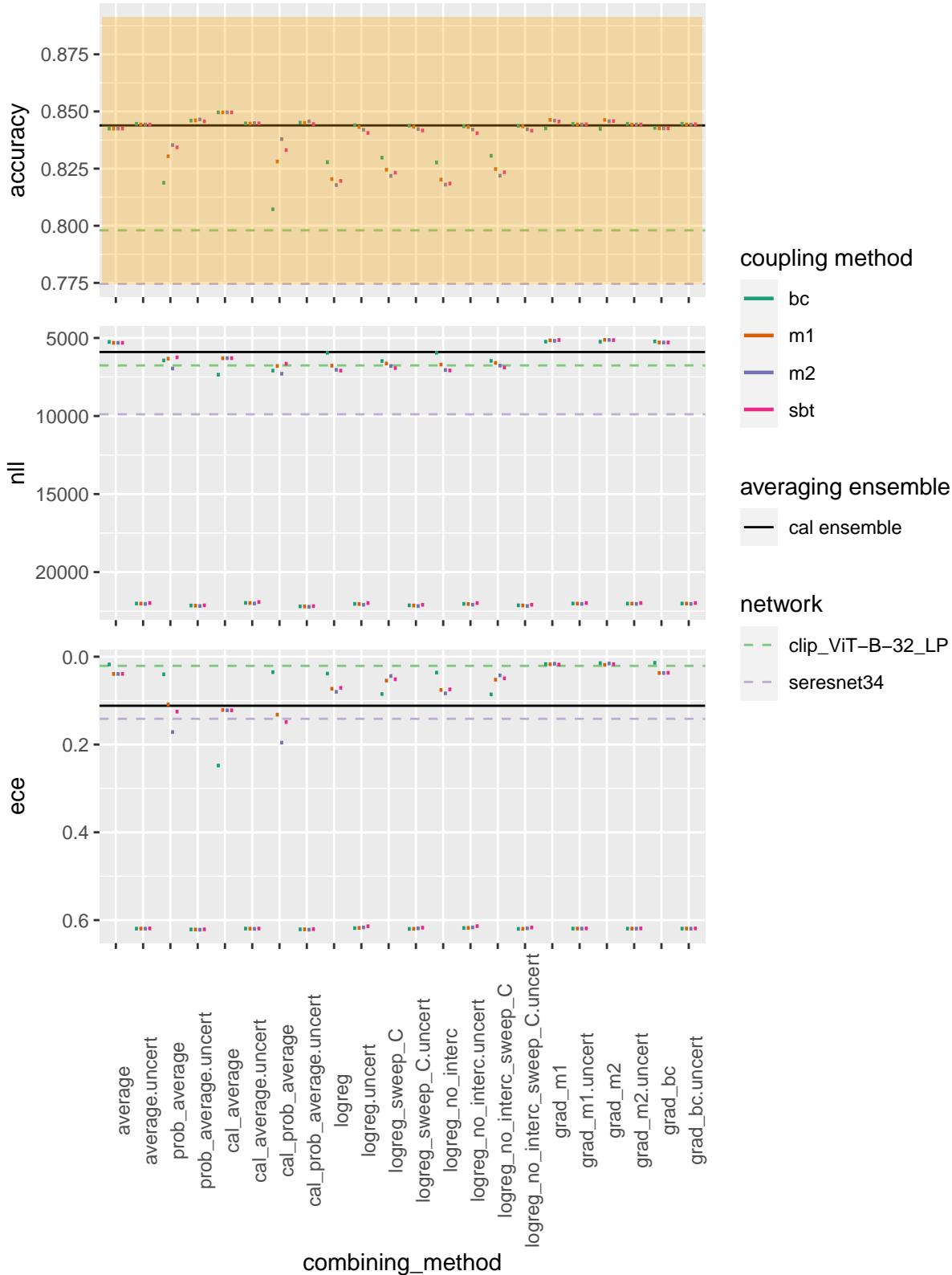
Ensemble metrics
Error inconsistency 0.134800001978874



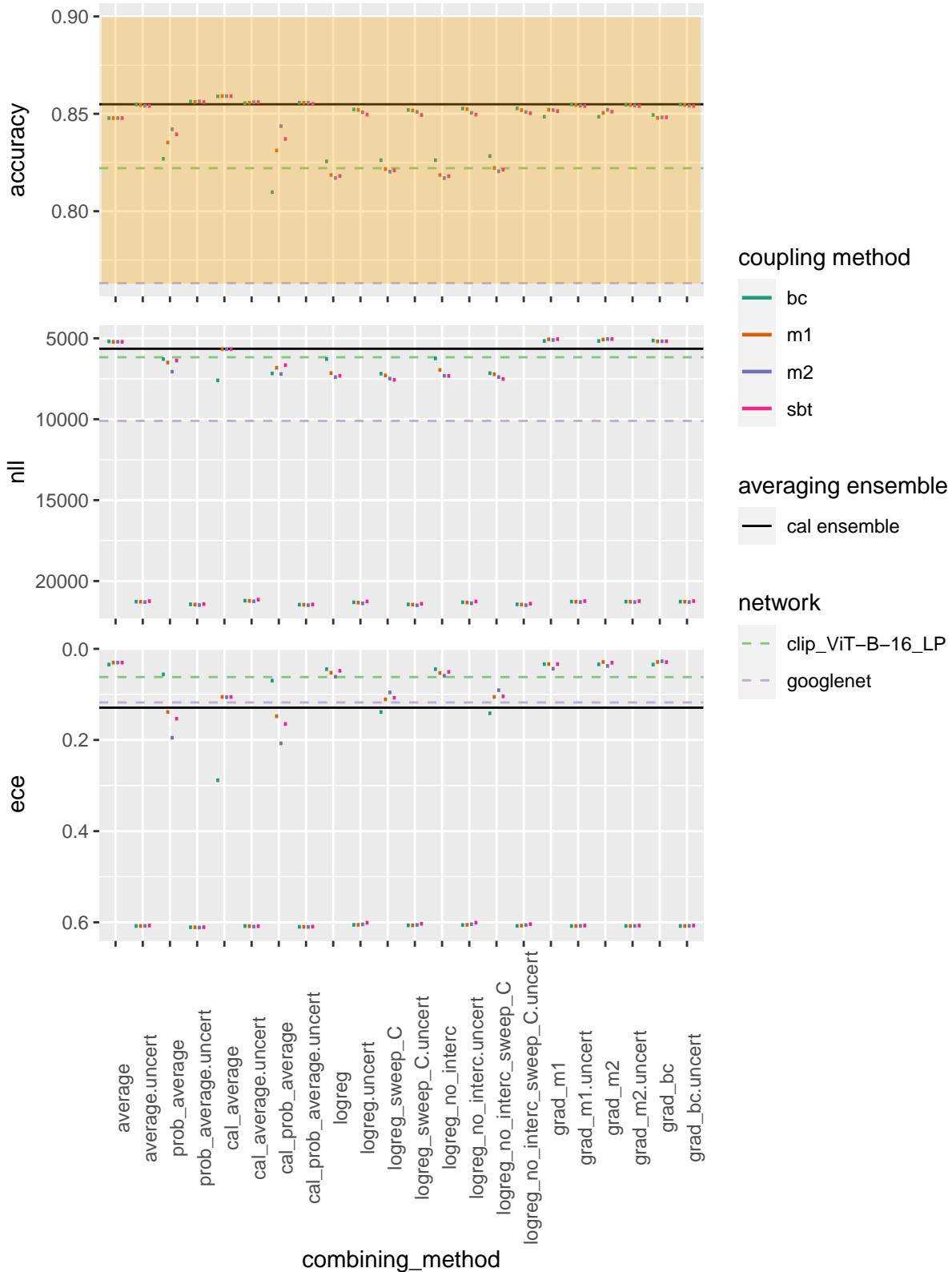
Ensemble metrics
Error inconsistency 0.203799992799759



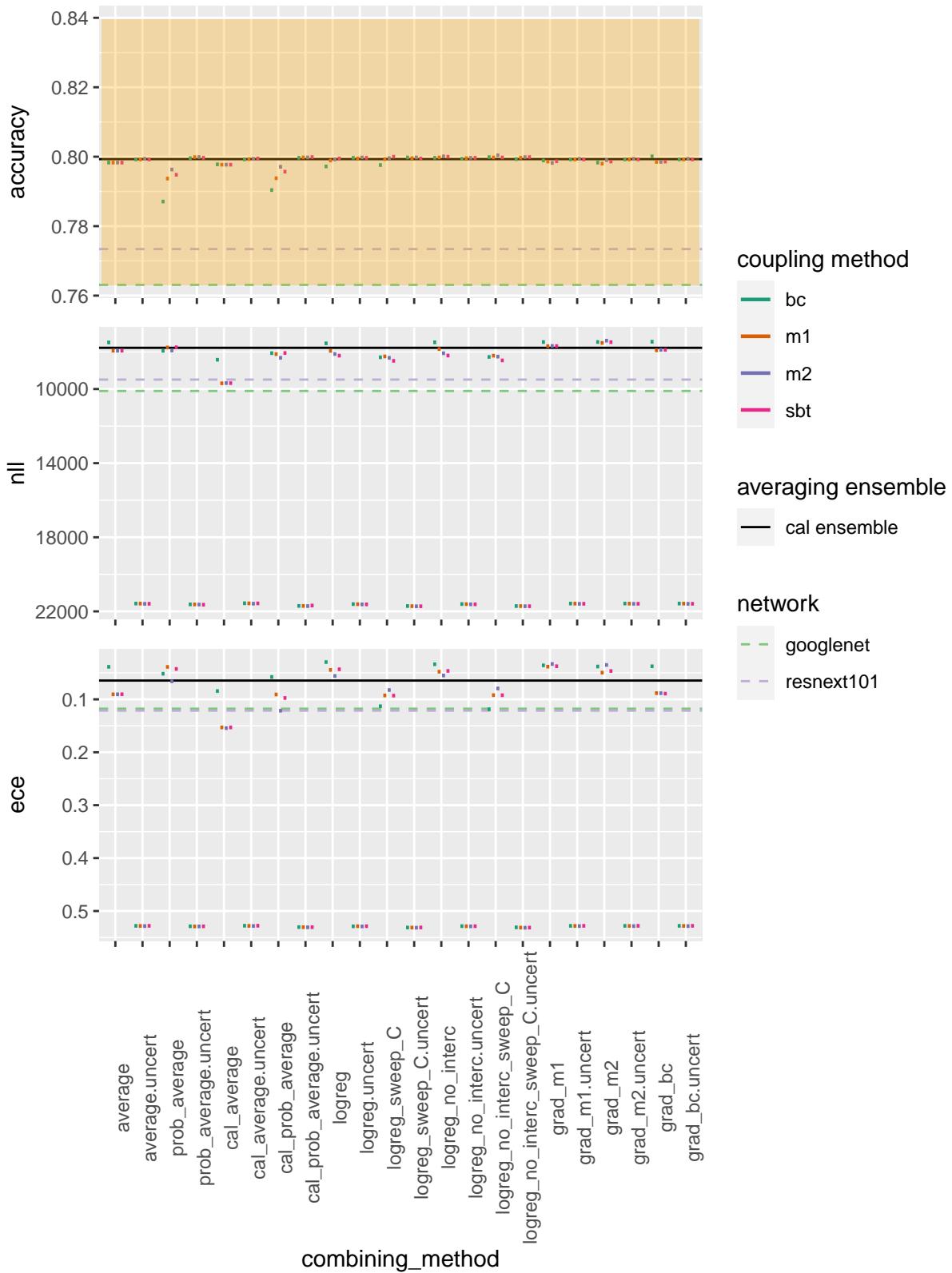
Ensemble metrics
Error inconsistency 0.210199996829033



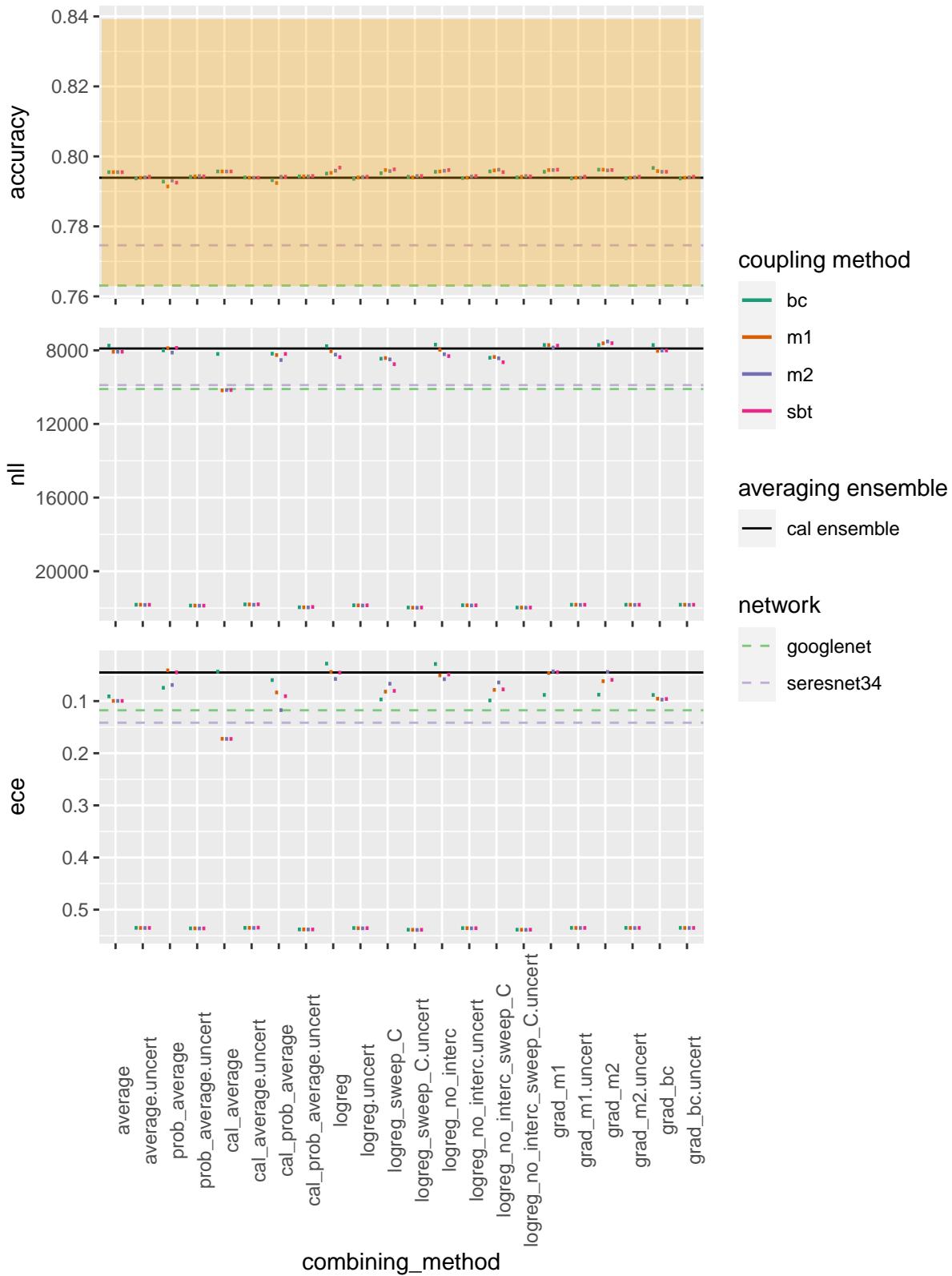
Ensemble metrics
Error inconsistency 0.214399993419647



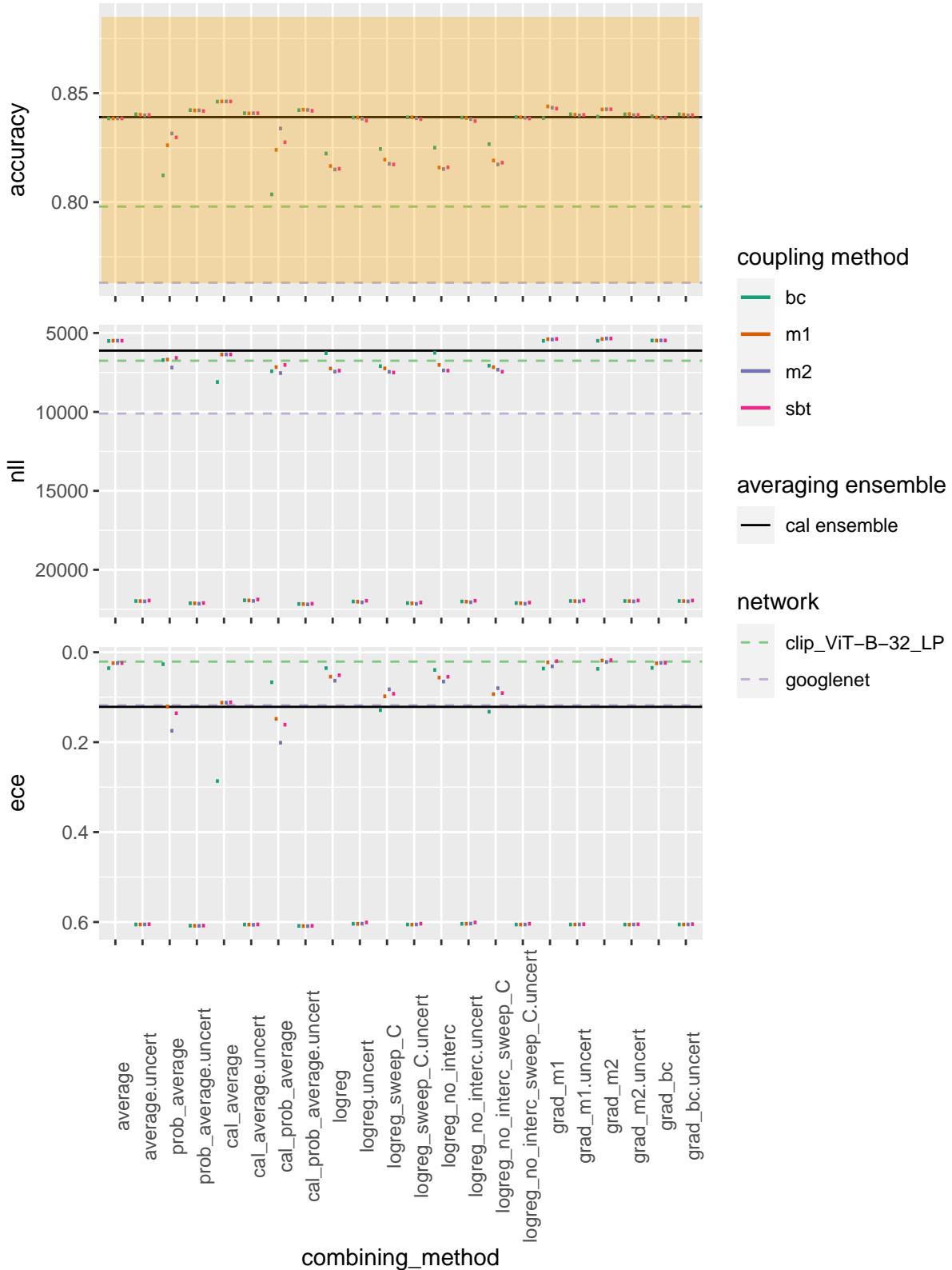
Ensemble metrics
Error inconsistency 0.142700001597404



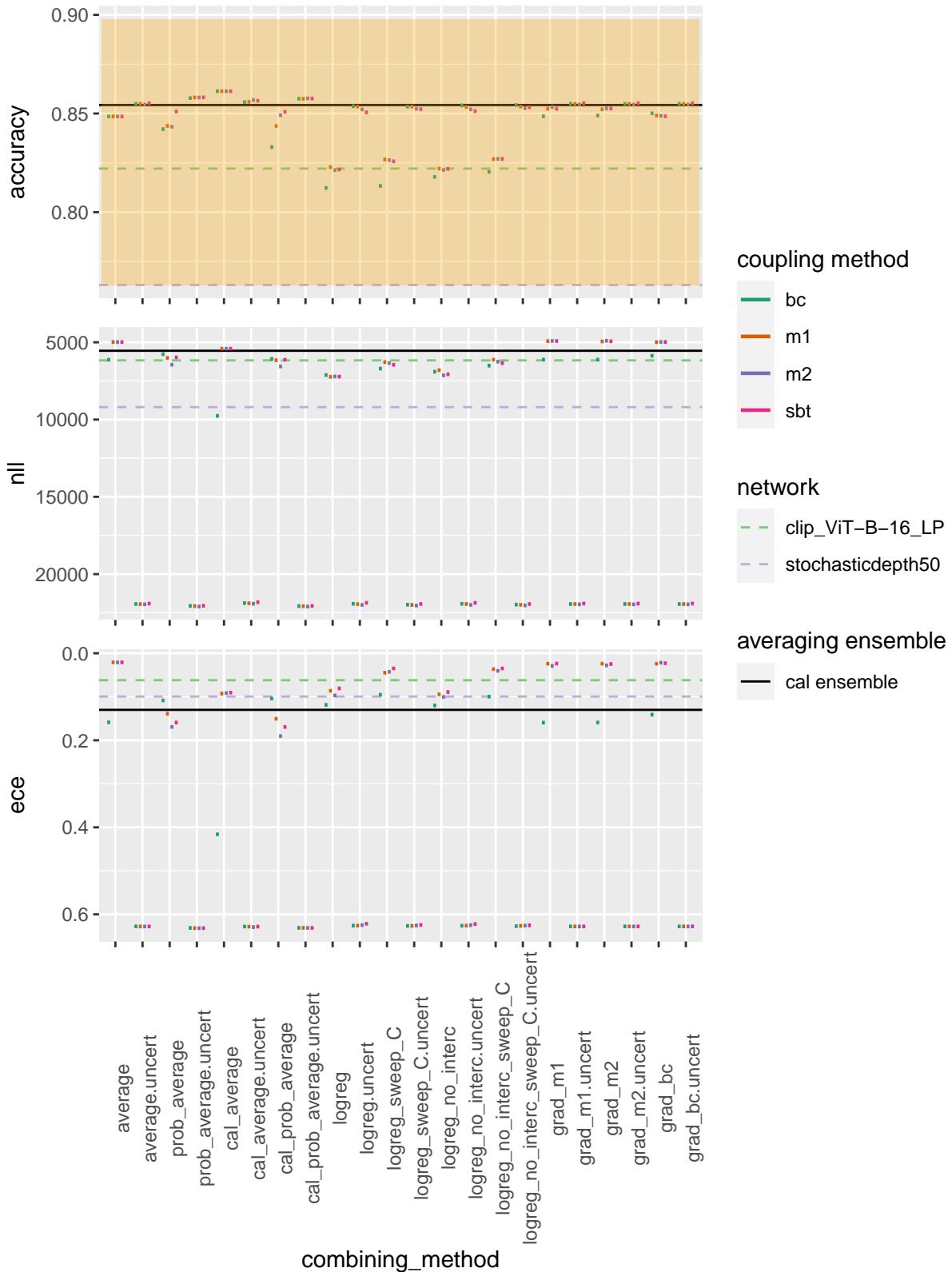
Ensemble metrics
Error inconsistency 0.140699997544289



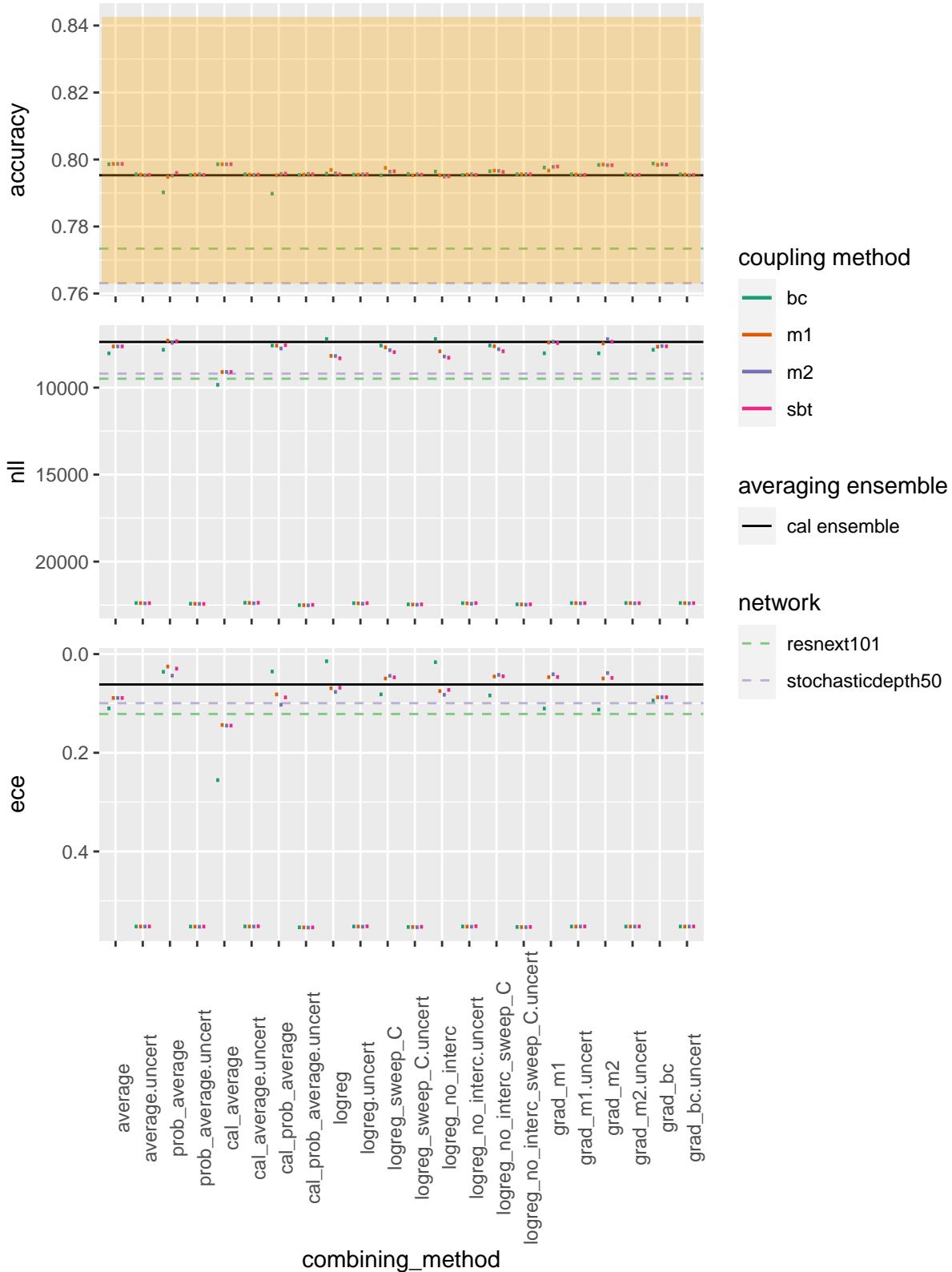
Ensemble metrics
Error inconsistency 0.208899989724159



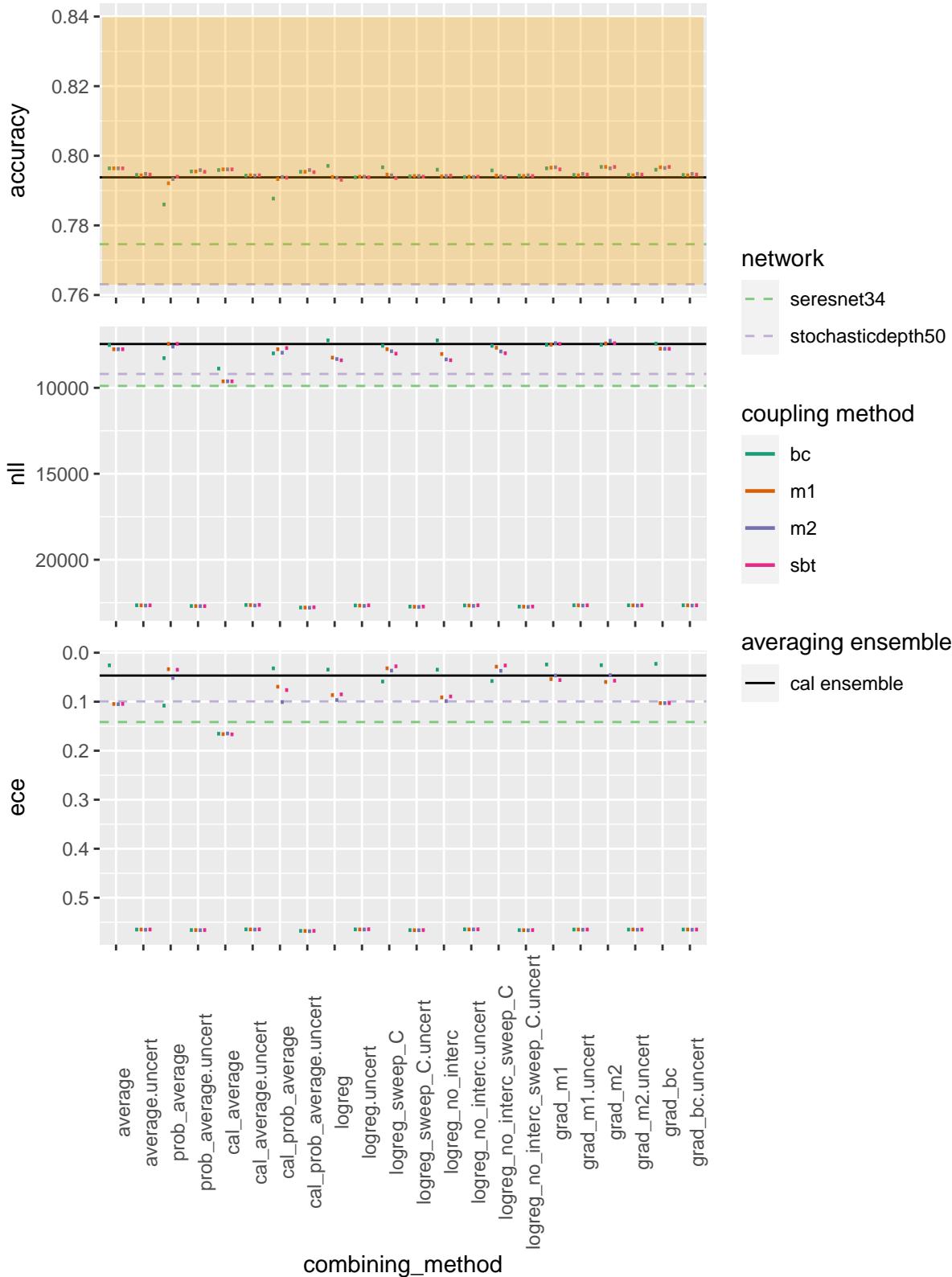
Ensemble metrics
Error inconsistency 0.210400000214577



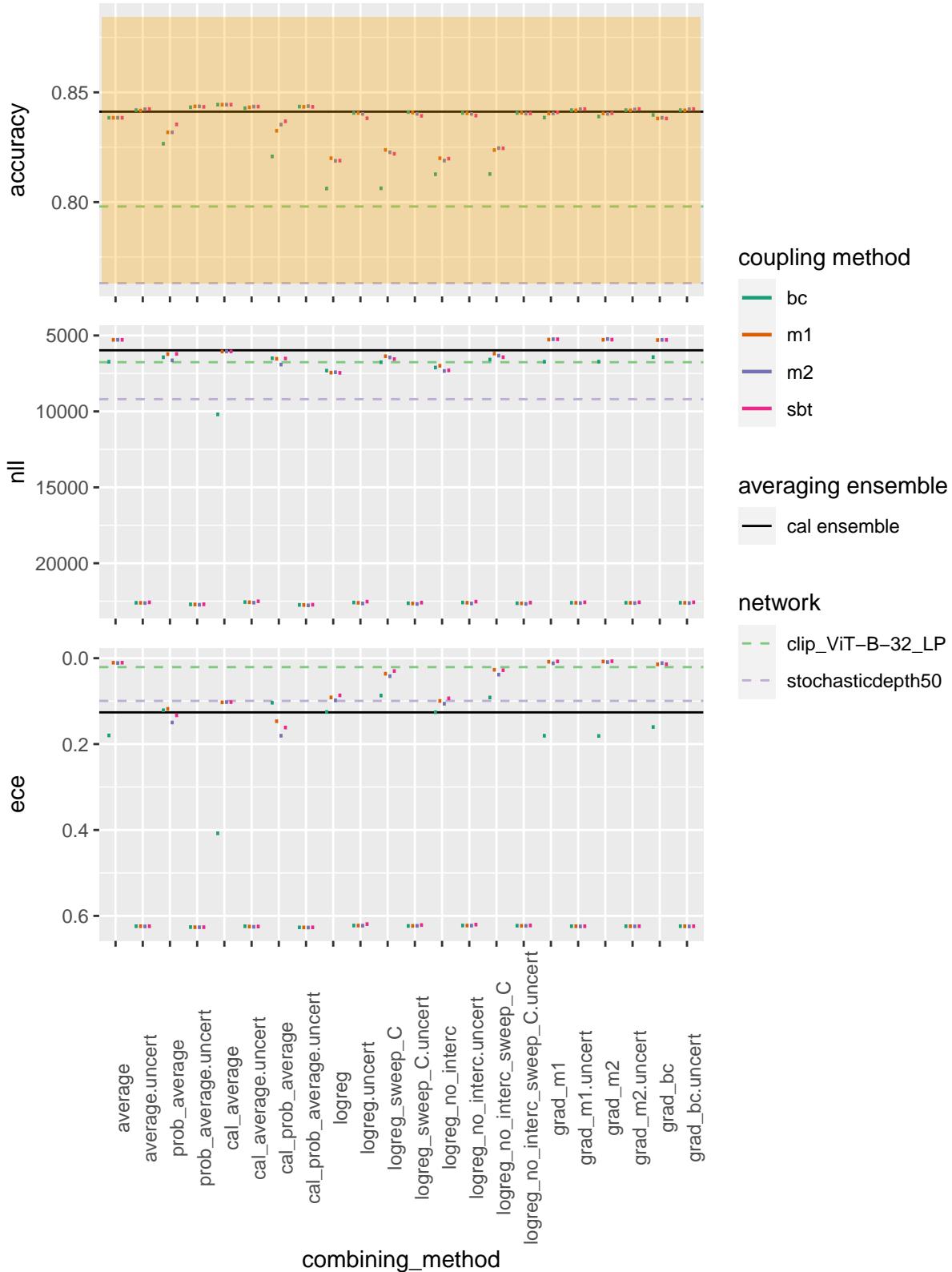
Ensemble metrics
Error inconsistency 0.148699998855591



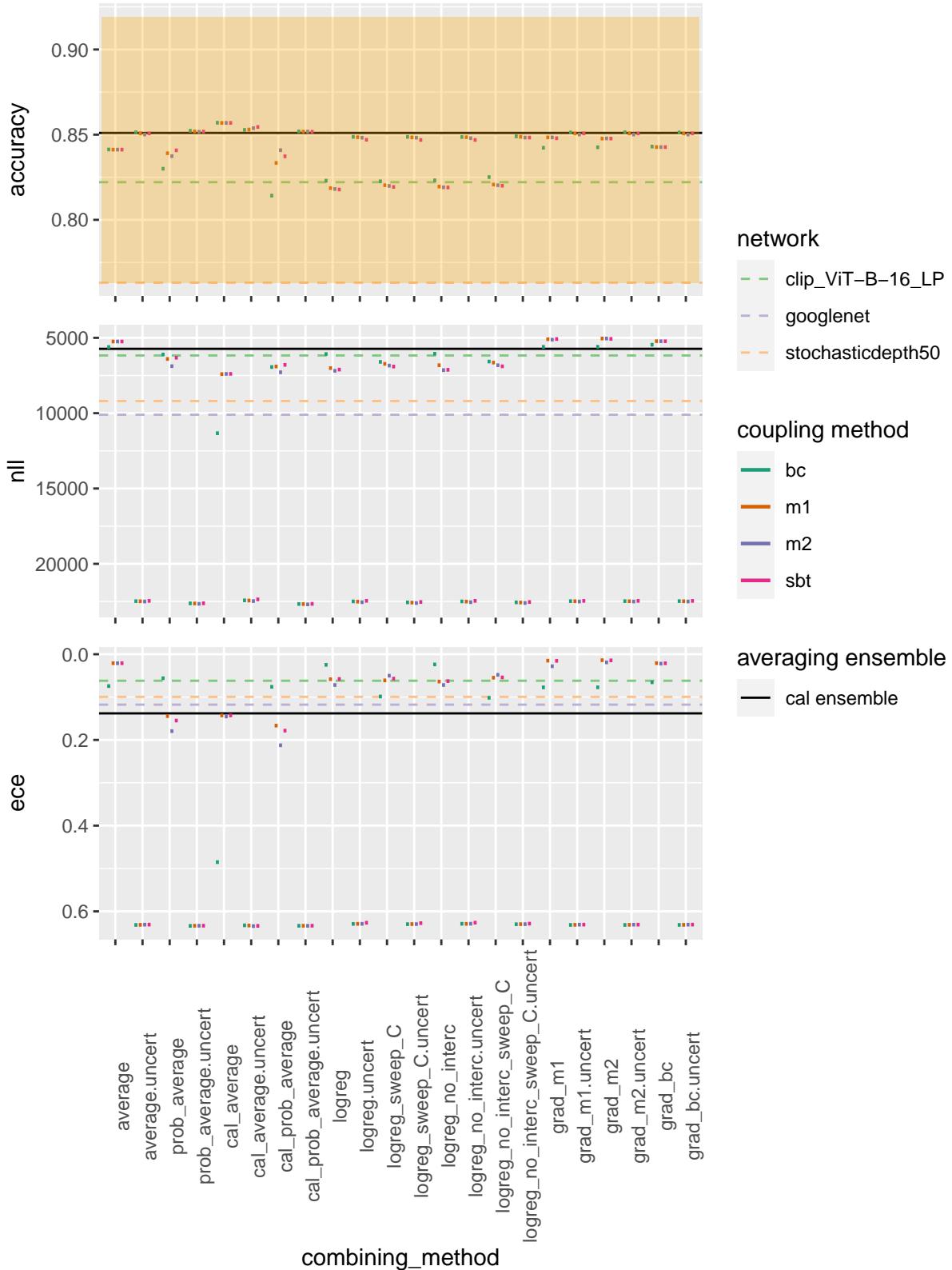
Ensemble metrics
Error inconsistency 0.142099991440773



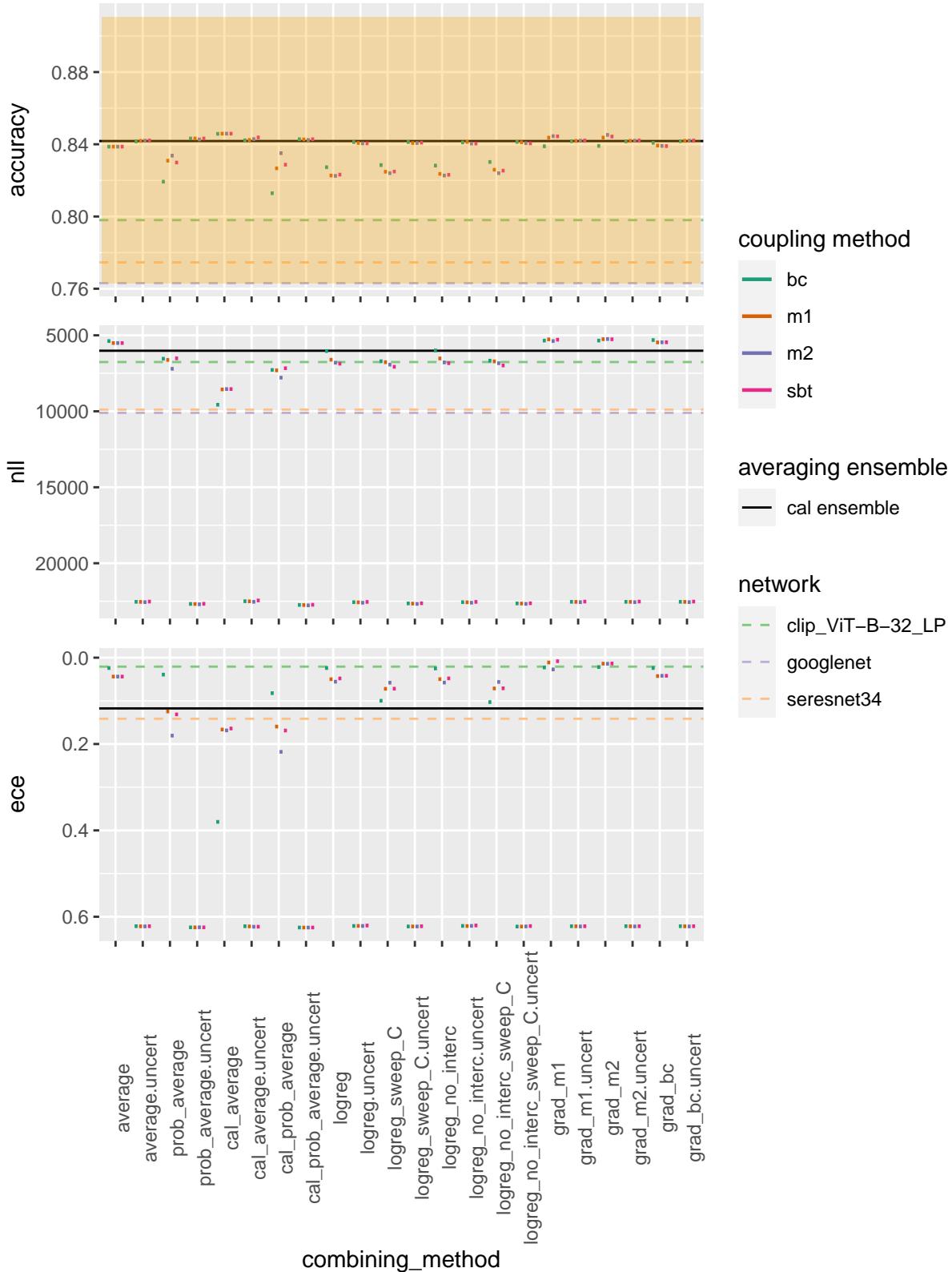
Ensemble metrics
Error inconsistency 0.207699999213219



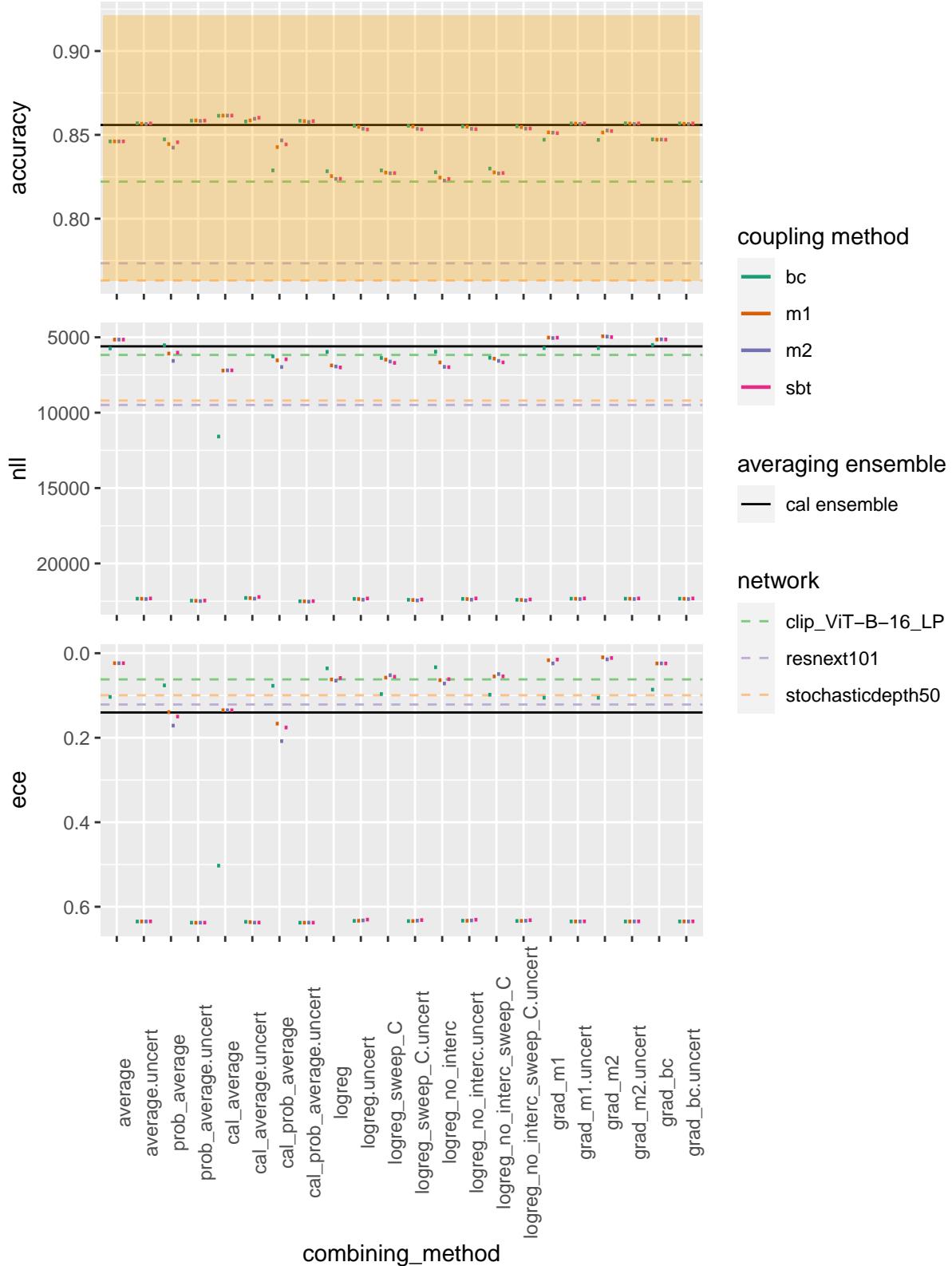
Ensemble metrics
Error inconsistency 0.286799997091293



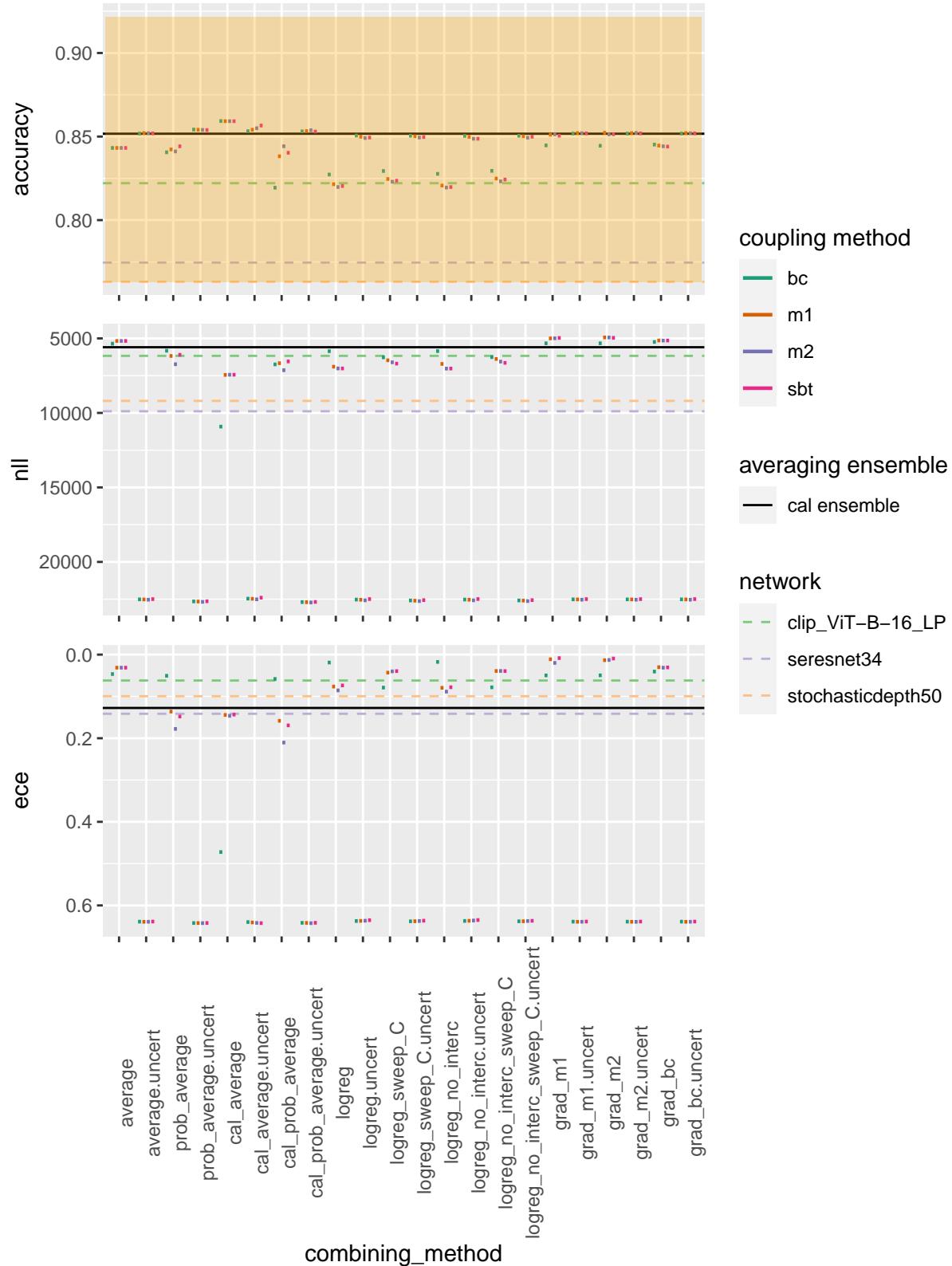
Ensemble metrics
Error inconsistency 0.27989998459816



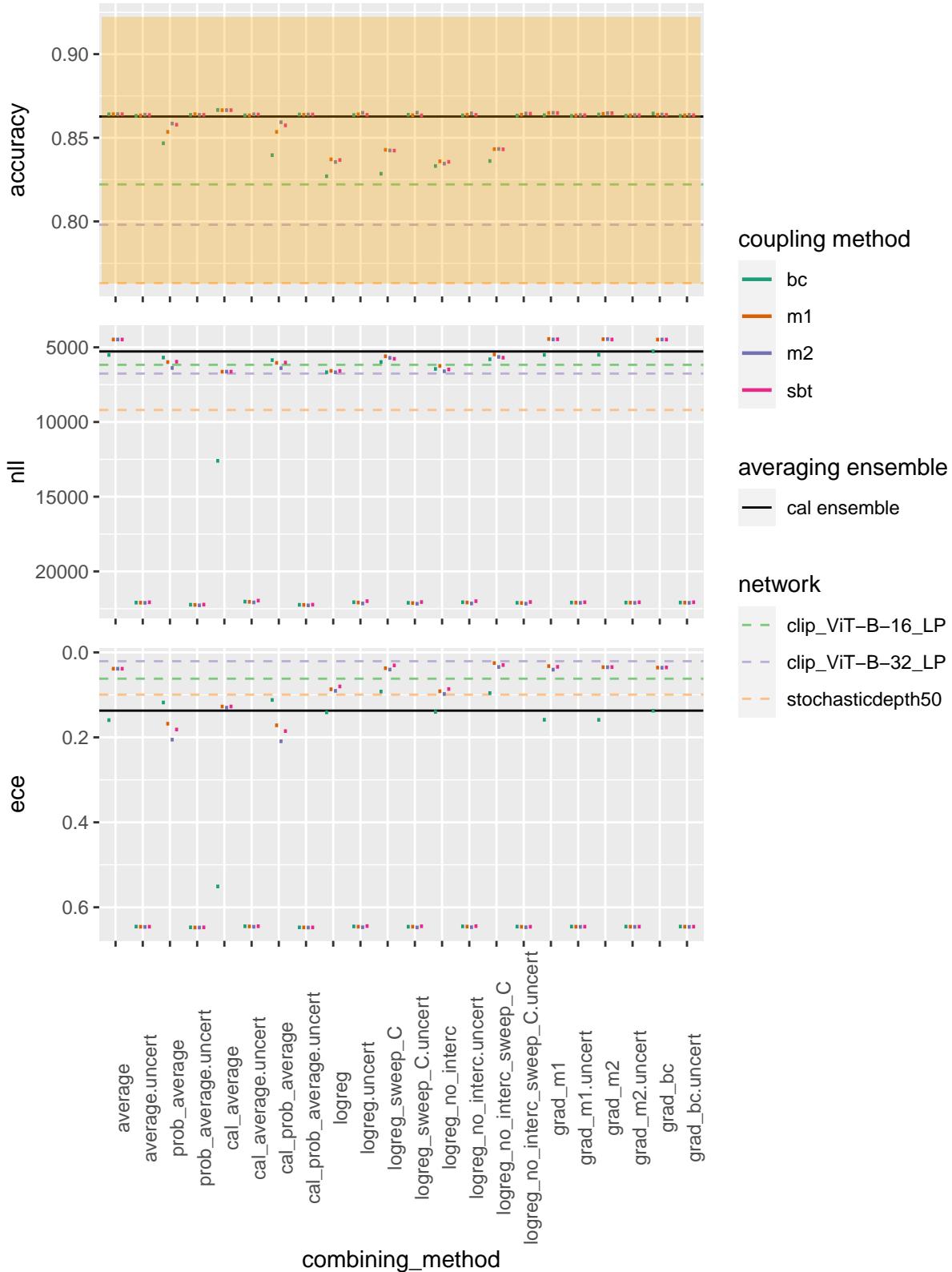
Ensemble metrics
Error inconsistency 0.282700002193451



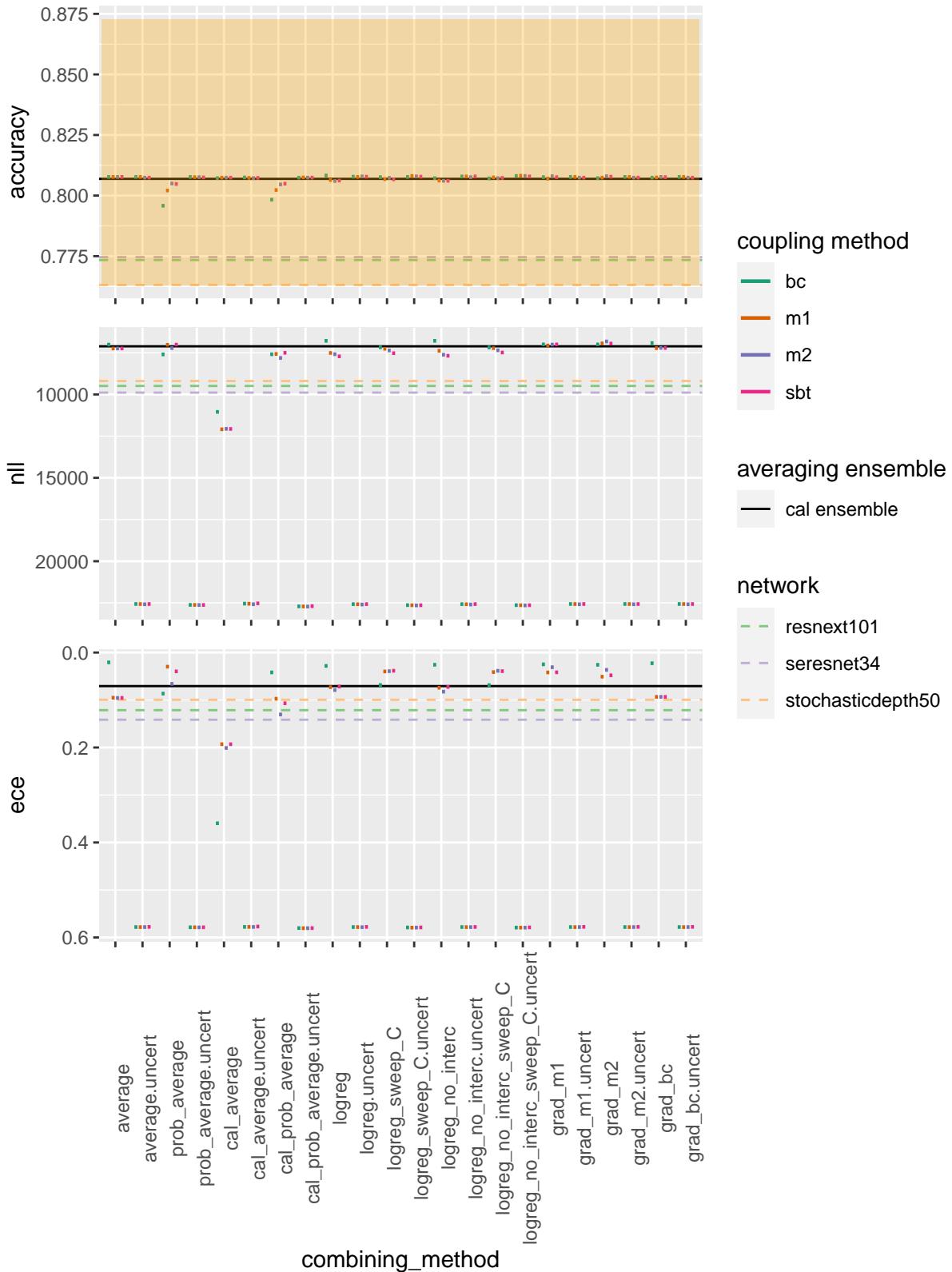
Ensemble metrics
Error inconsistency 0.282299995422363



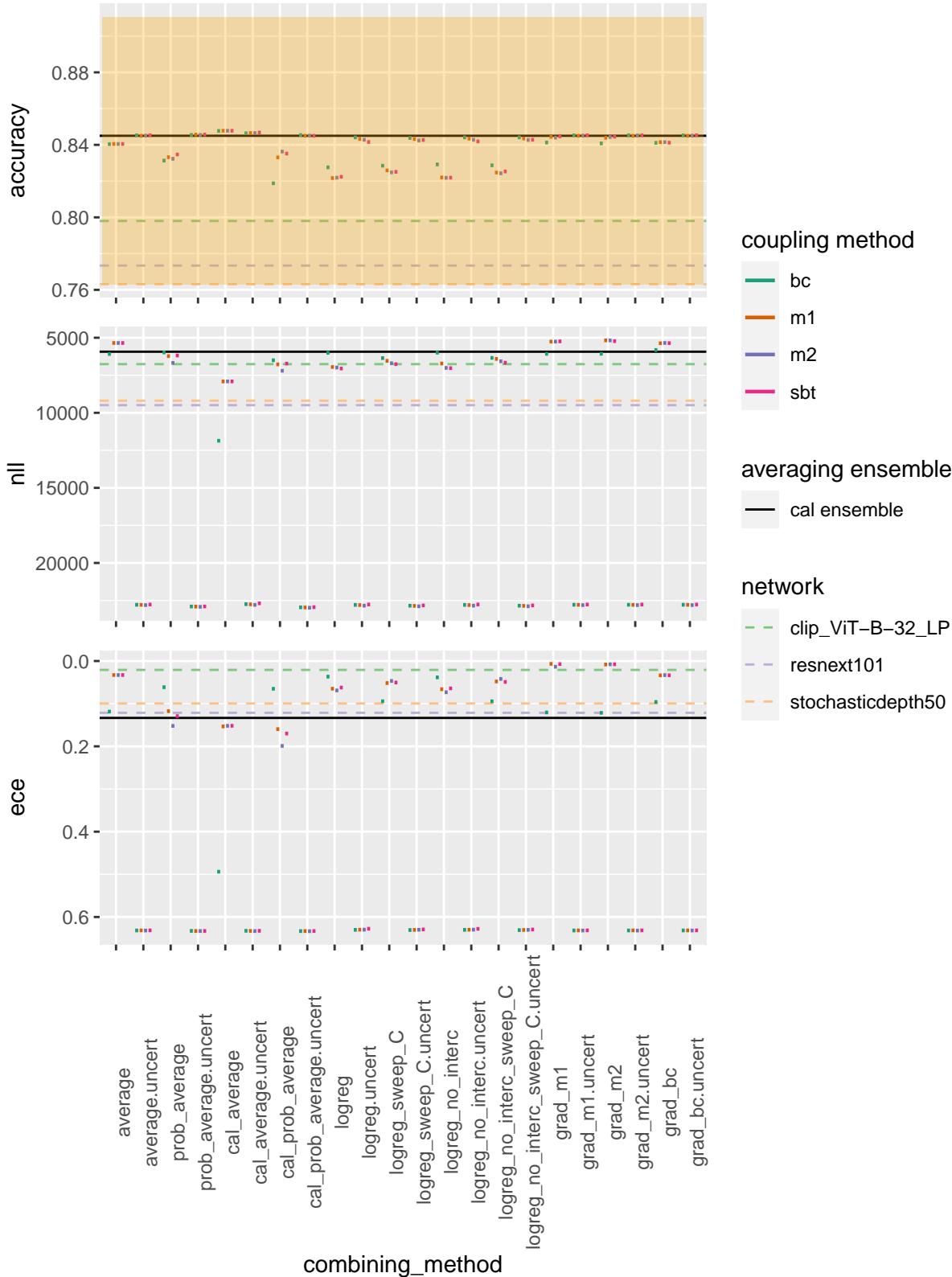
Ensemble metrics
Error inconsistency 0.284999996423721



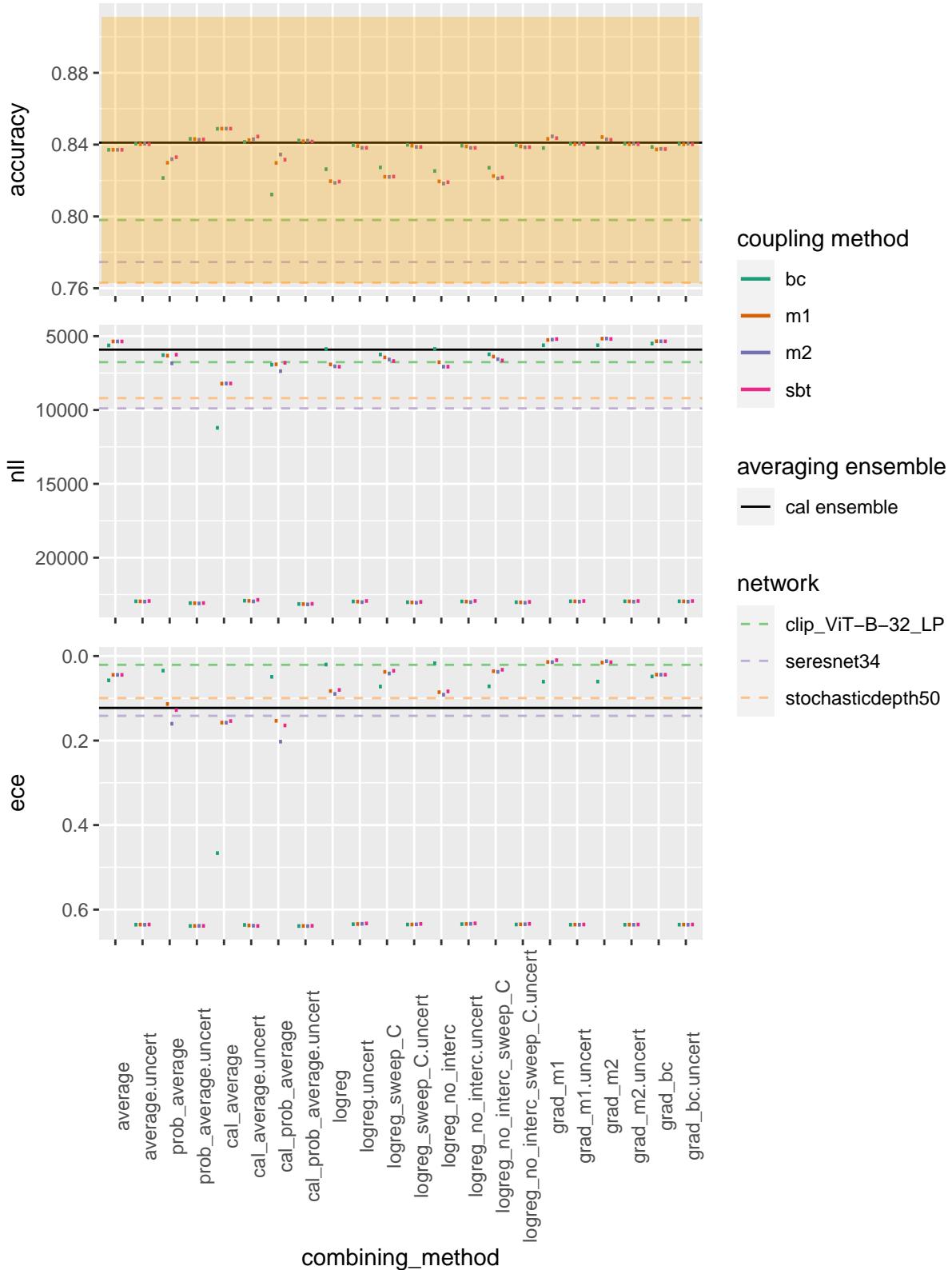
Ensemble metrics
Error inconsistency 0.212799996137619



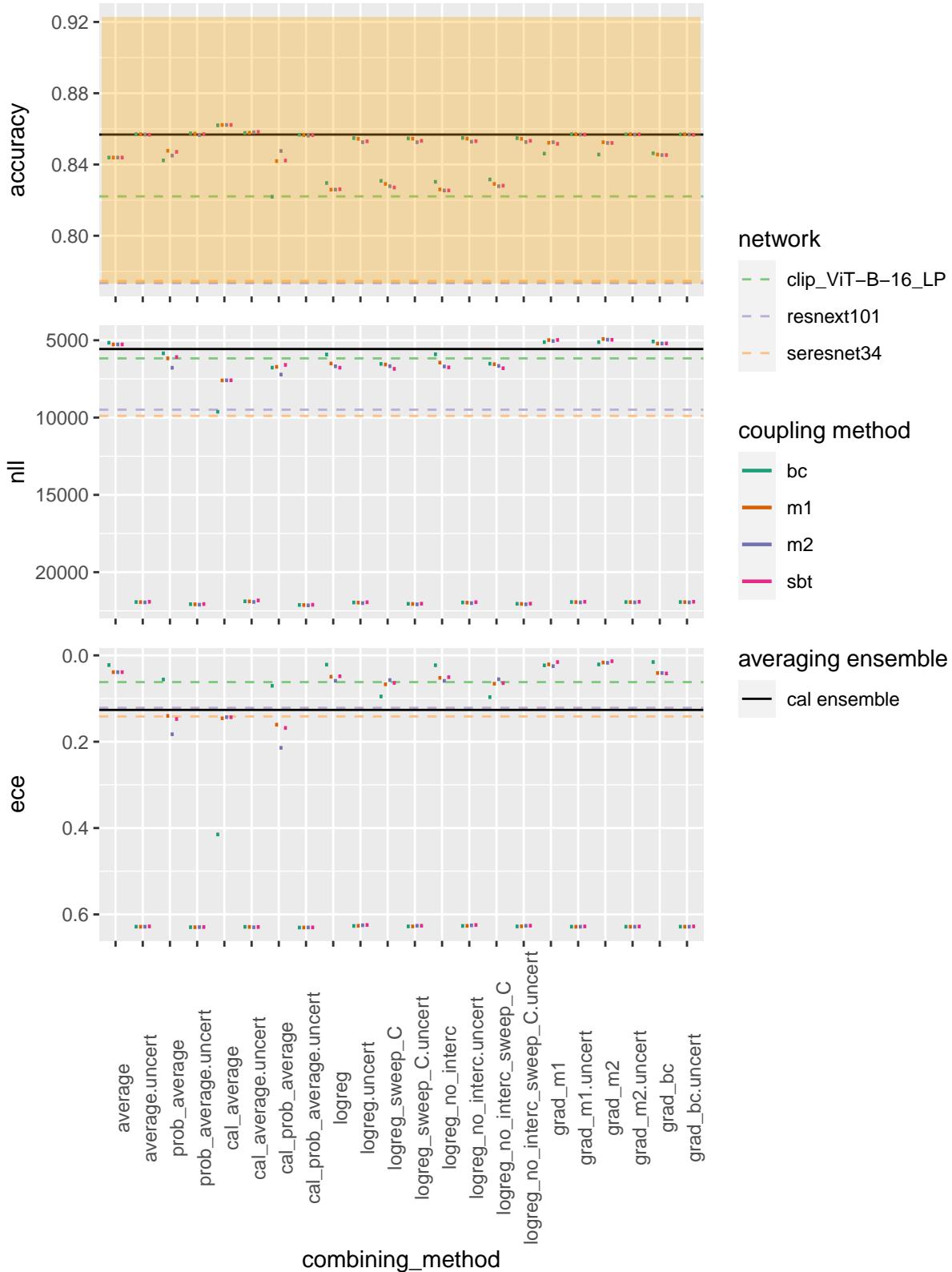
Ensemble metrics
Error inconsistency 0.280099987983704



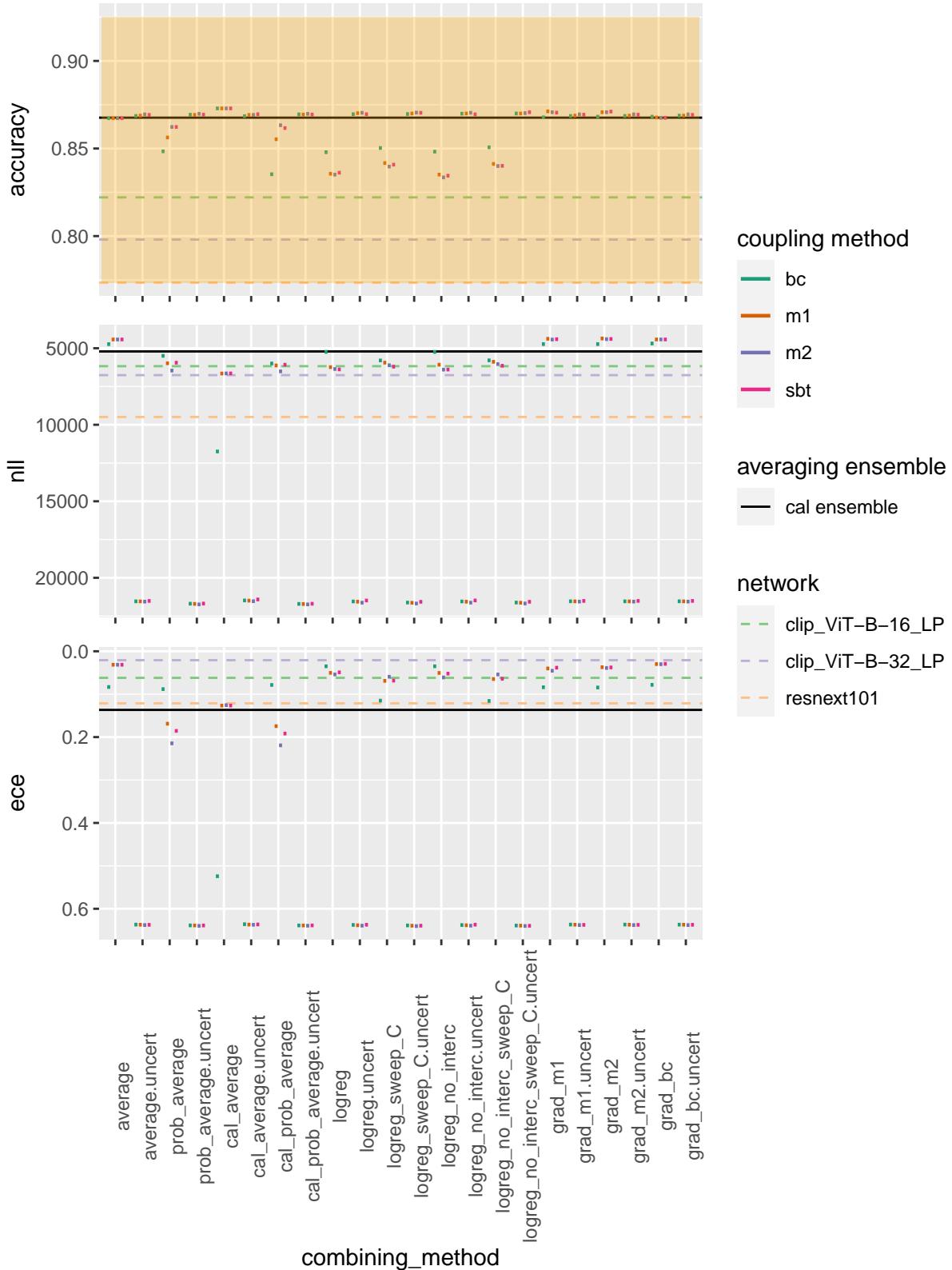
Ensemble metrics
Error inconsistency 0.280000001192093



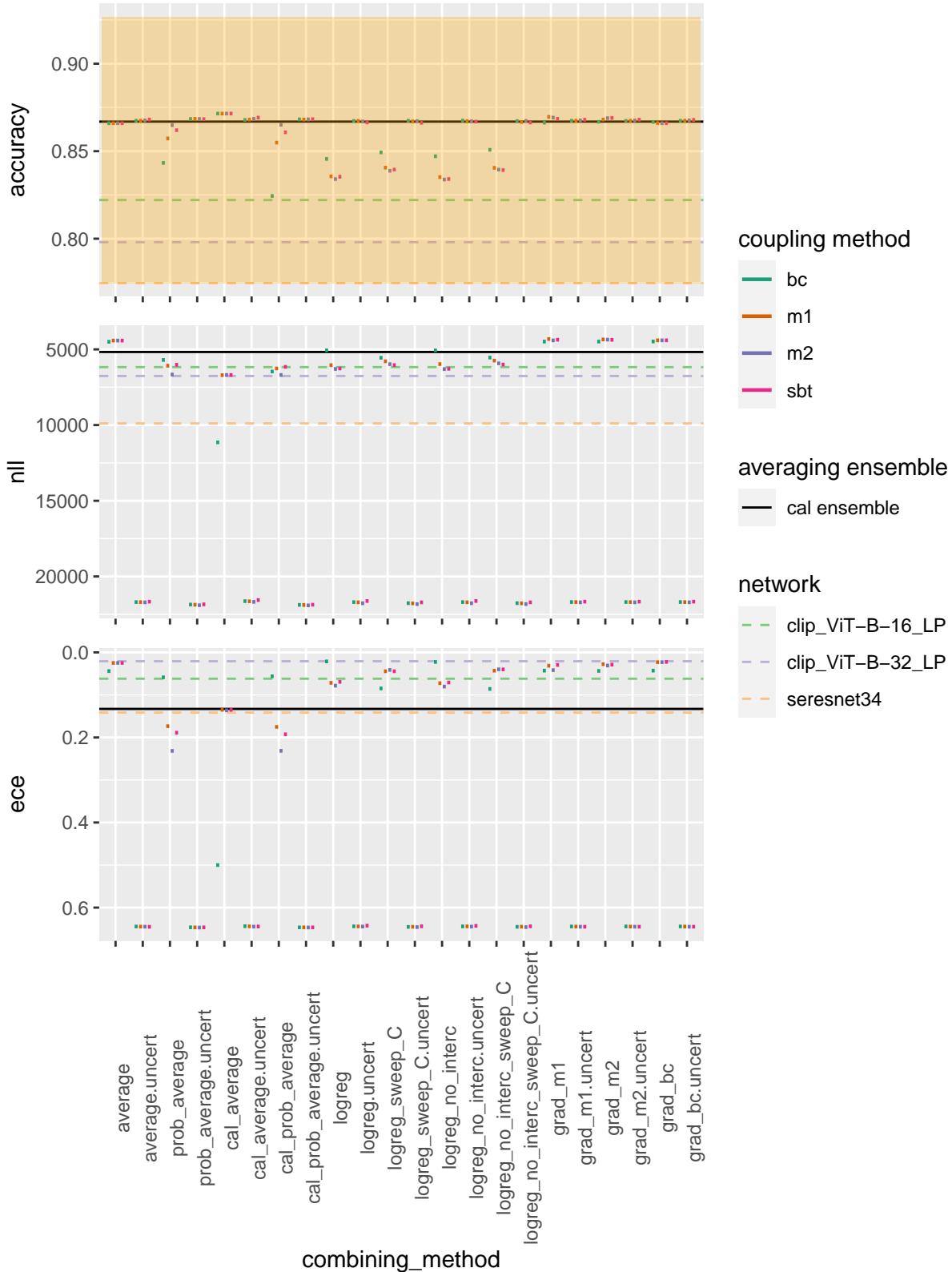
Ensemble metrics
Error inconsistency 0.276600003242493



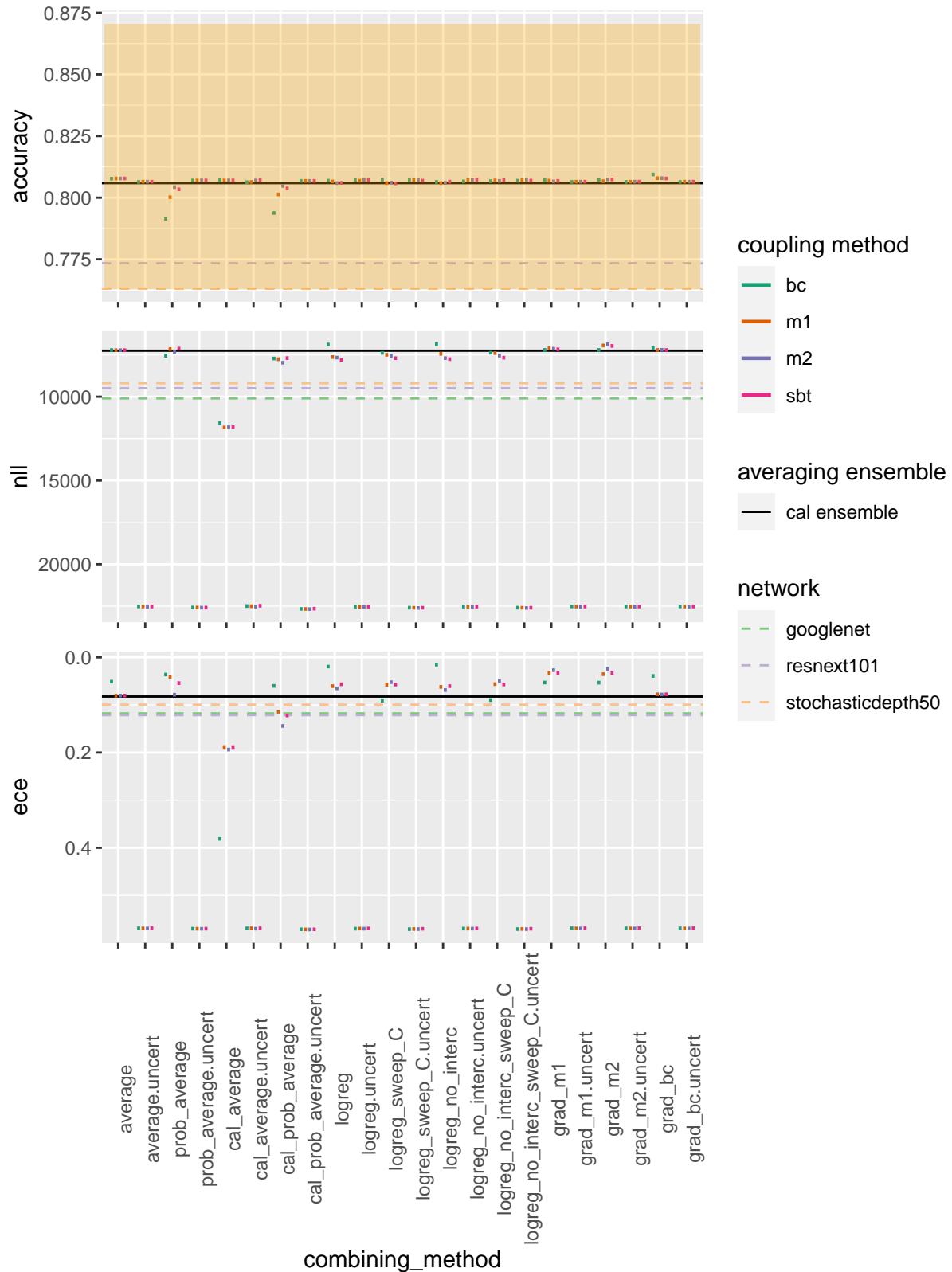
Ensemble metrics
Error inconsistency 0.28099998831749



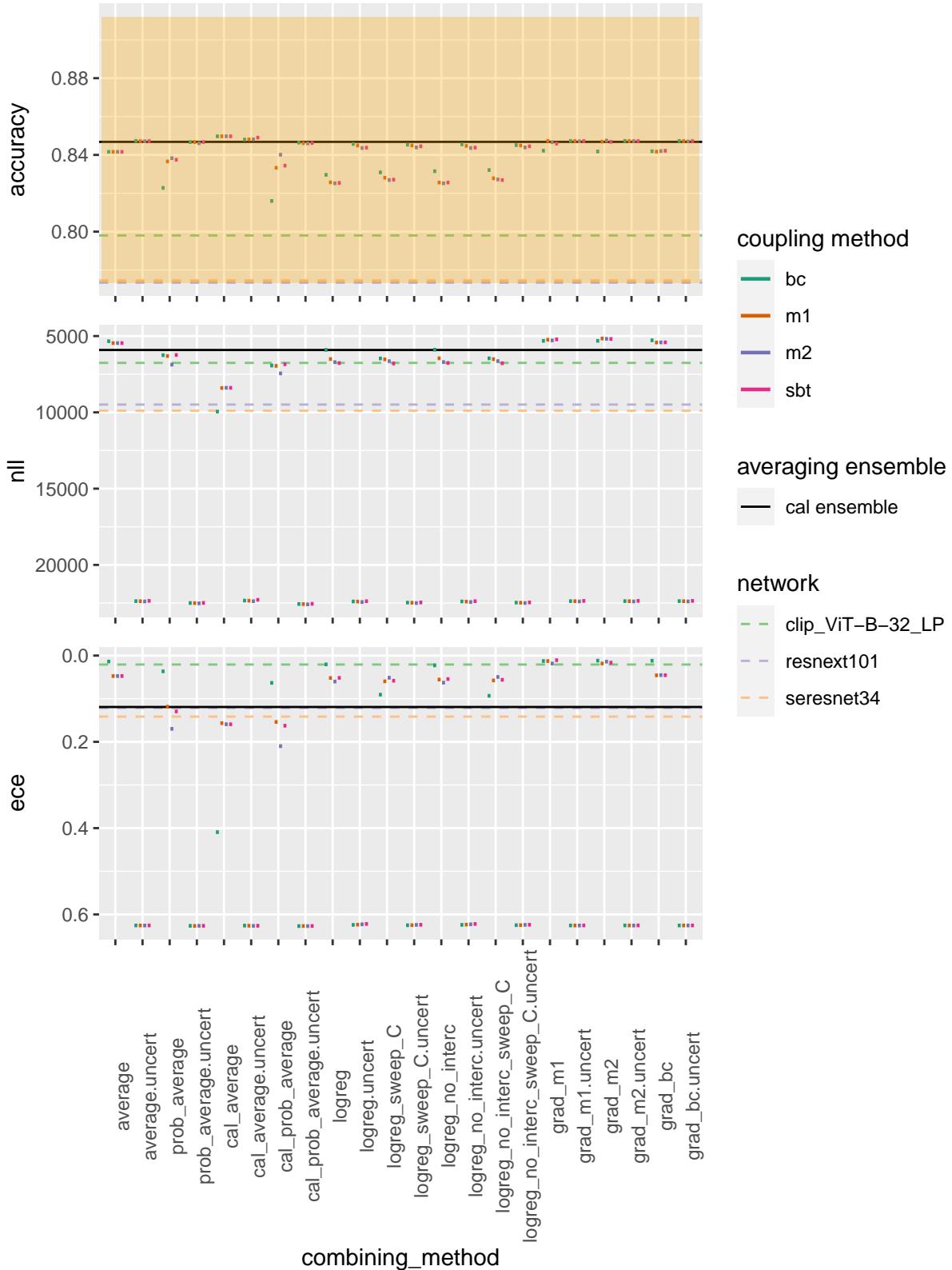
Ensemble metrics
Error inconsistency 0.287099987268448



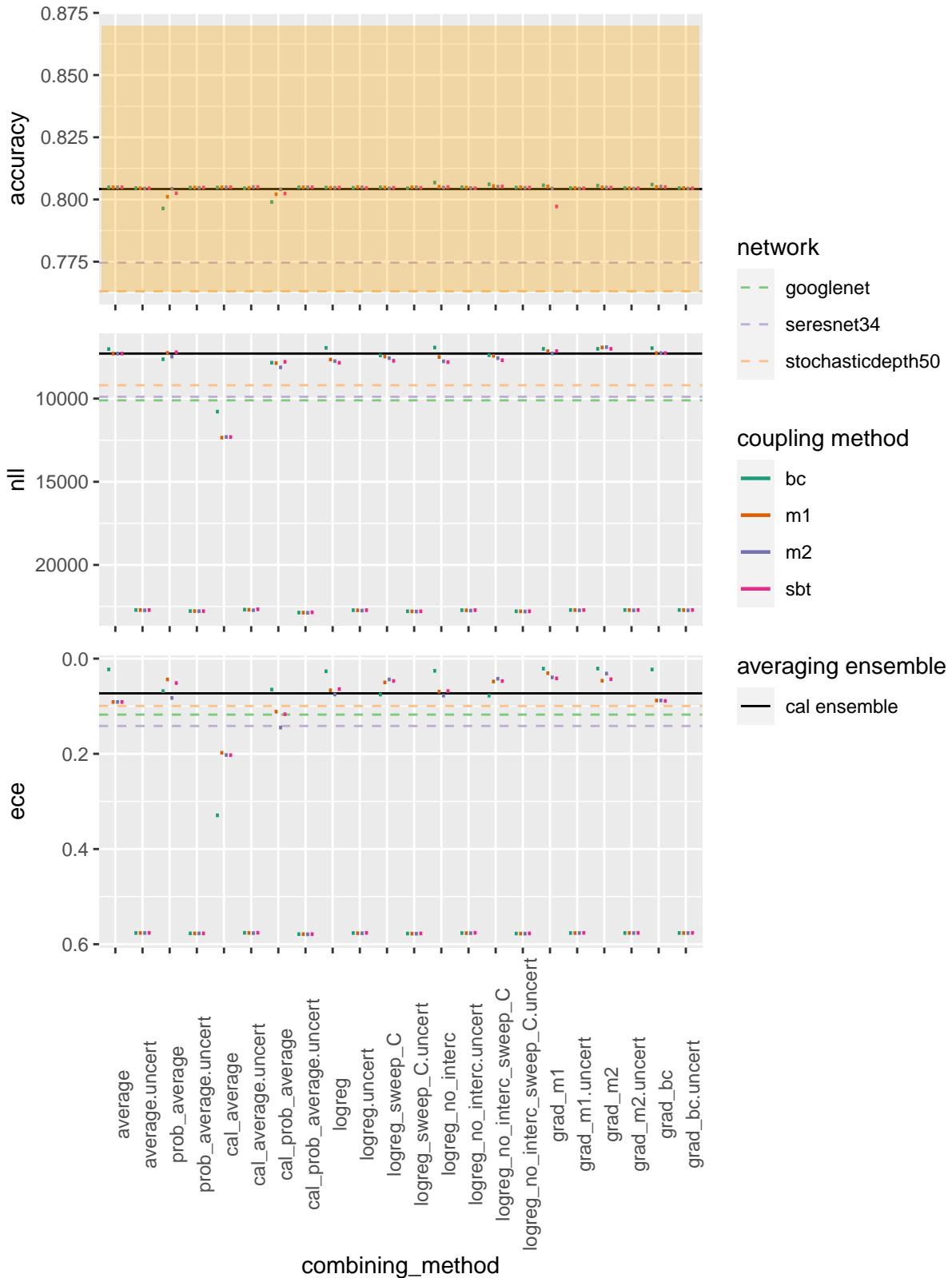
Ensemble metrics
Error inconsistency 0.220100000500679



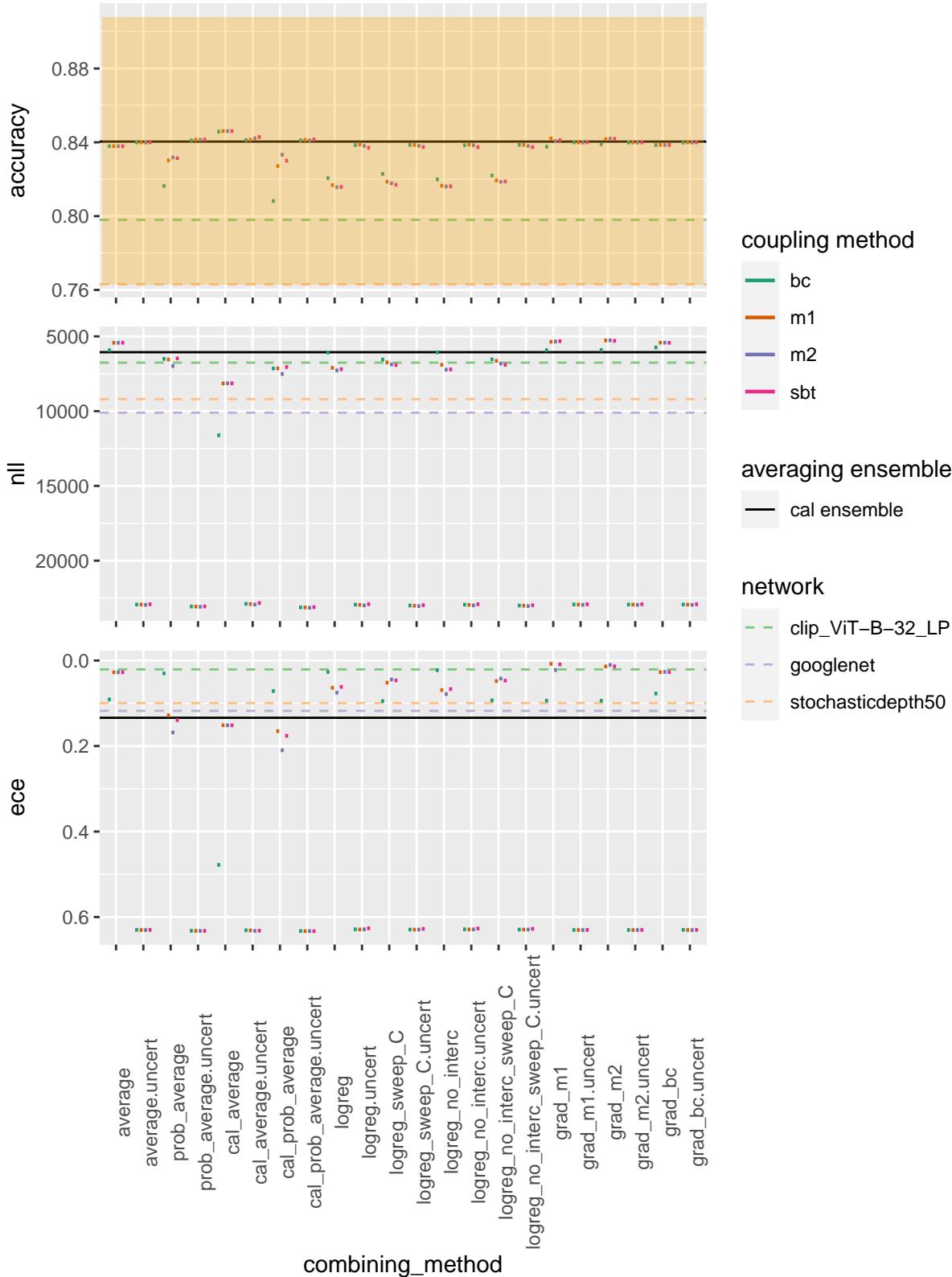
Ensemble metrics
Error inconsistency 0.274399995803833



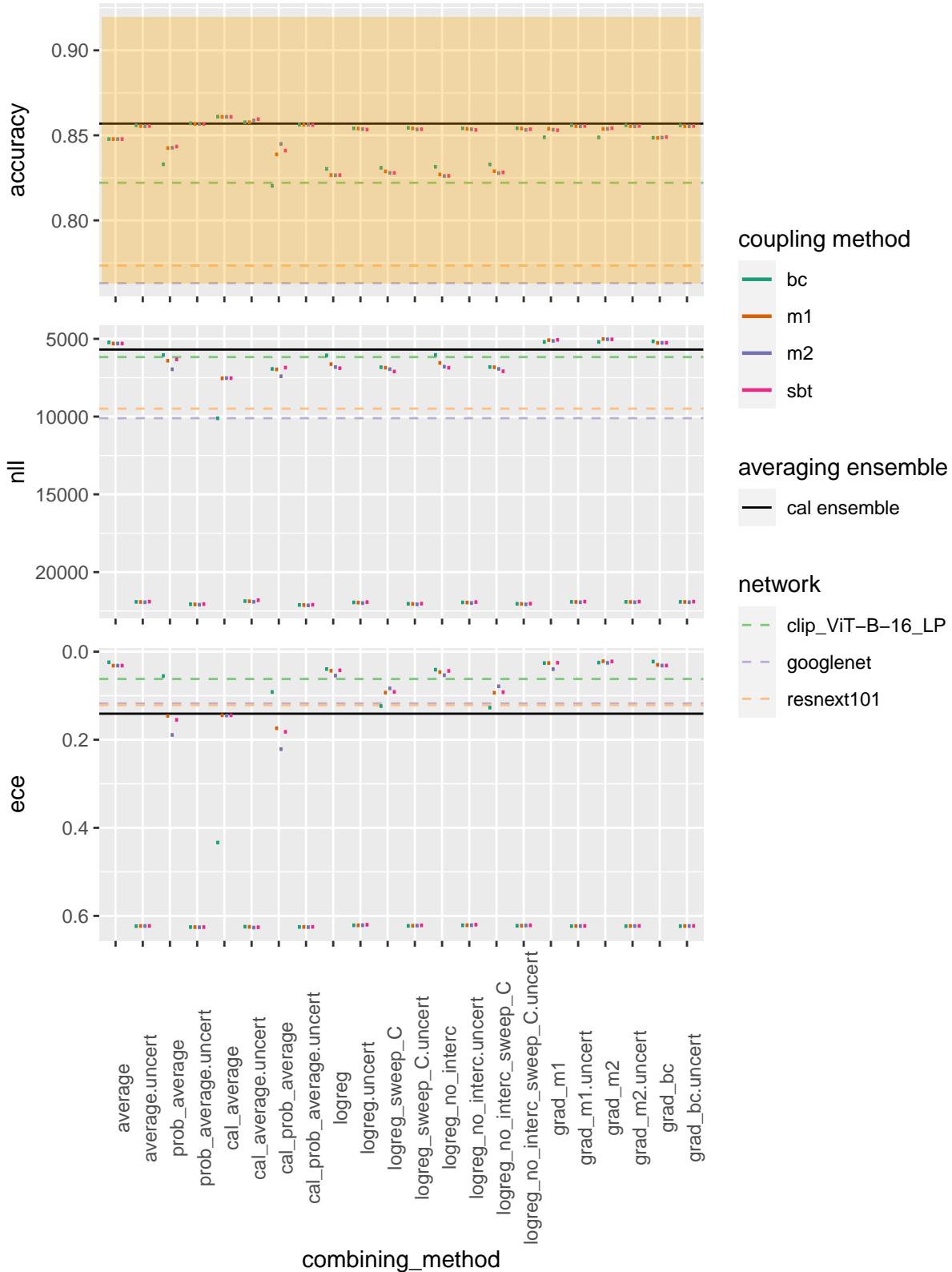
Ensemble metrics
Error inconsistency 0.215799987316132



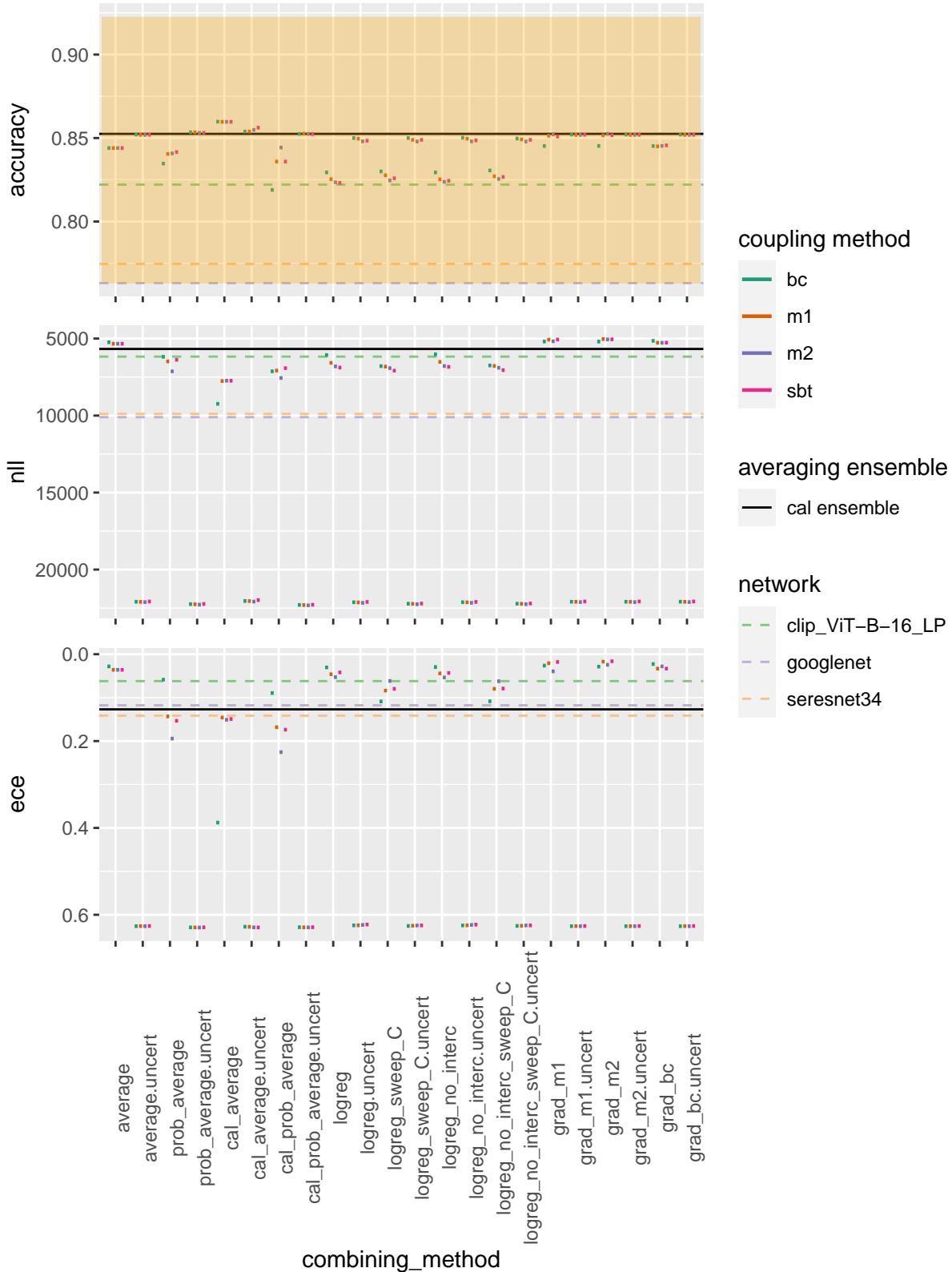
Ensemble metrics
Error inconsistency 0.282700002193451



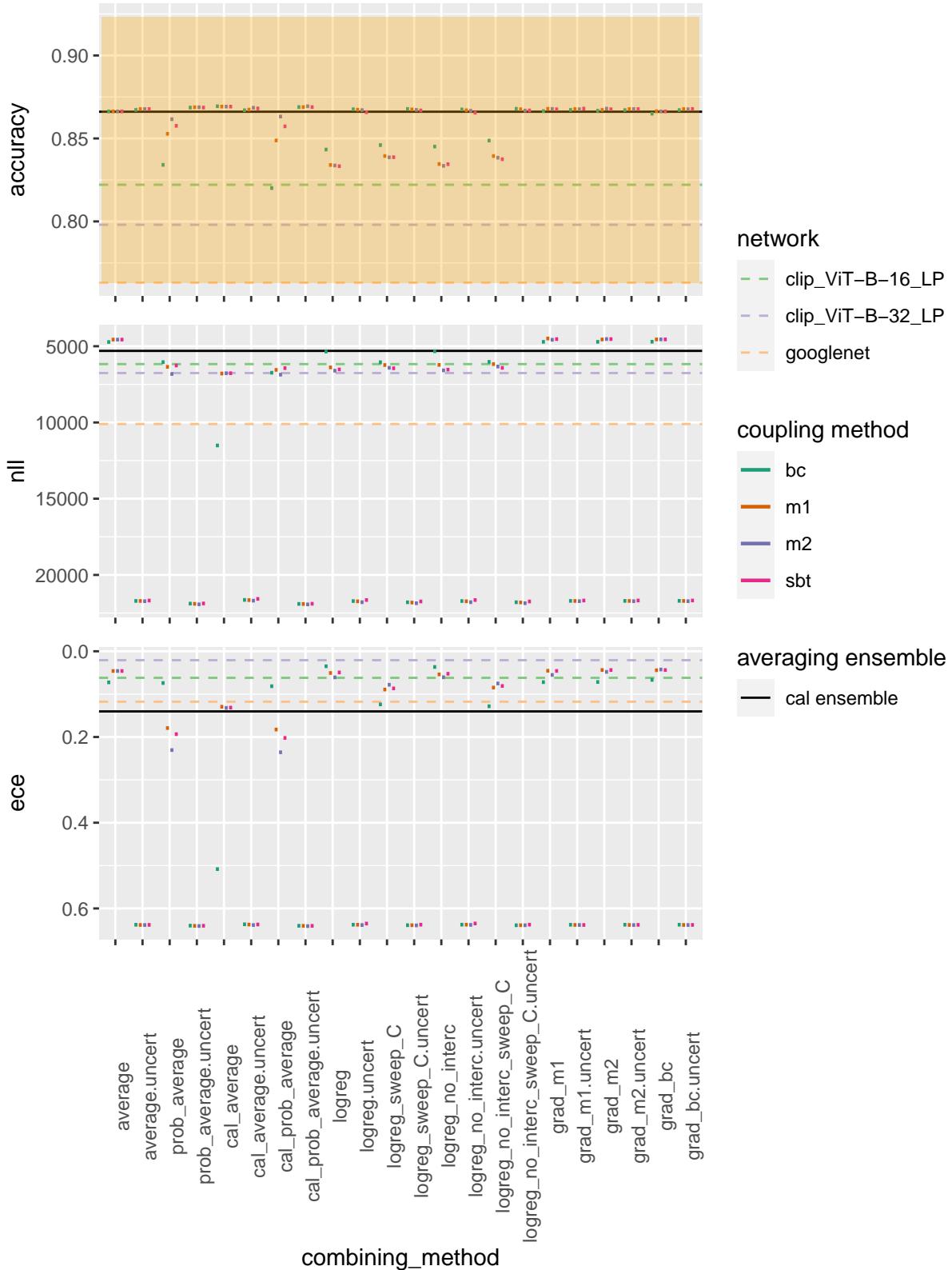
Ensemble metrics
Error inconsistency 0.281699985265732



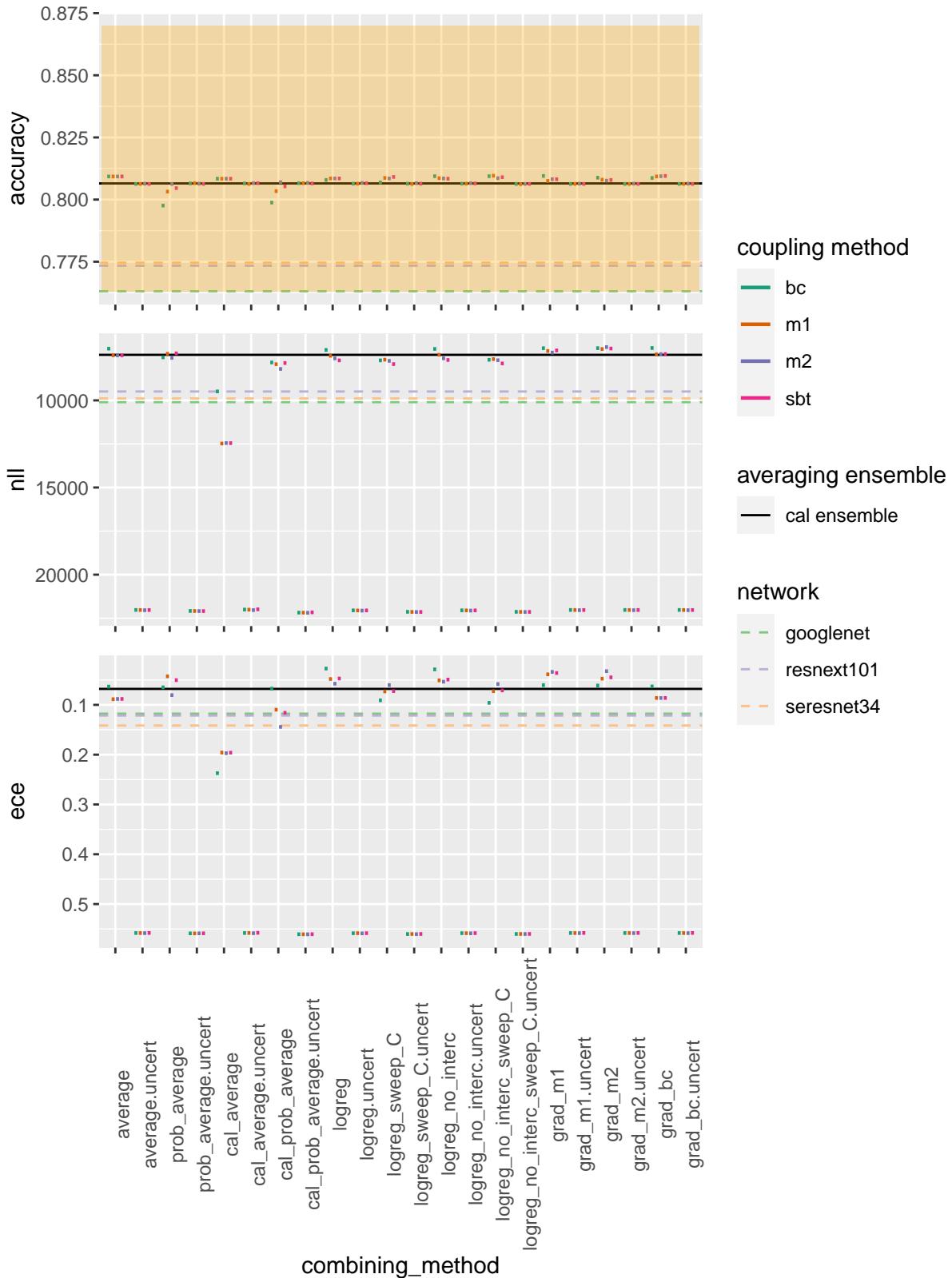
Ensemble metrics
Error inconsistency 0.283600002527237



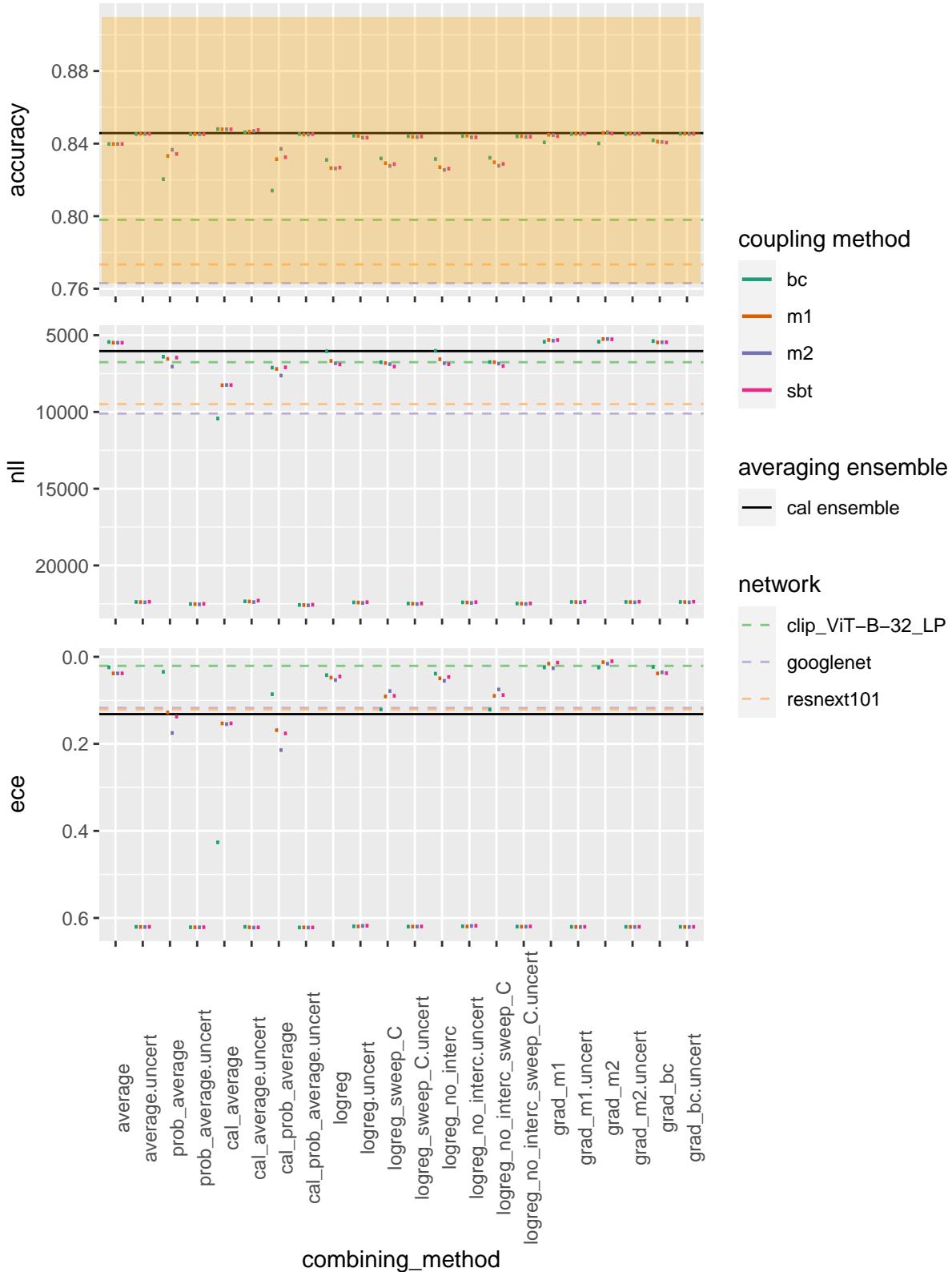
Ensemble metrics
Error inconsistency 0.287599980831146



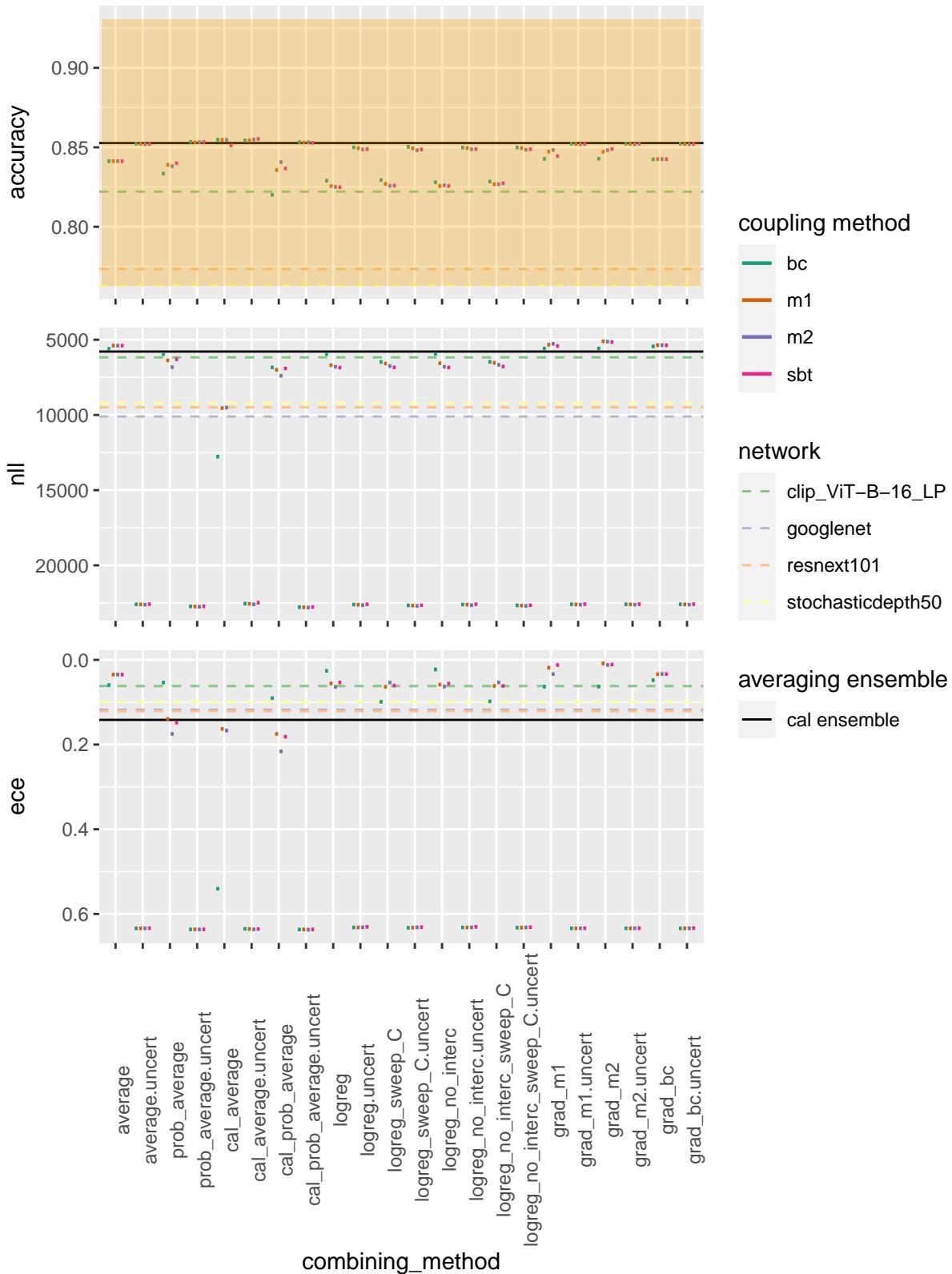
Ensemble metrics
Error inconsistency 0.209099993109703



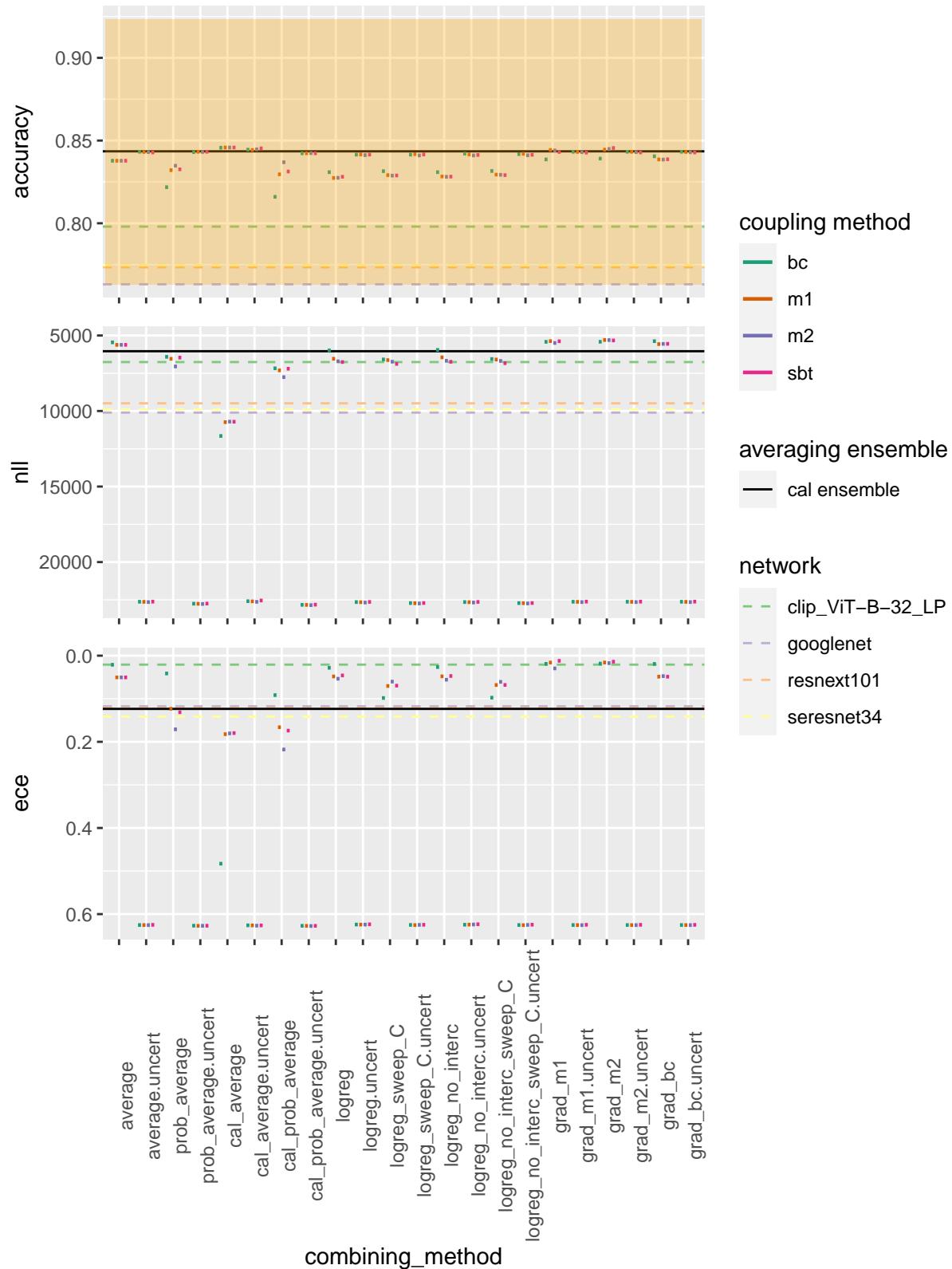
Ensemble metrics
Error inconsistency 0.277700006961823



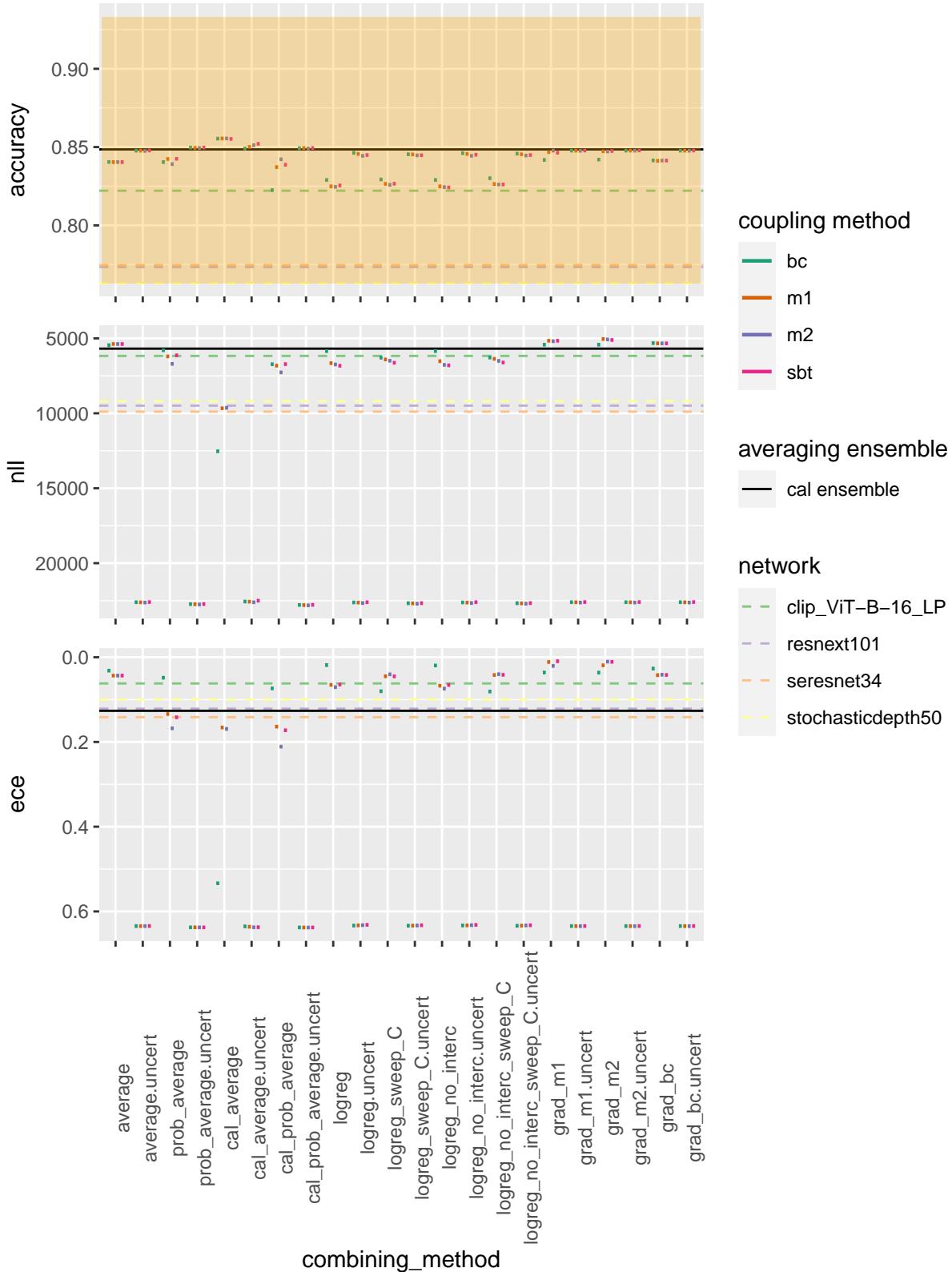
Ensemble metrics
Error inconsistency 0.326900005340576



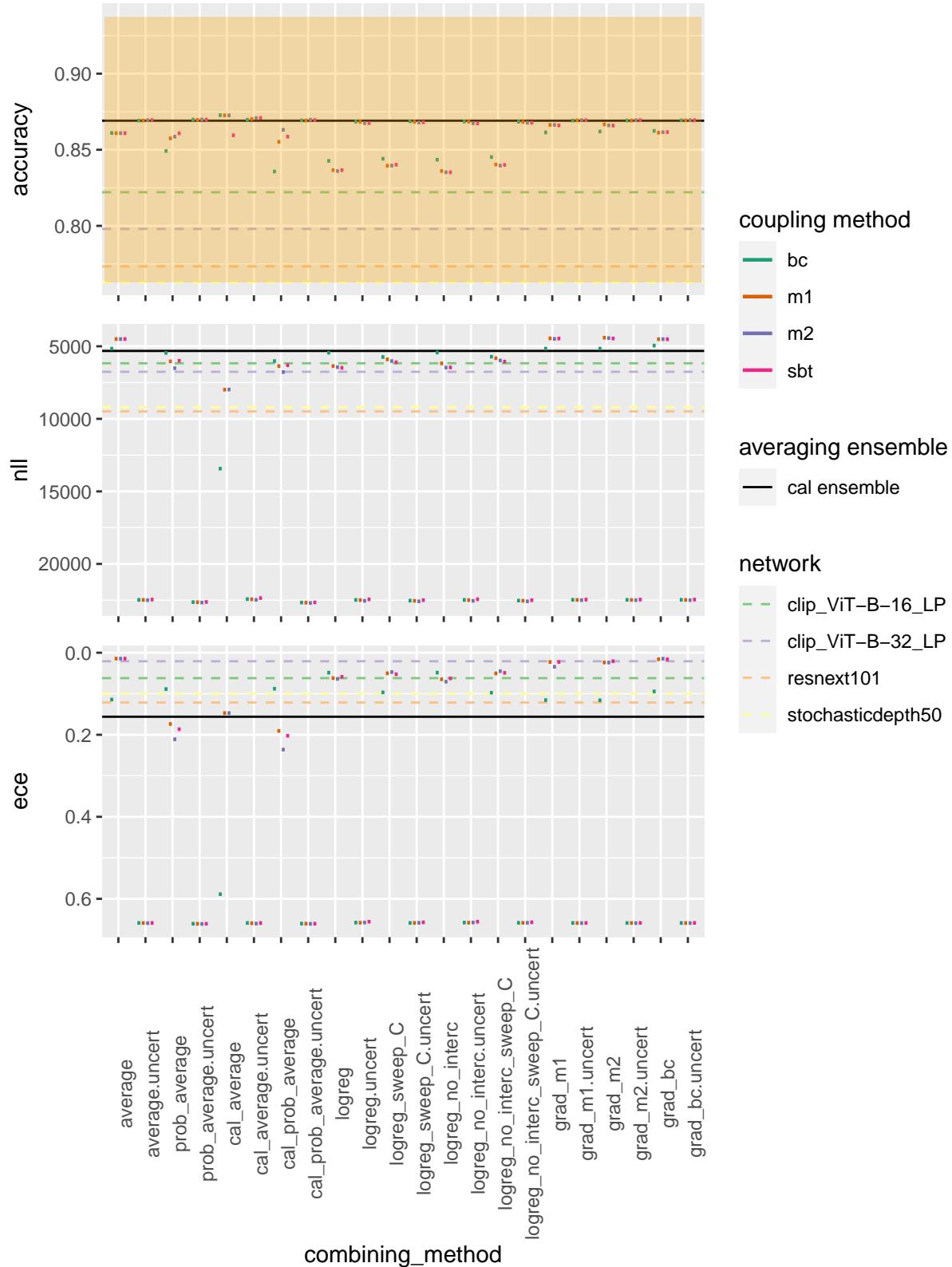
Ensemble metrics
Error inconsistency 0.319399982690811



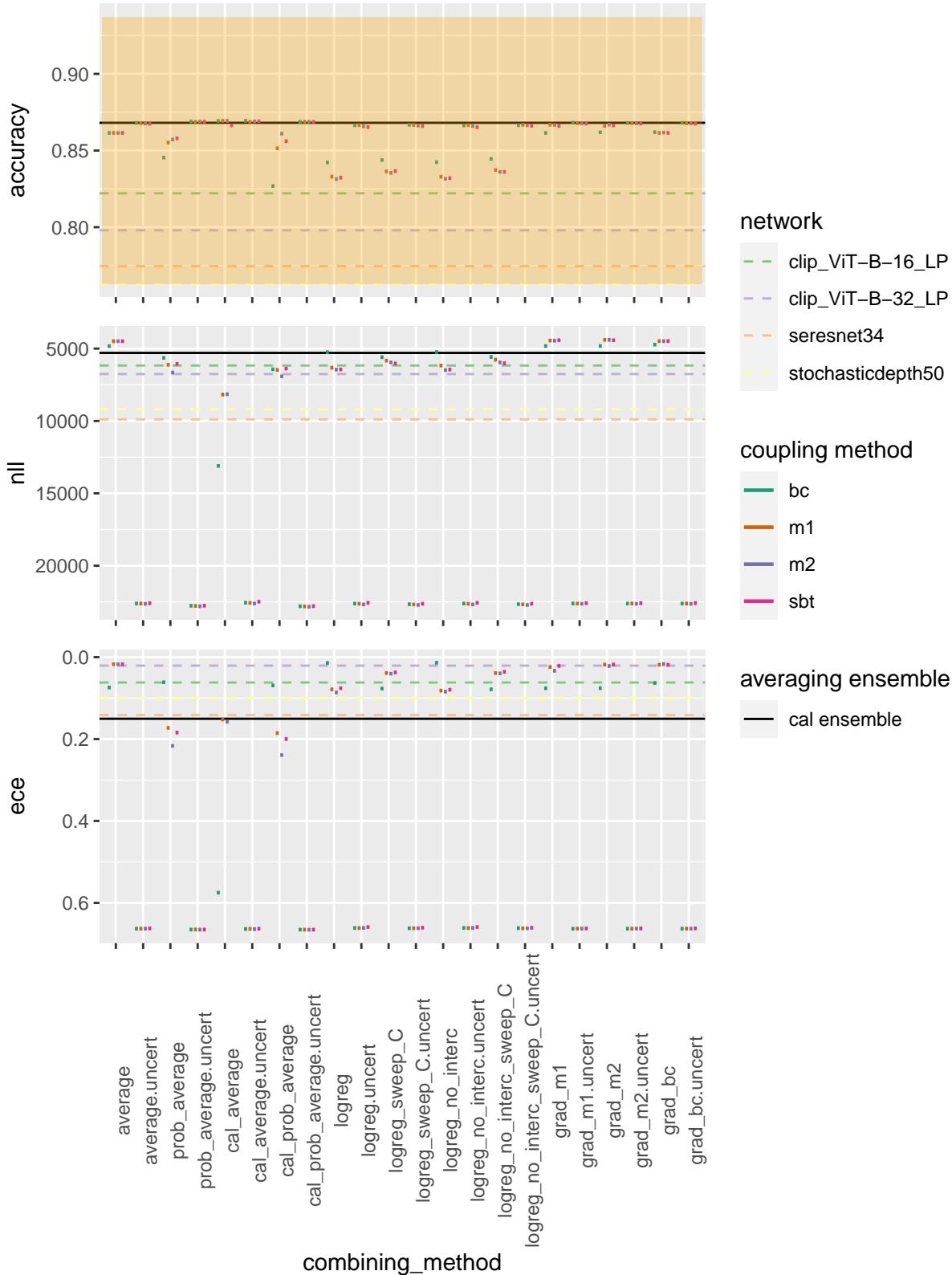
Ensemble metrics
Error inconsistency 0.321399986743927



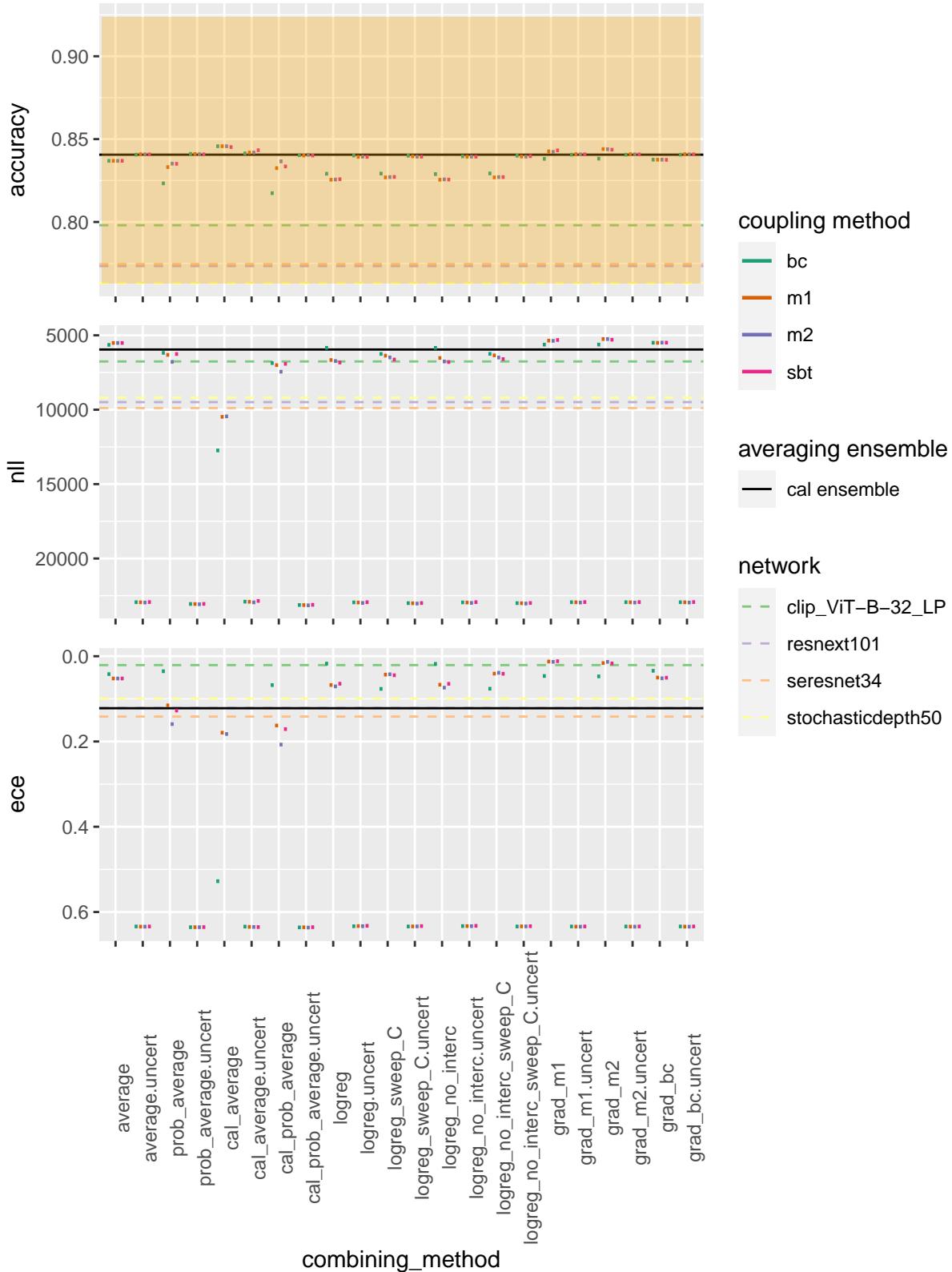
Ensemble metrics
Error inconsistency 0.337999999523163



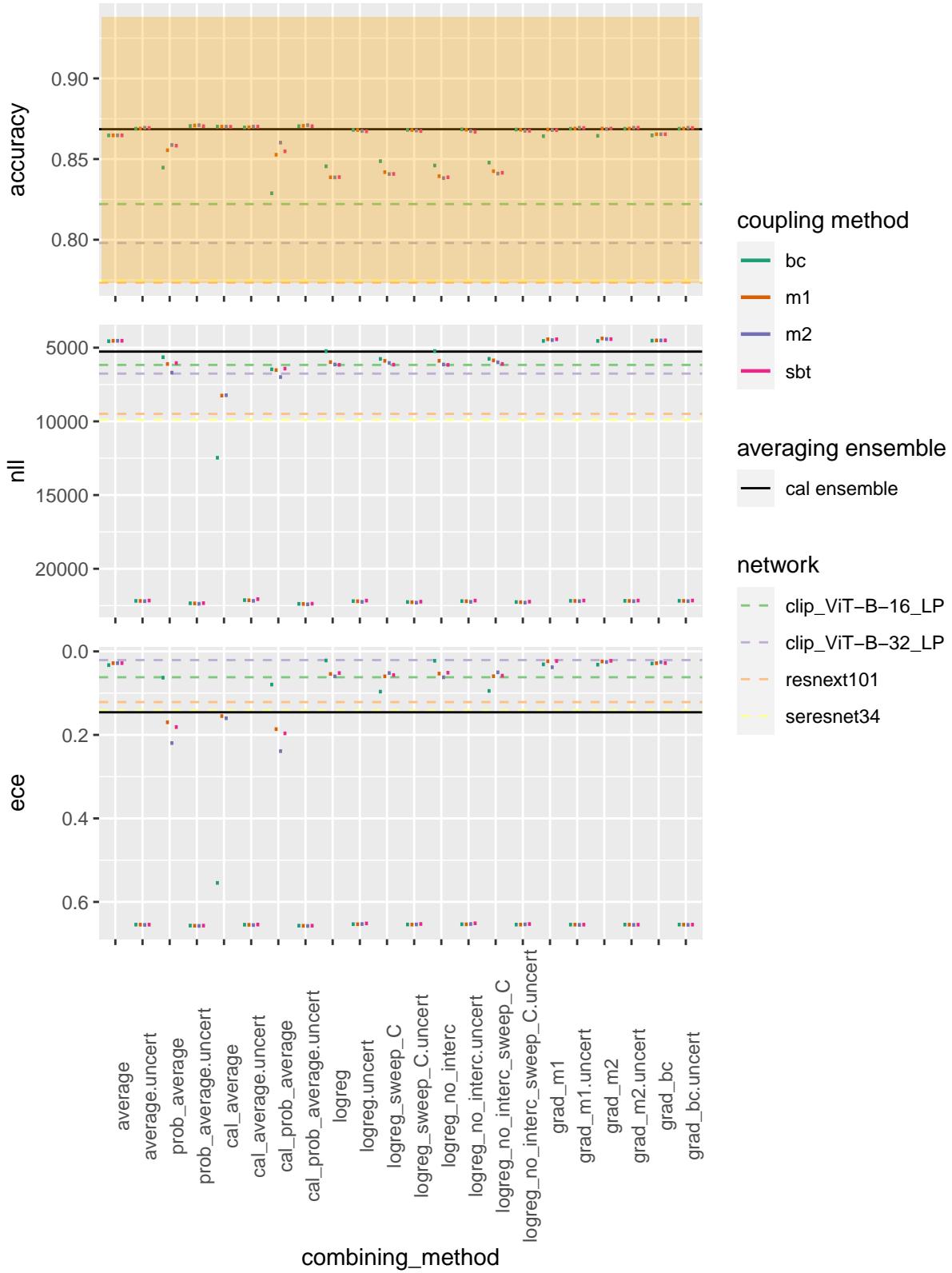
Ensemble metrics
Error inconsistency 0.338499993085861



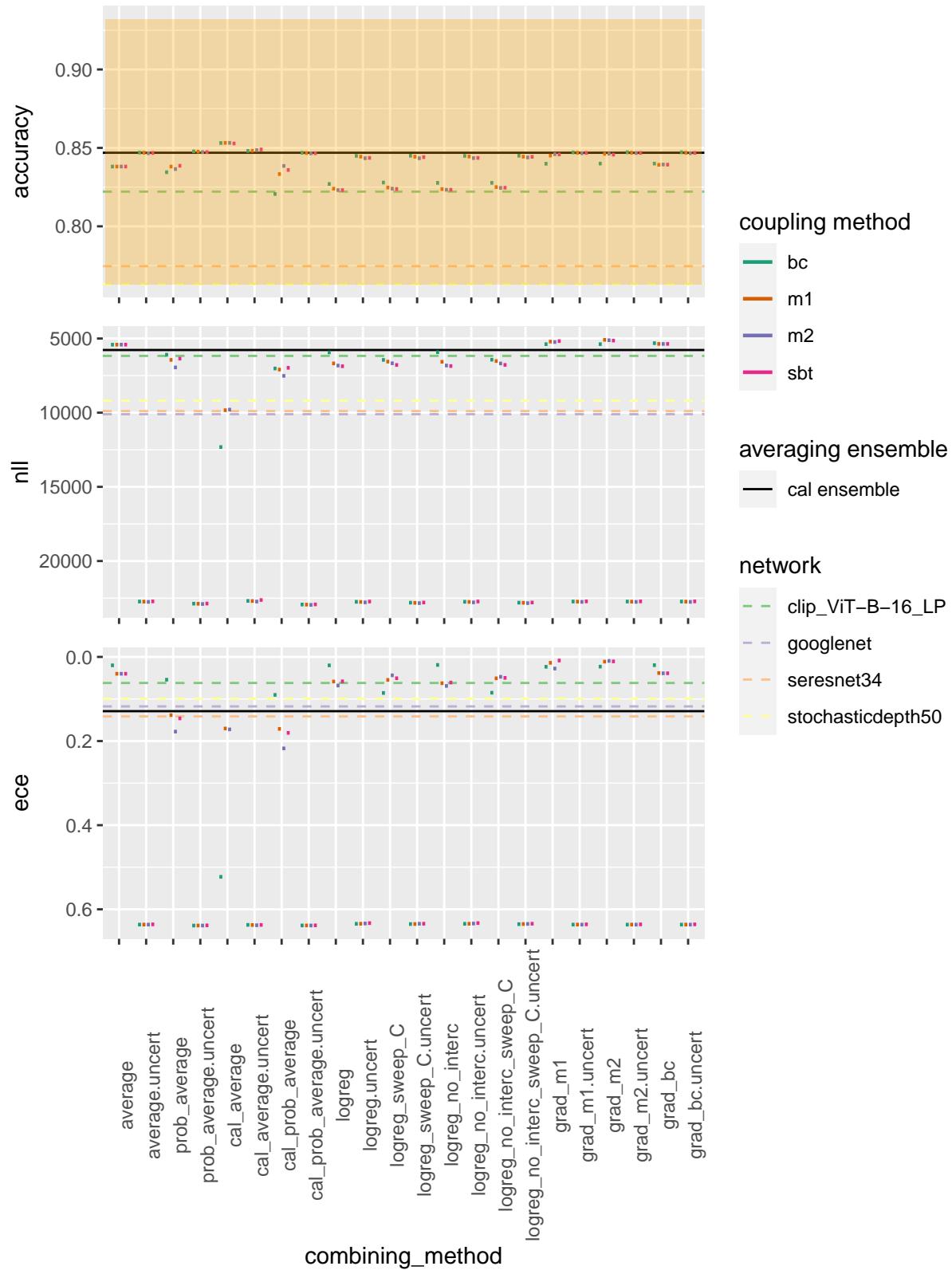
Ensemble metrics
Error inconsistency 0.319199979305267



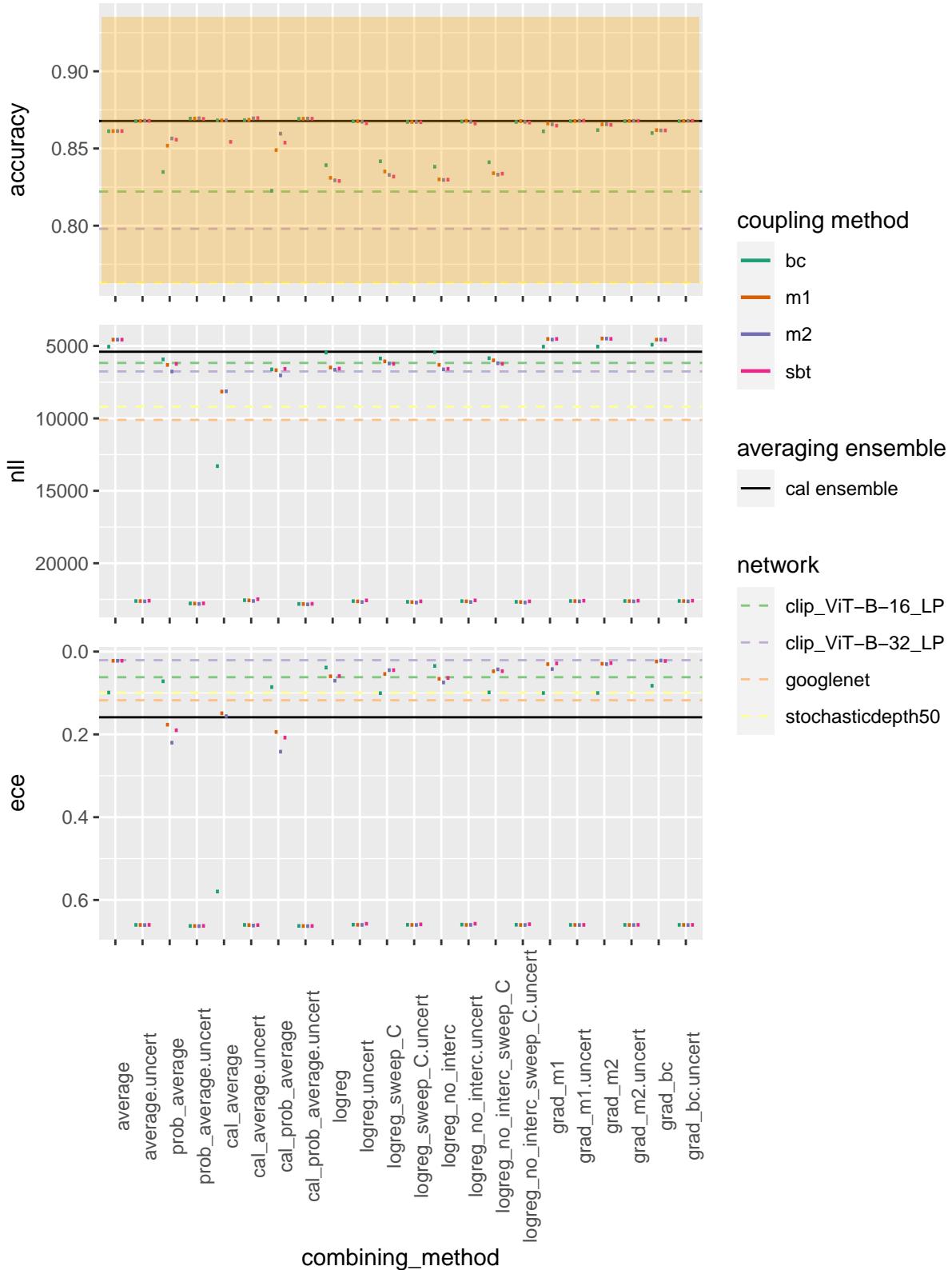
Ensemble metrics
Error inconsistency 0.333099991083145



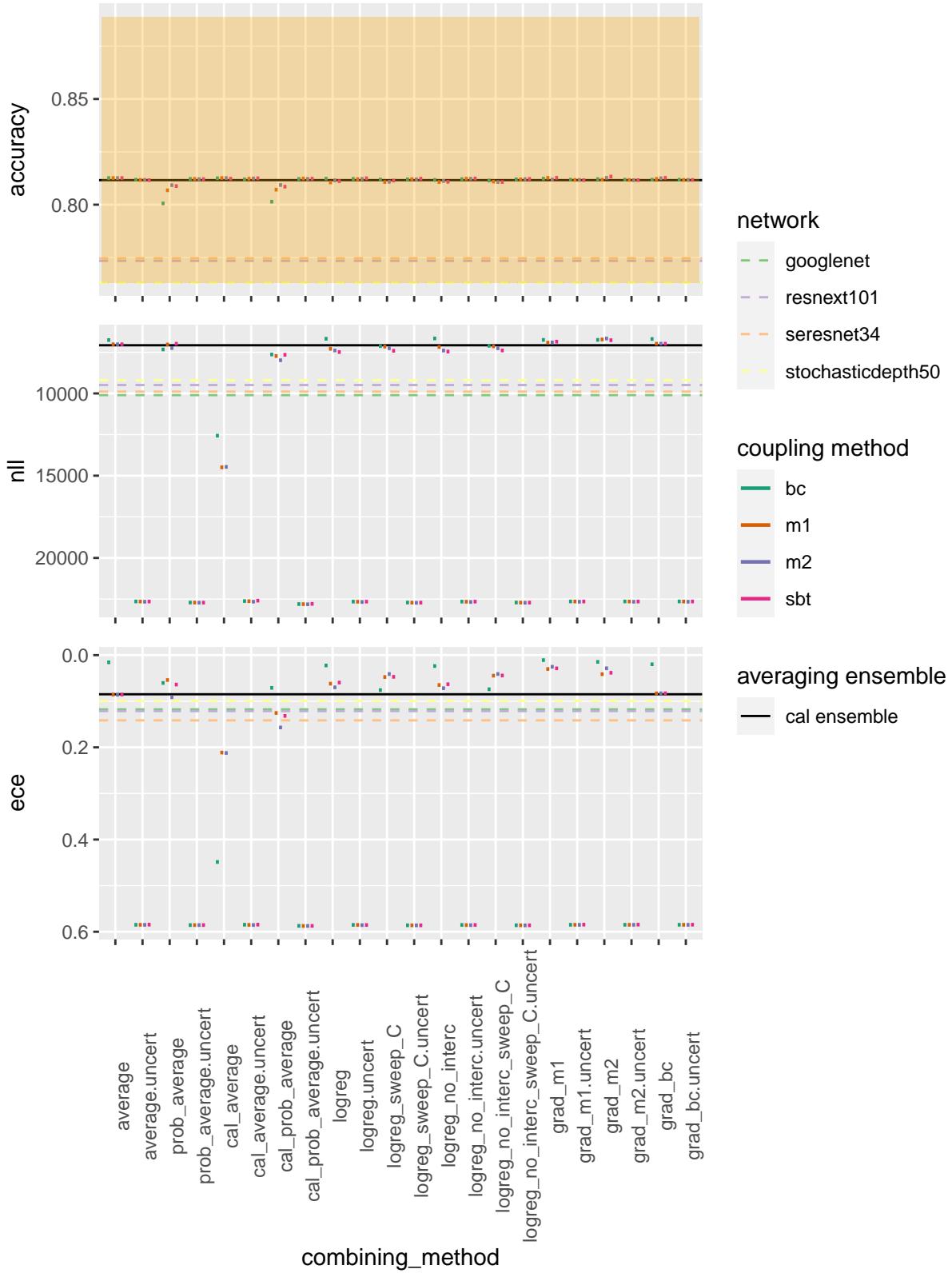
Ensemble metrics
Error inconsistency 0.326499998569489



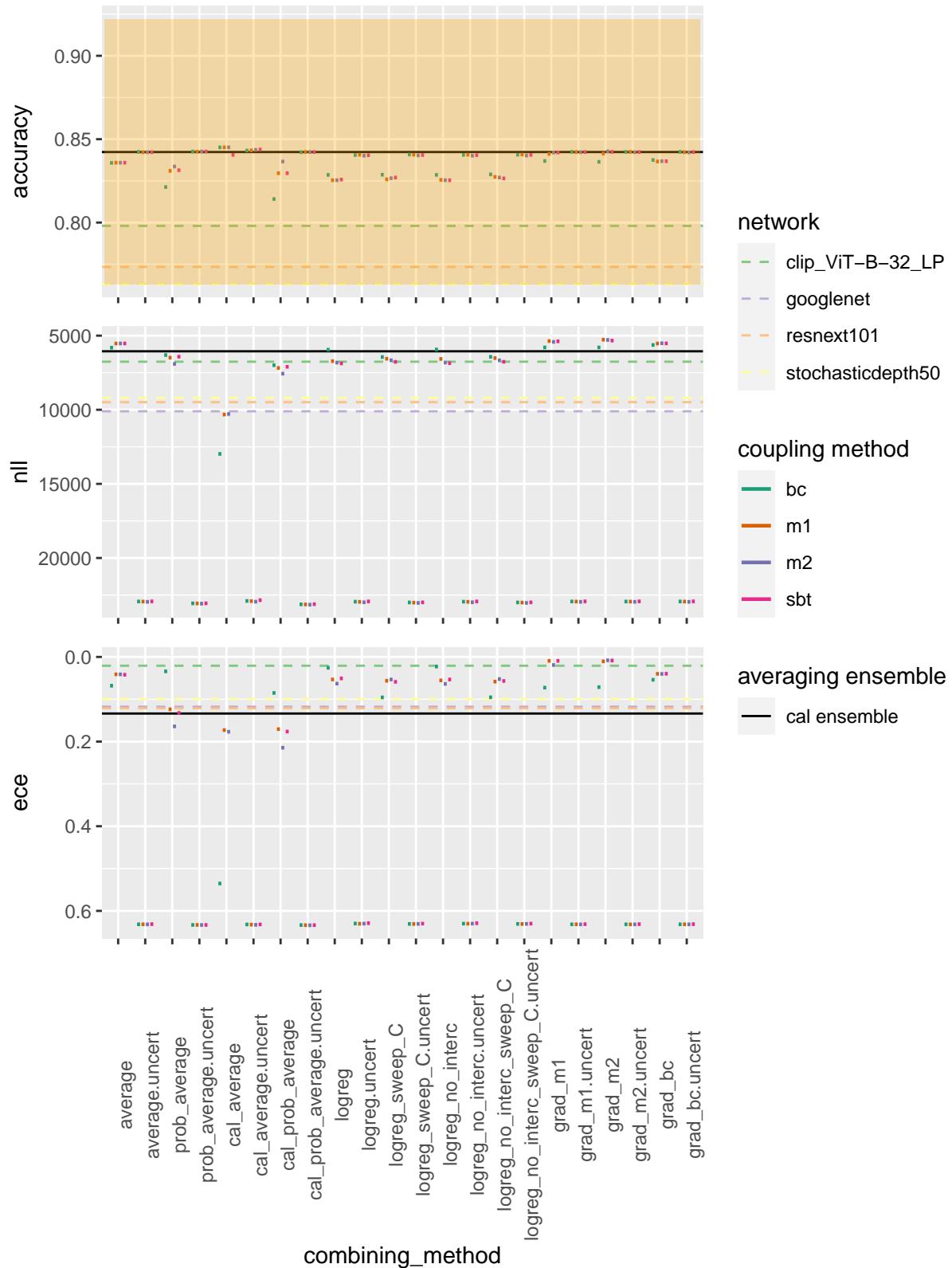
Ensemble metrics
Error inconsistency 0.341999977827072



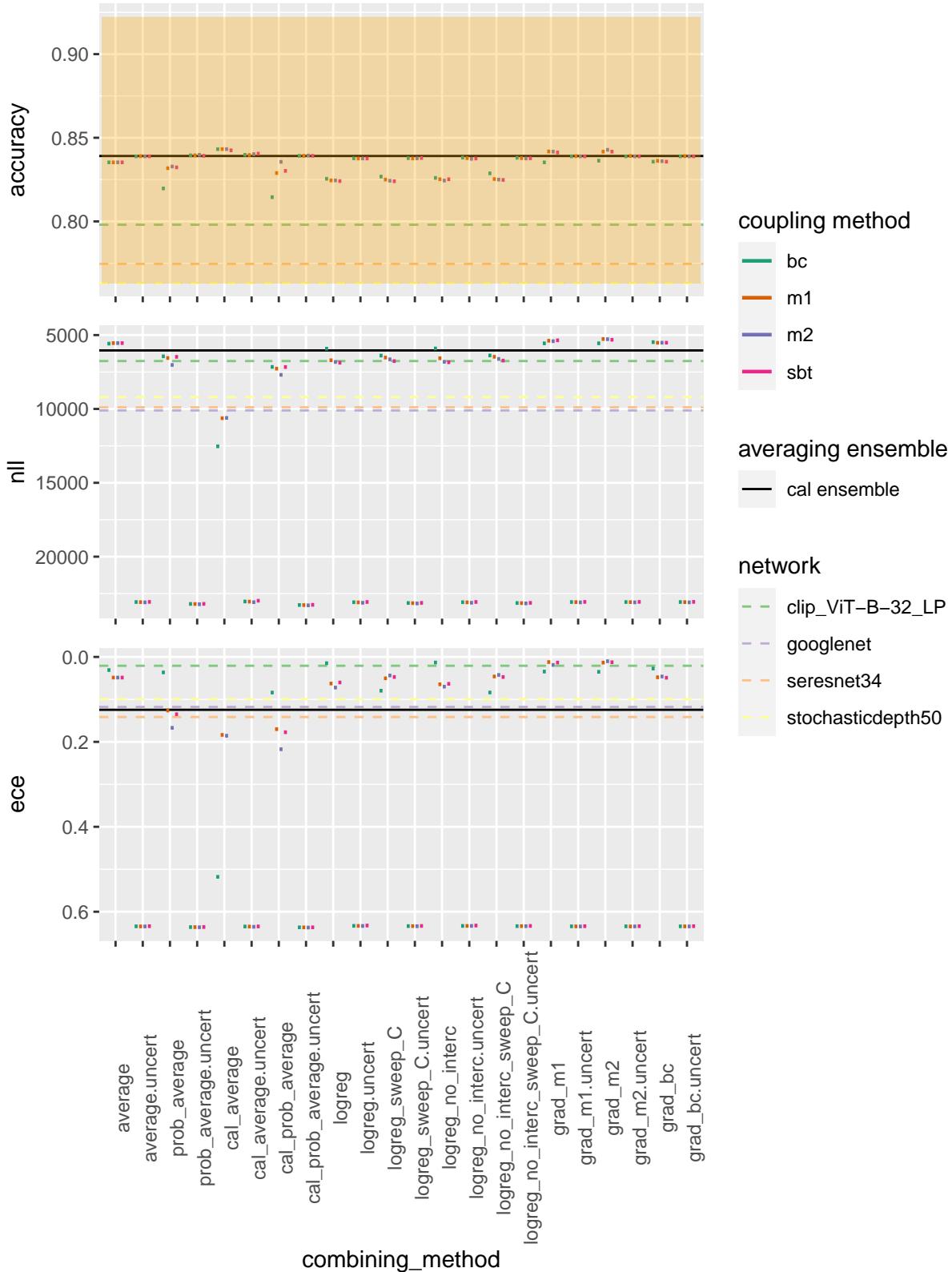
Ensemble metrics
Error inconsistency 0.260600000619888



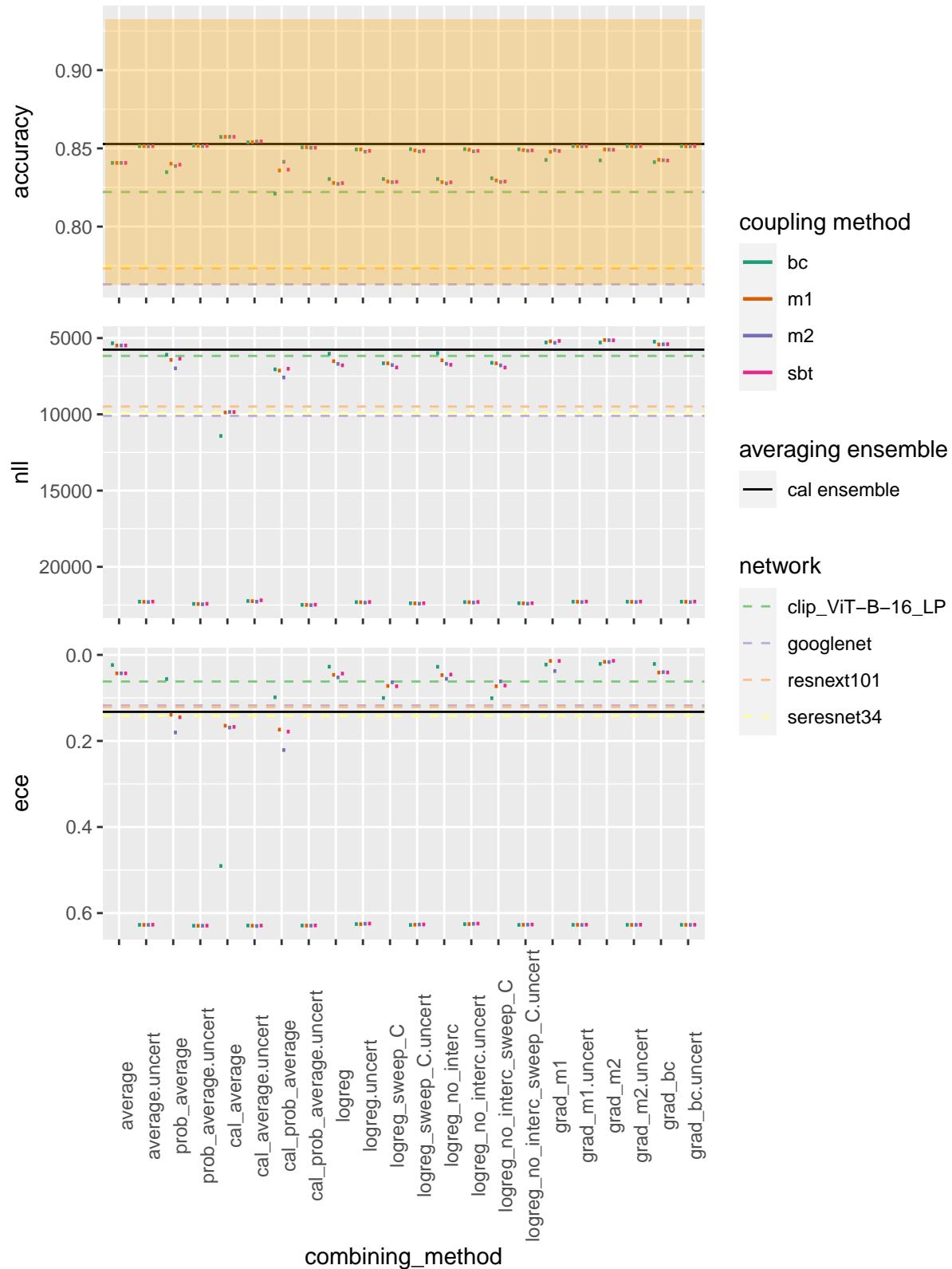
Ensemble metrics
Error inconsistency 0.324099987745285



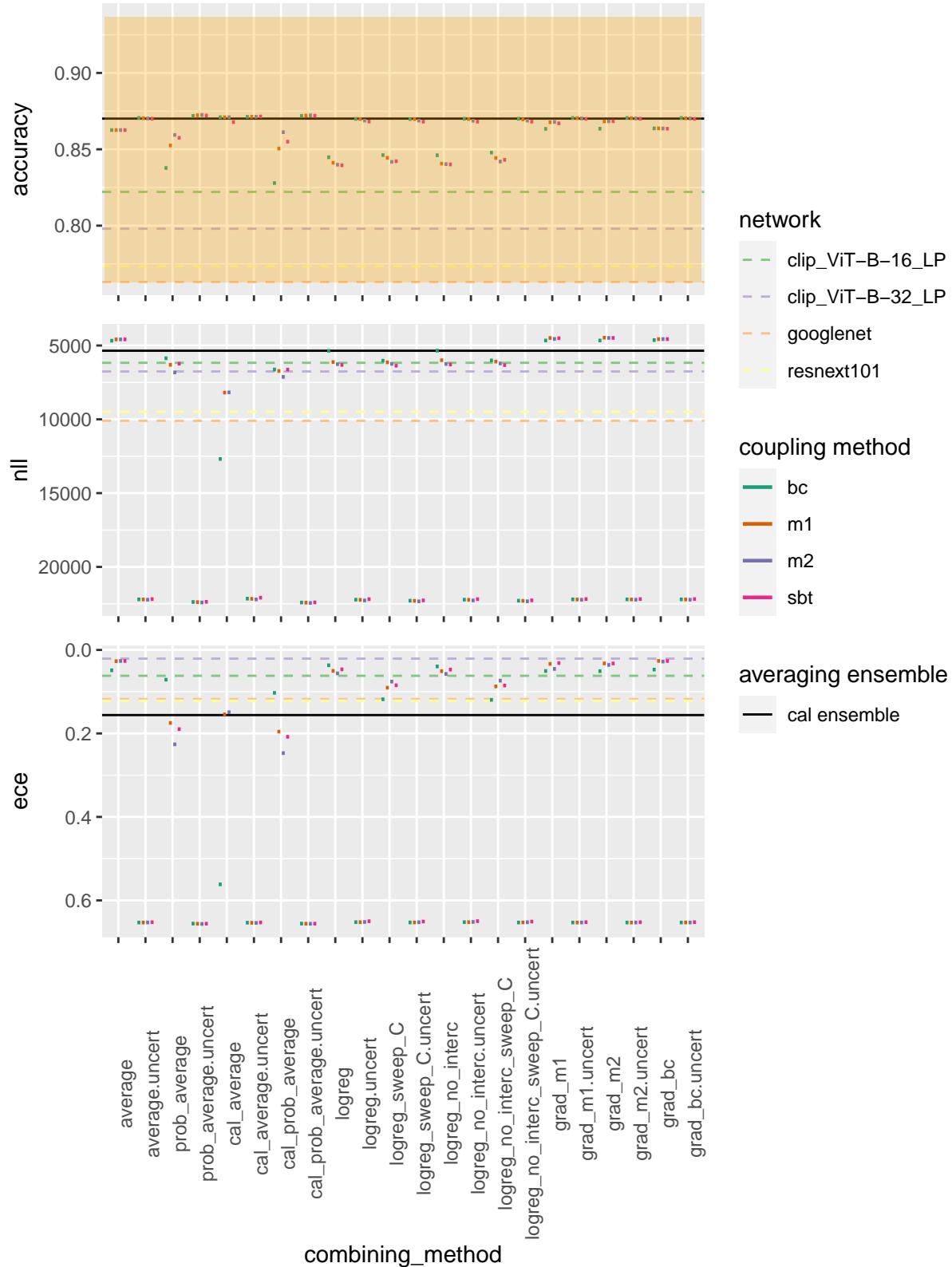
Ensemble metrics
Error inconsistency 0.323499977588654



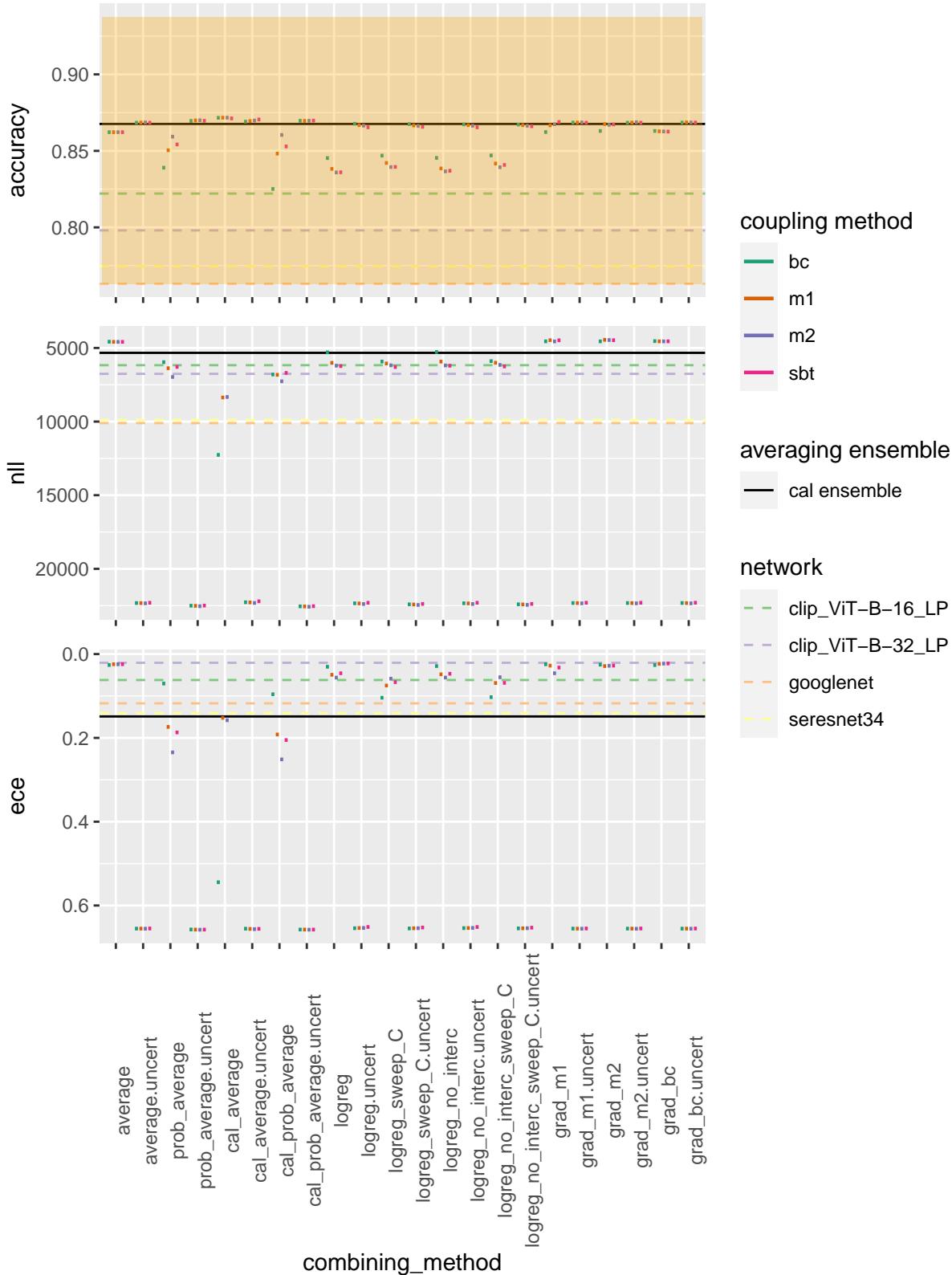
Ensemble metrics
Error inconsistency 0.322099983692169



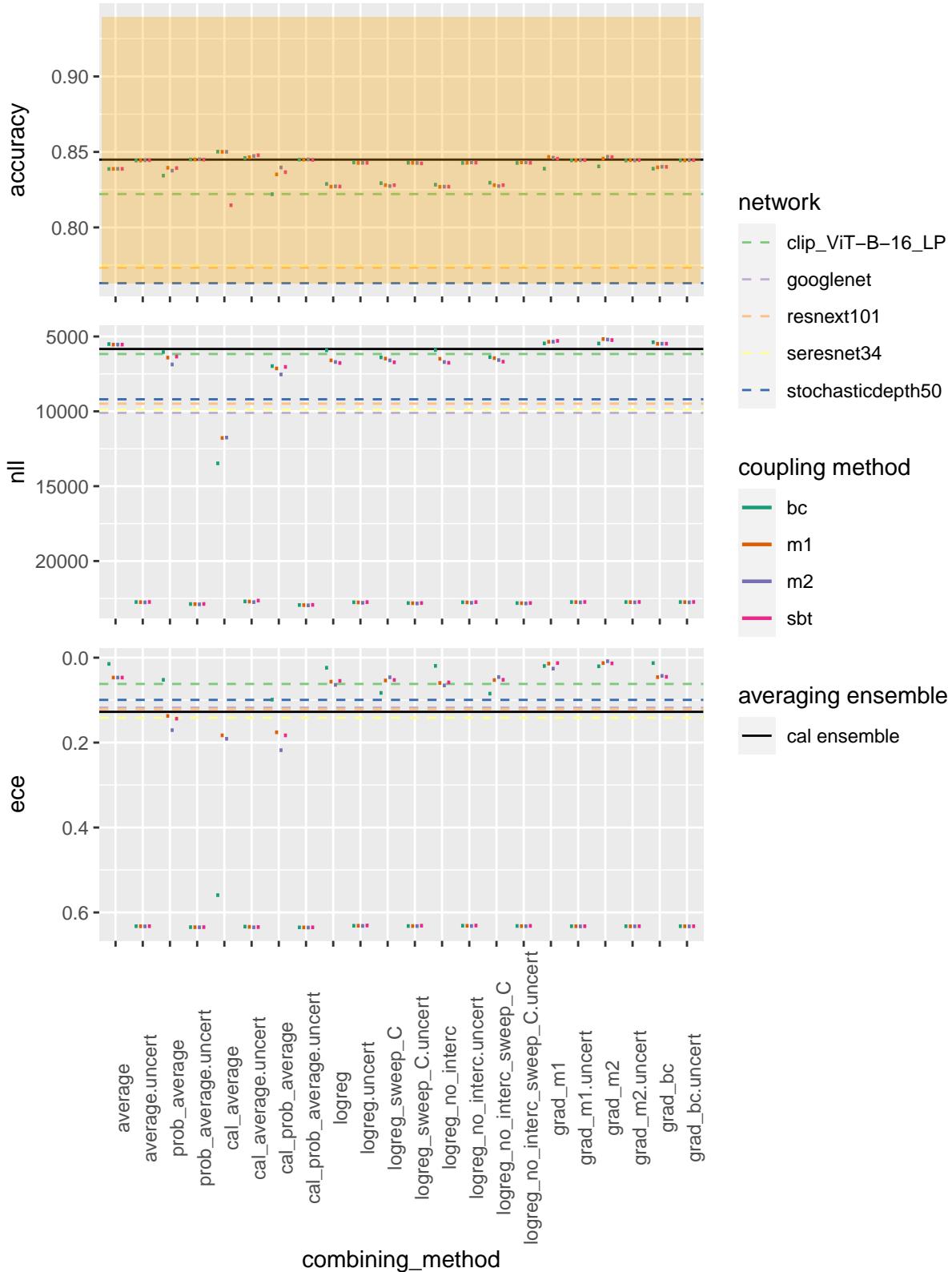
Ensemble metrics
Error inconsistency 0.337799996137619



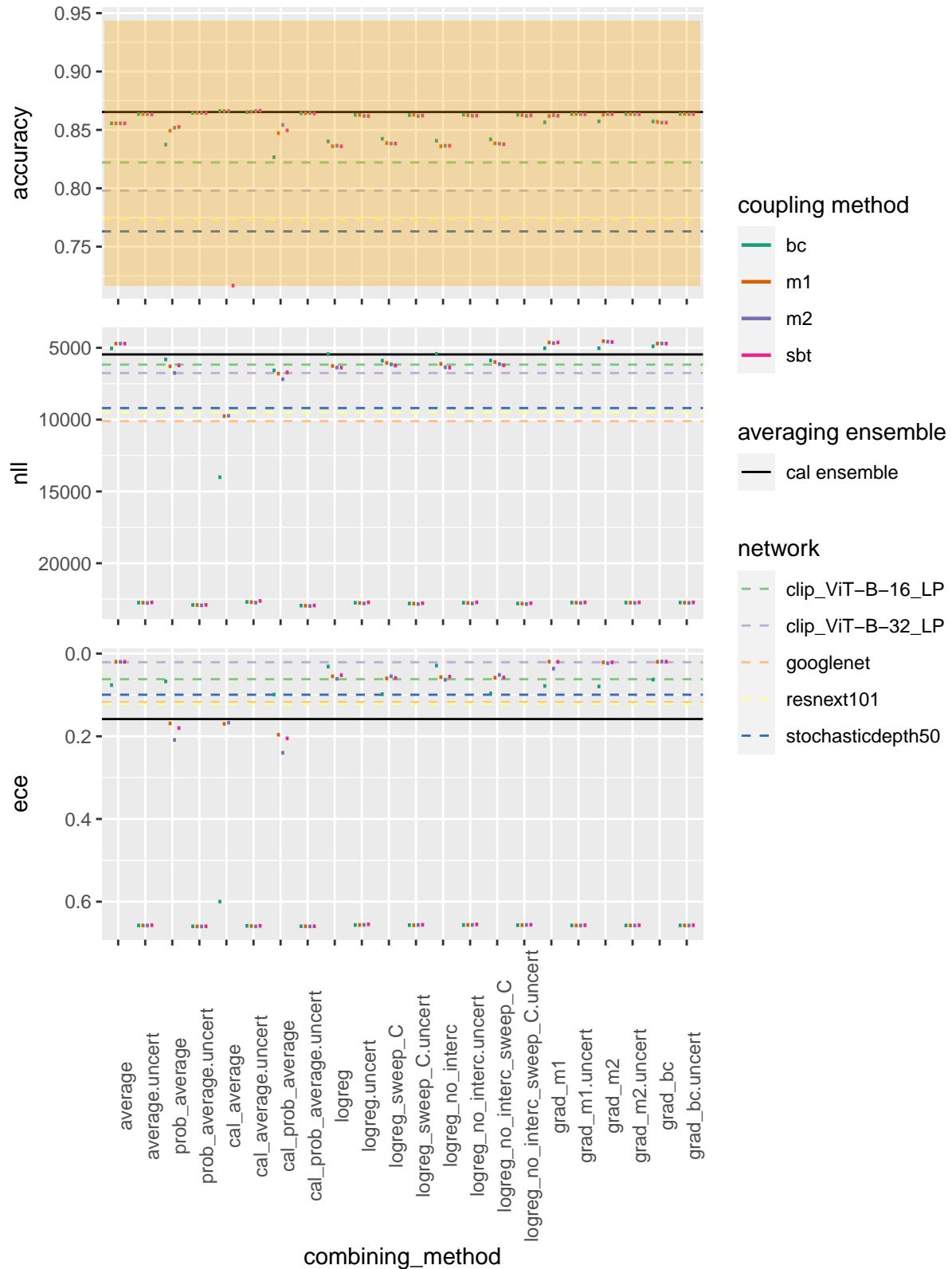
Ensemble metrics
Error inconsistency 0.339899986982346



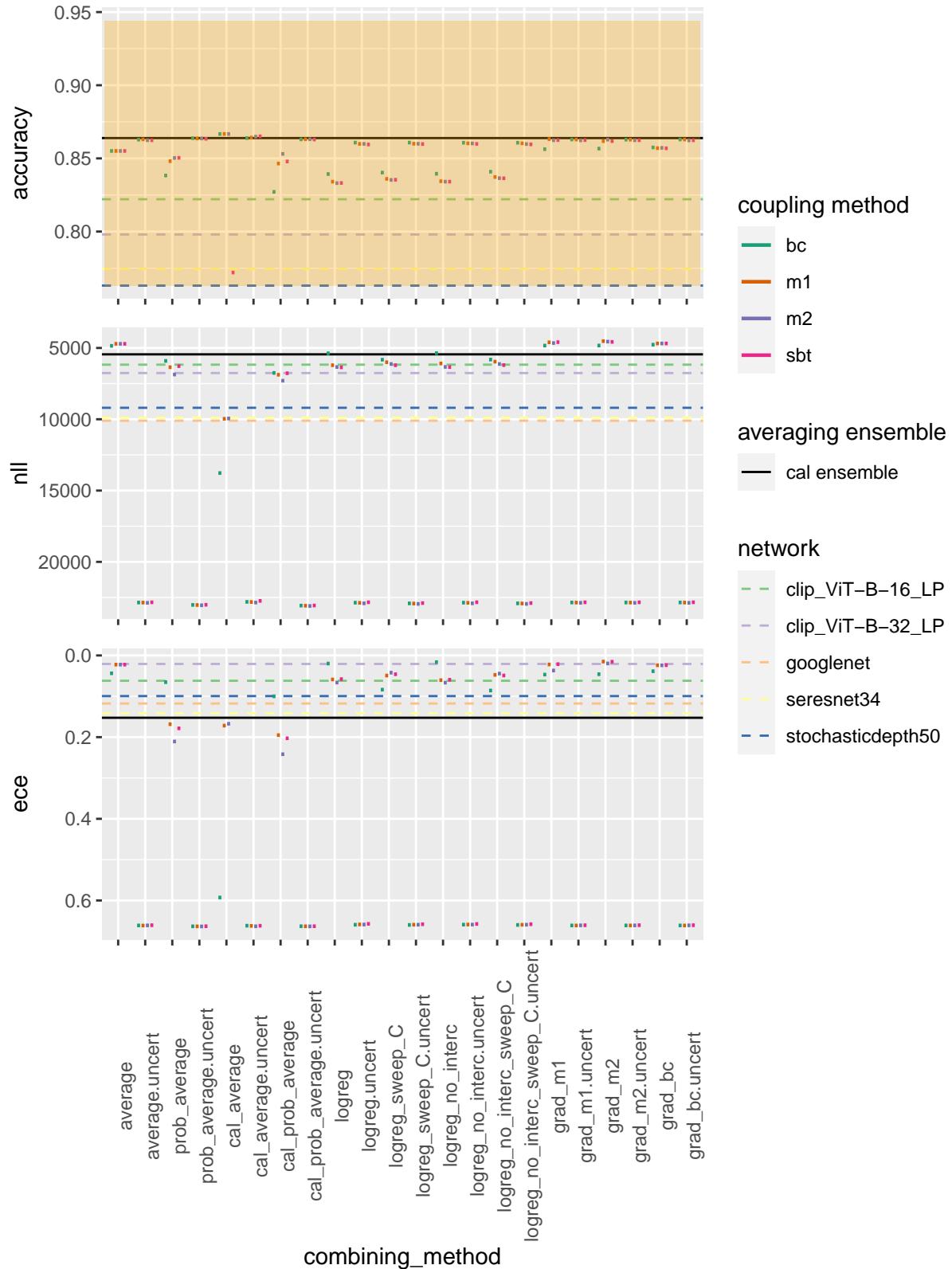
Ensemble metrics
Error inconsistency 0.353499978780746



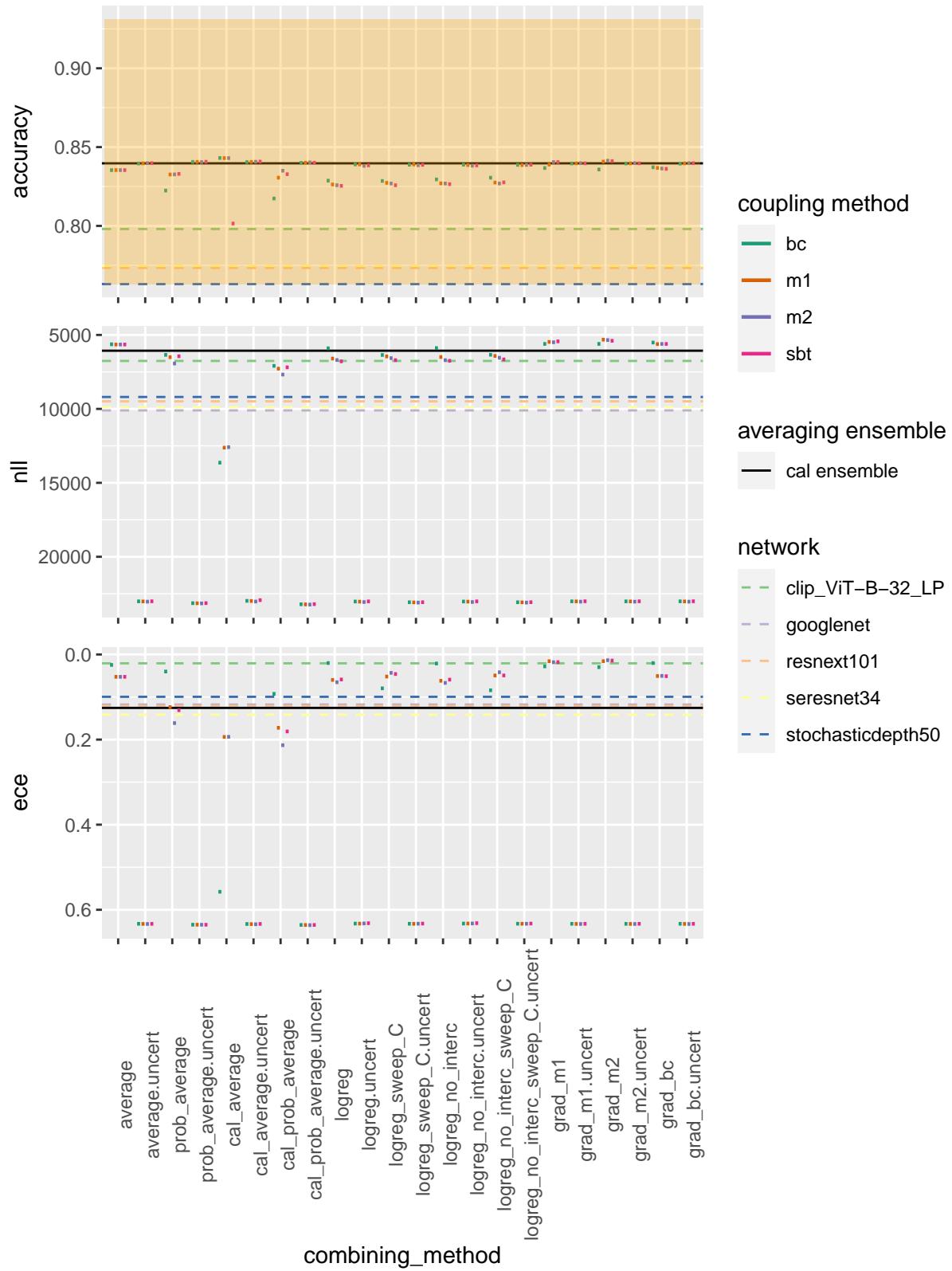
Ensemble metrics
Error inconsistency 0.373499989509583



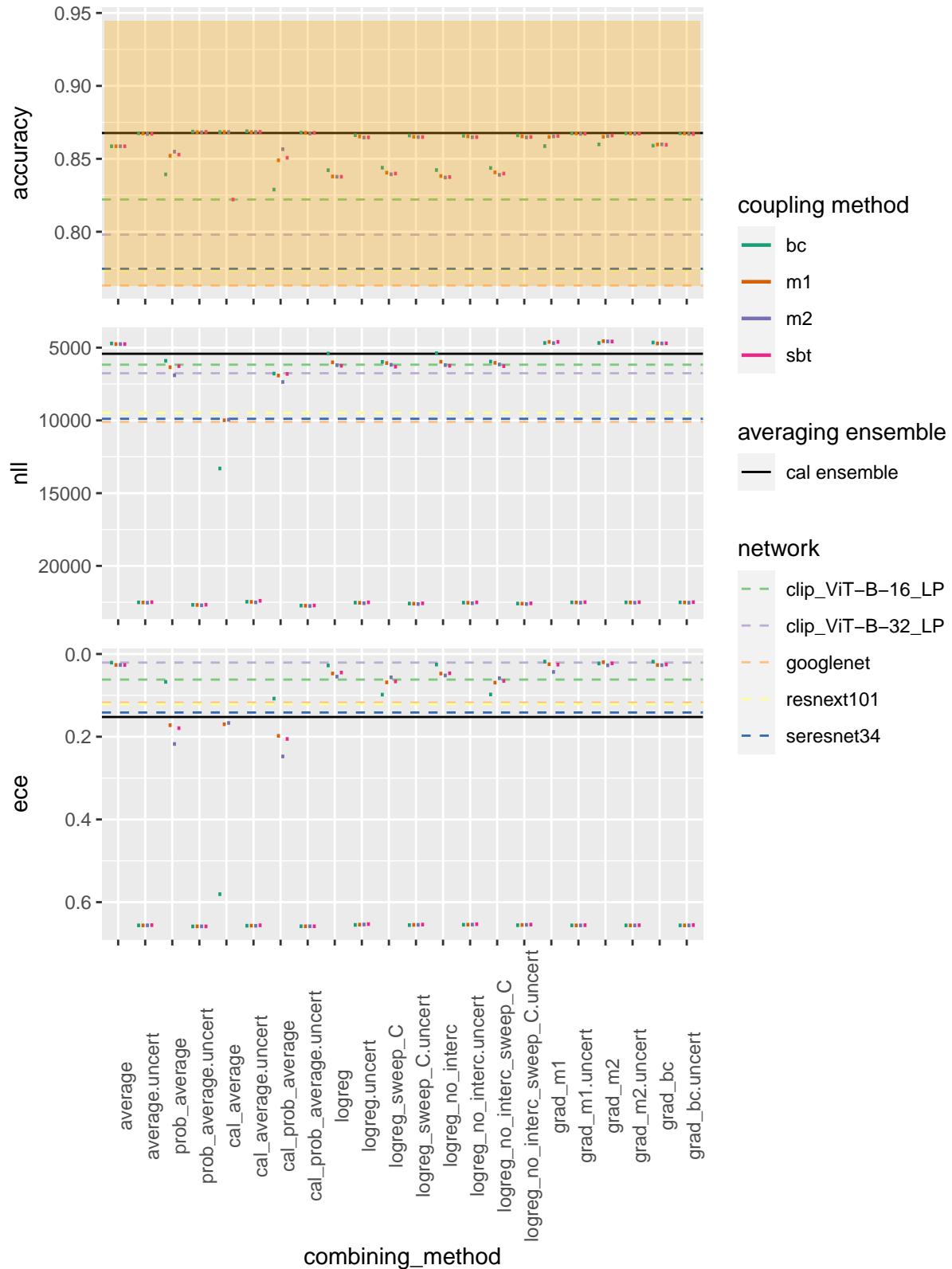
Ensemble metrics
Error inconsistency 0.373400002717972



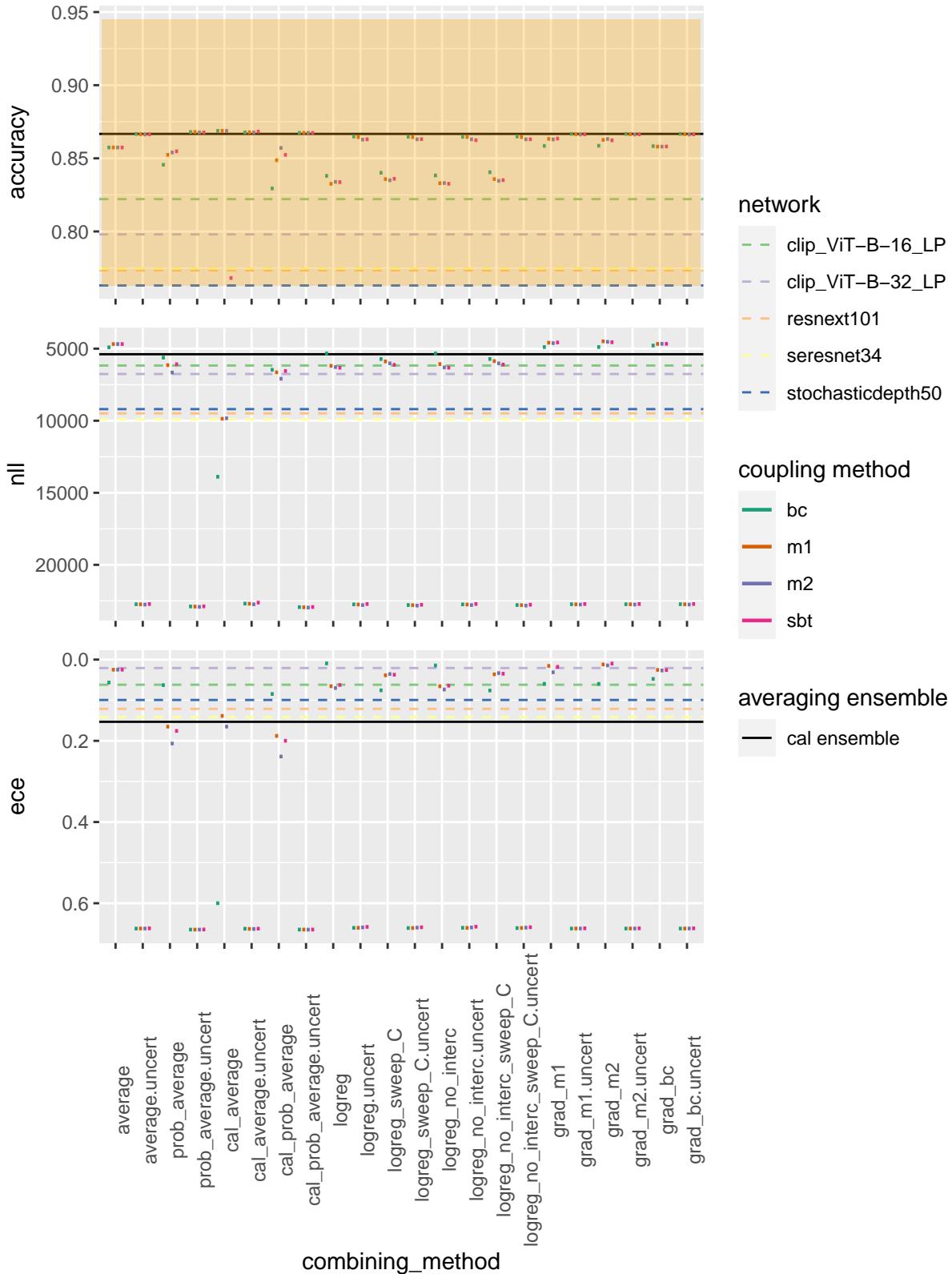
Ensemble metrics
Error inconsistency 0.351300001144409



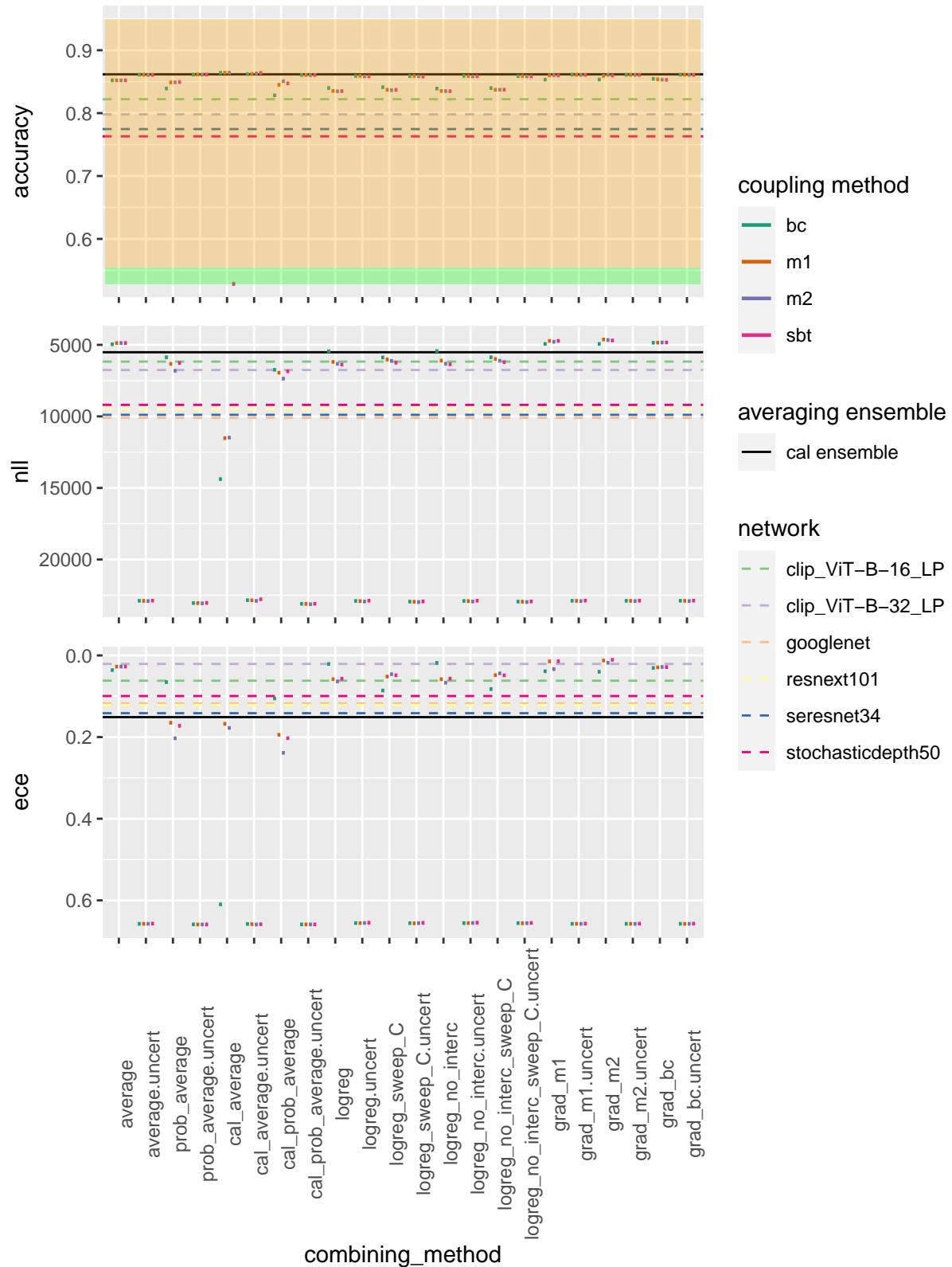
Ensemble metrics
Error inconsistency 0.369699984788895



Ensemble metrics
Error inconsistency 0.368099987506866



Ensemble metrics
Error inconsistency 0.394899994134903

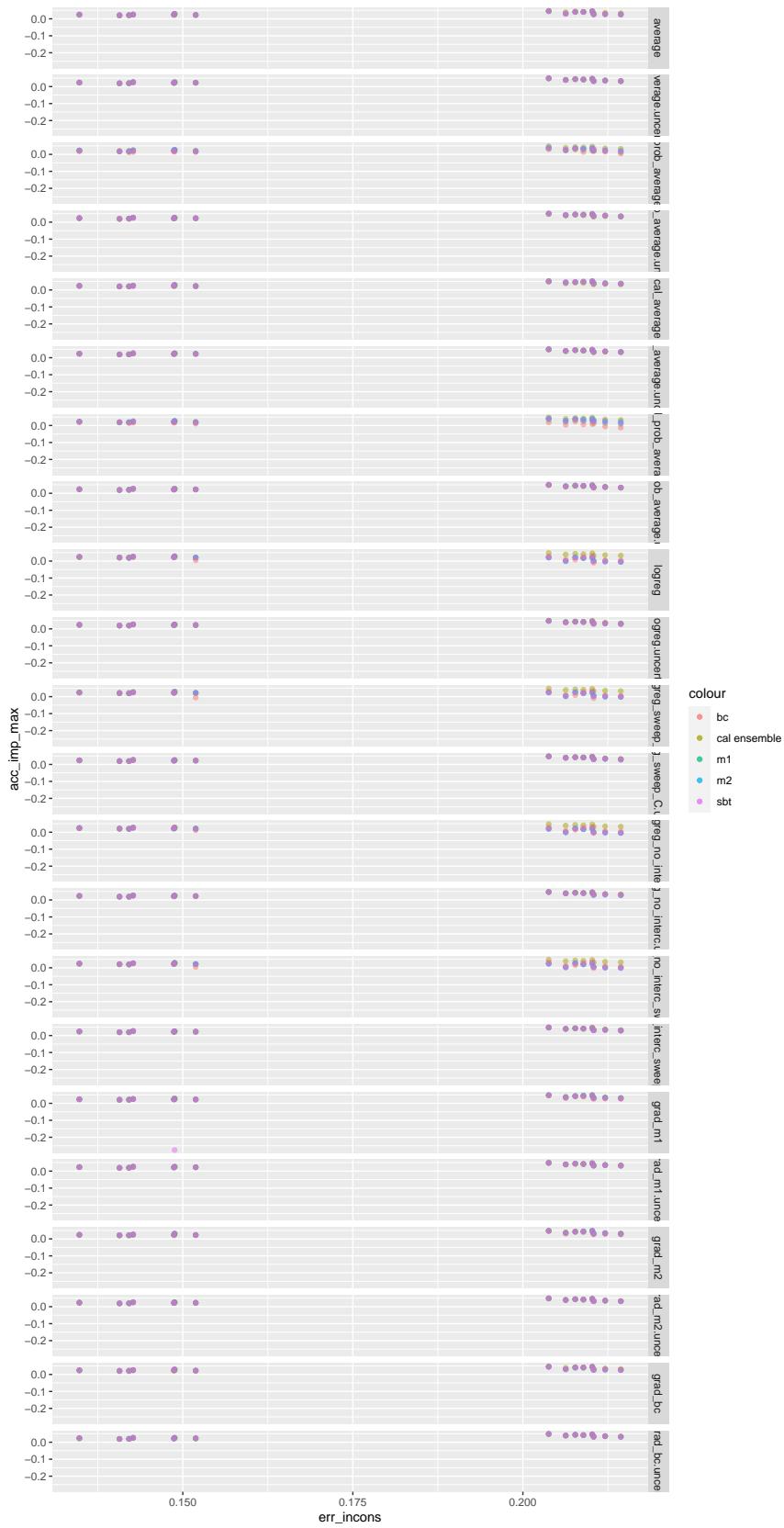


```

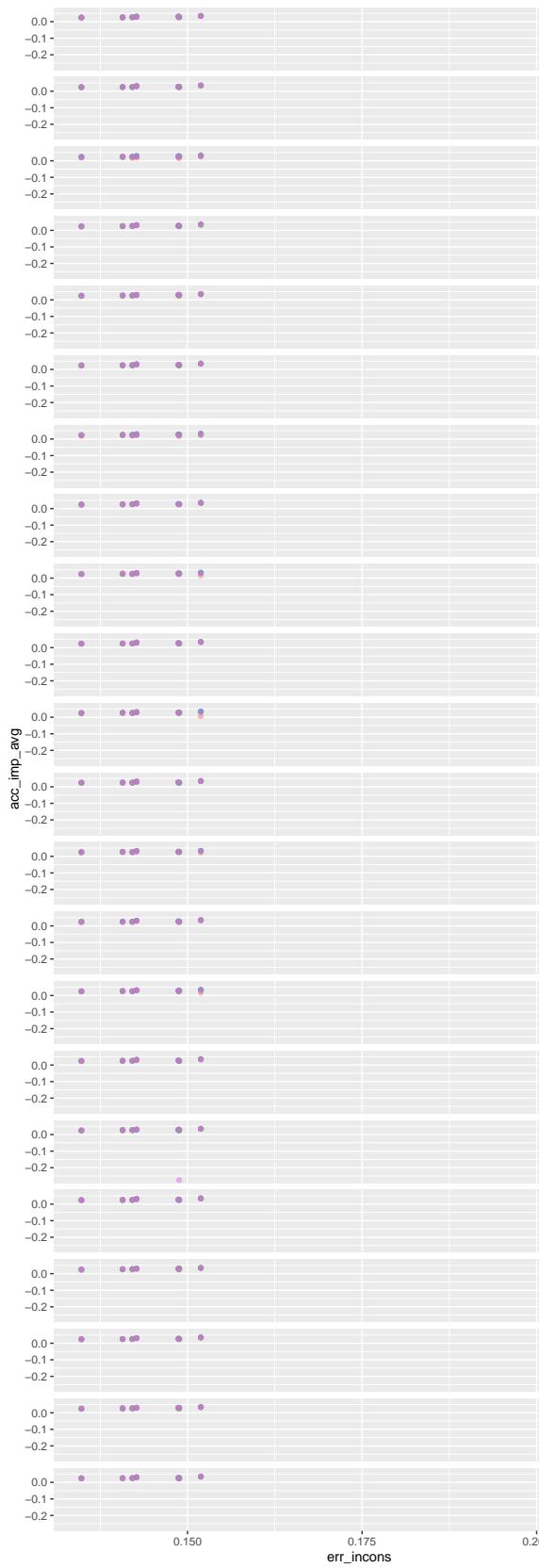
xax <- c(
  "err_incons",
  "nll_var", "nll_min", "nll_max", "nll_avg",
  "acc_var", "acc_min", "acc_max", "acc_avg",
  "ece_var", "ece_min", "ece_max", "ece_avg"
)
yax <- c("acc_imp_max", "acc_imp_avg")
for (sss in unique(ens_cal_plt_df$combination_size))
{
  cur_ens_cal_plt_df <- ens_cal_plt_df %>%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

```

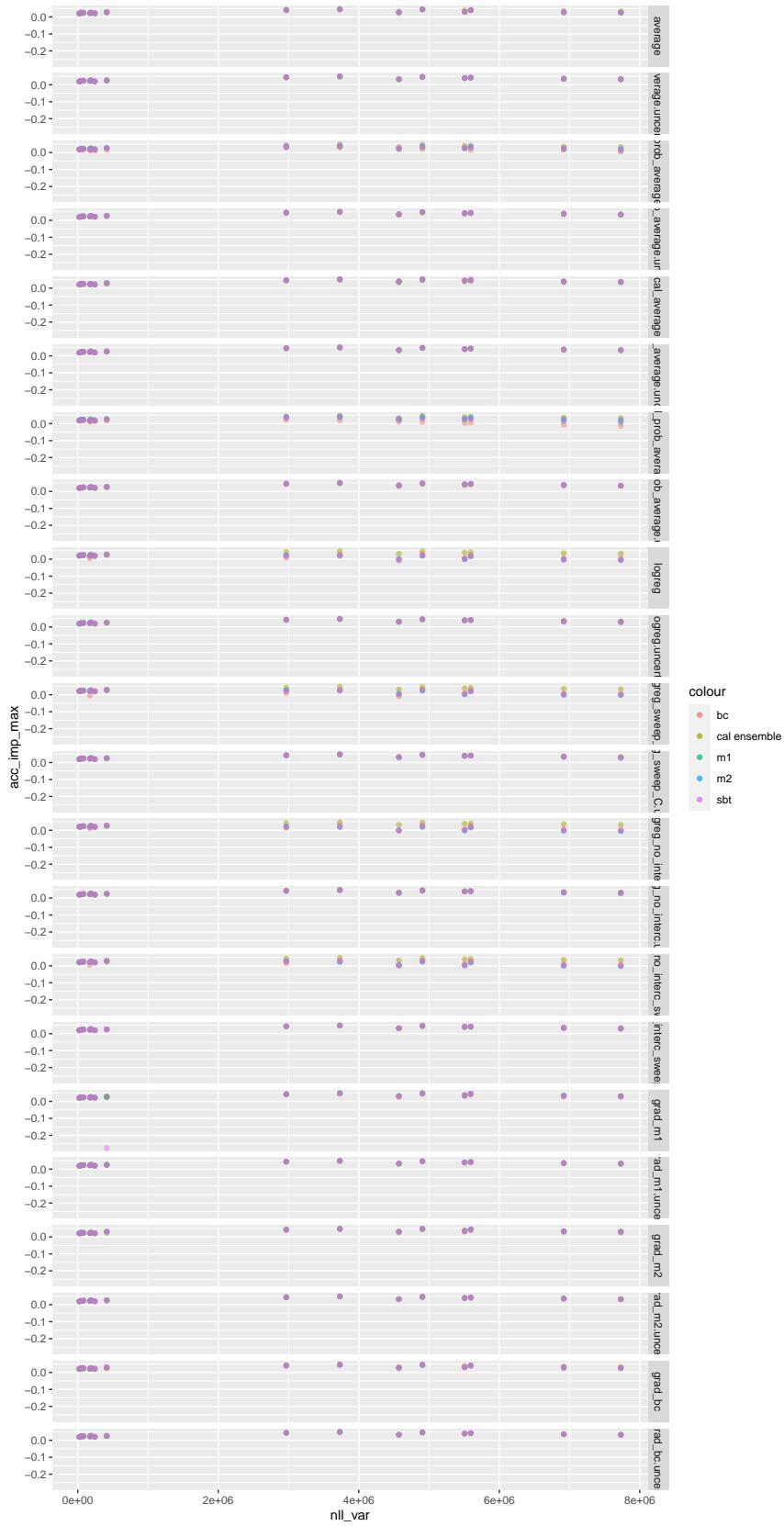
Accuracy improvement of ensemble over
the best of networks vs err_incons.
Ensemble size 2



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



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



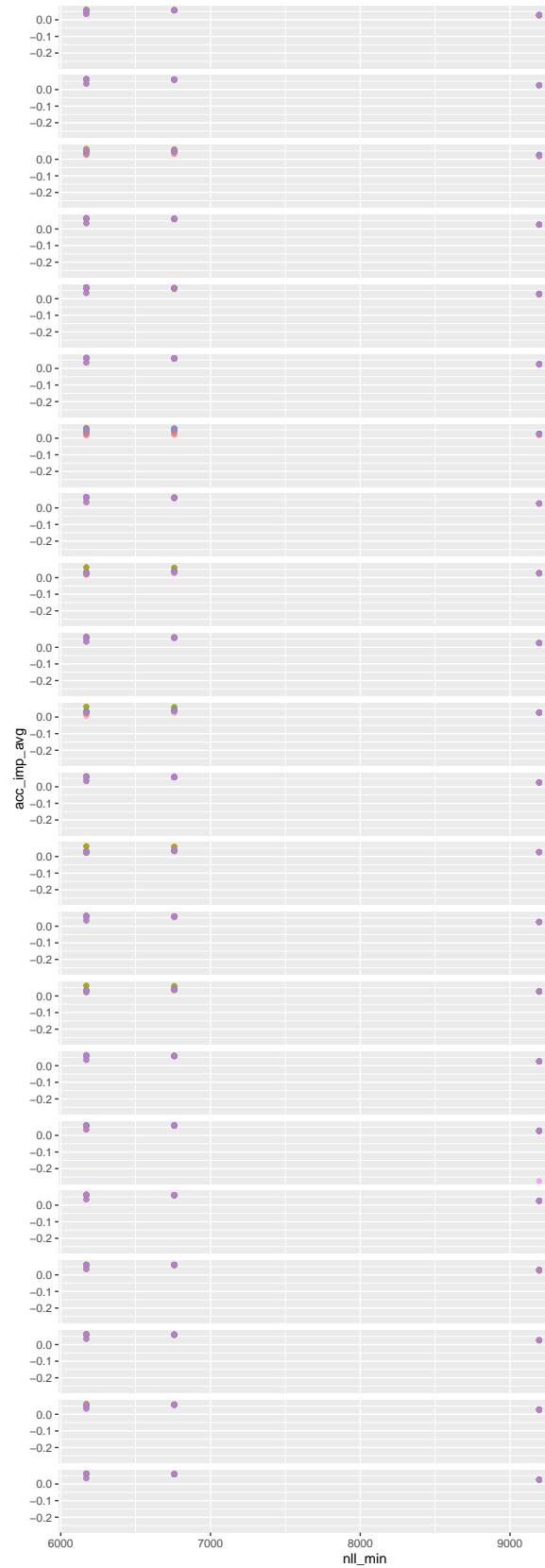
Accuracy improvement of ensemble over
the average of networks vs nll_var.
Ensemble size 2



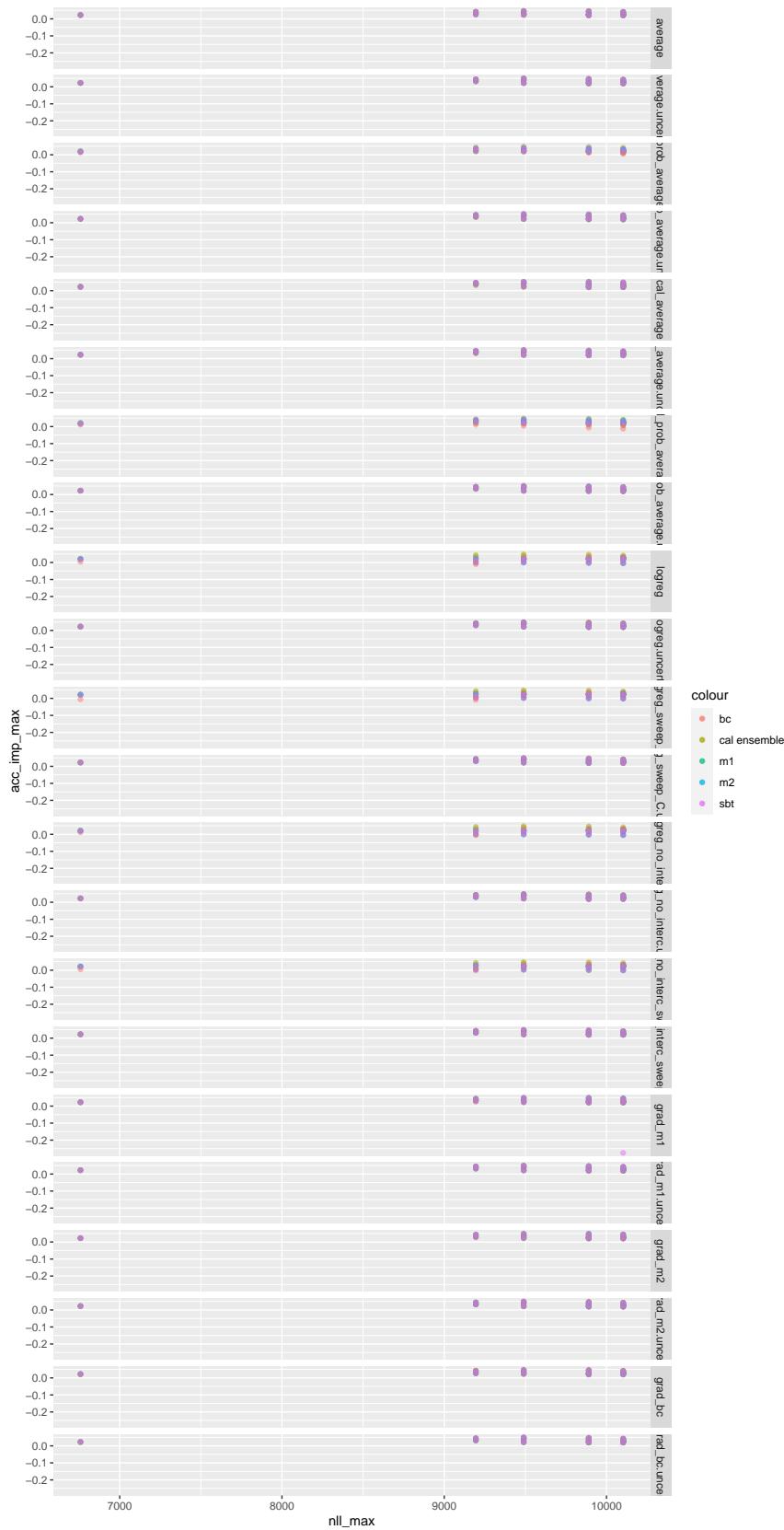
Accuracy improvement of ensemble over
the best of networks vs nll_min.
Ensemble size 2



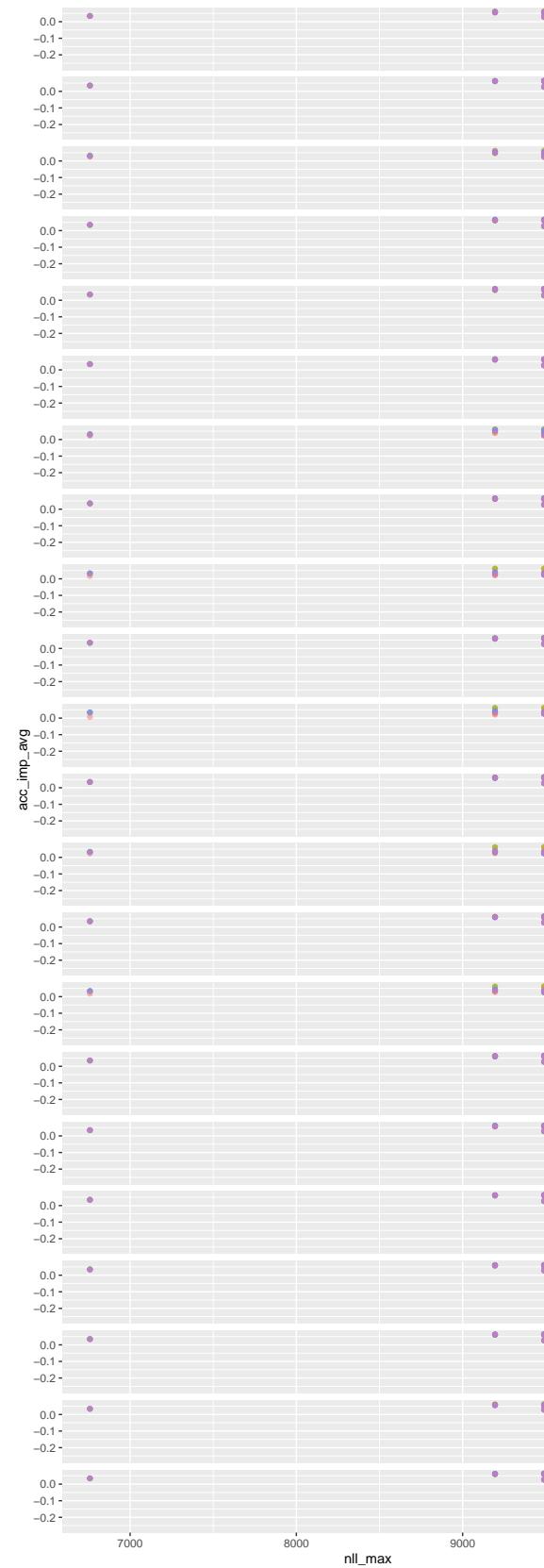
Accuracy improvement of ensemble over
the average of networks vs nll_min.
Ensemble size 2



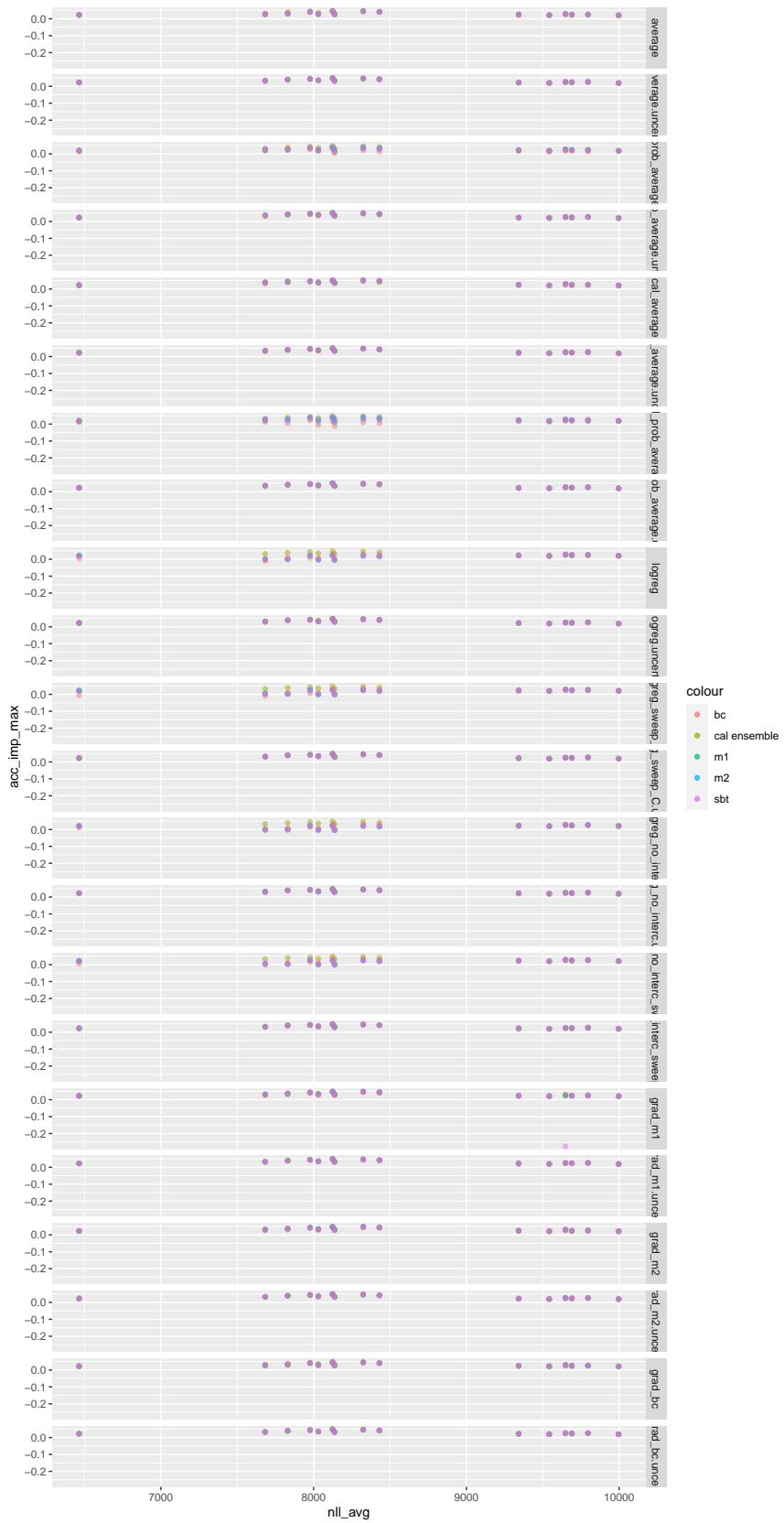
Accuracy improvement of ensemble over
the best of networks vs nll_max.
Ensemble size 2



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



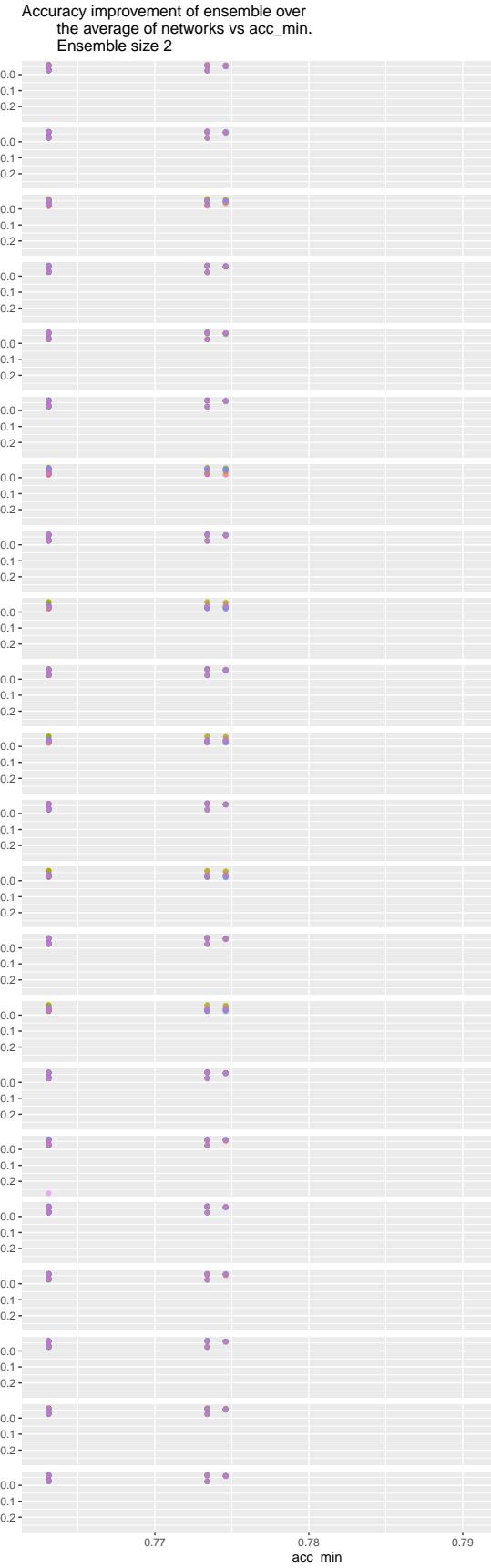
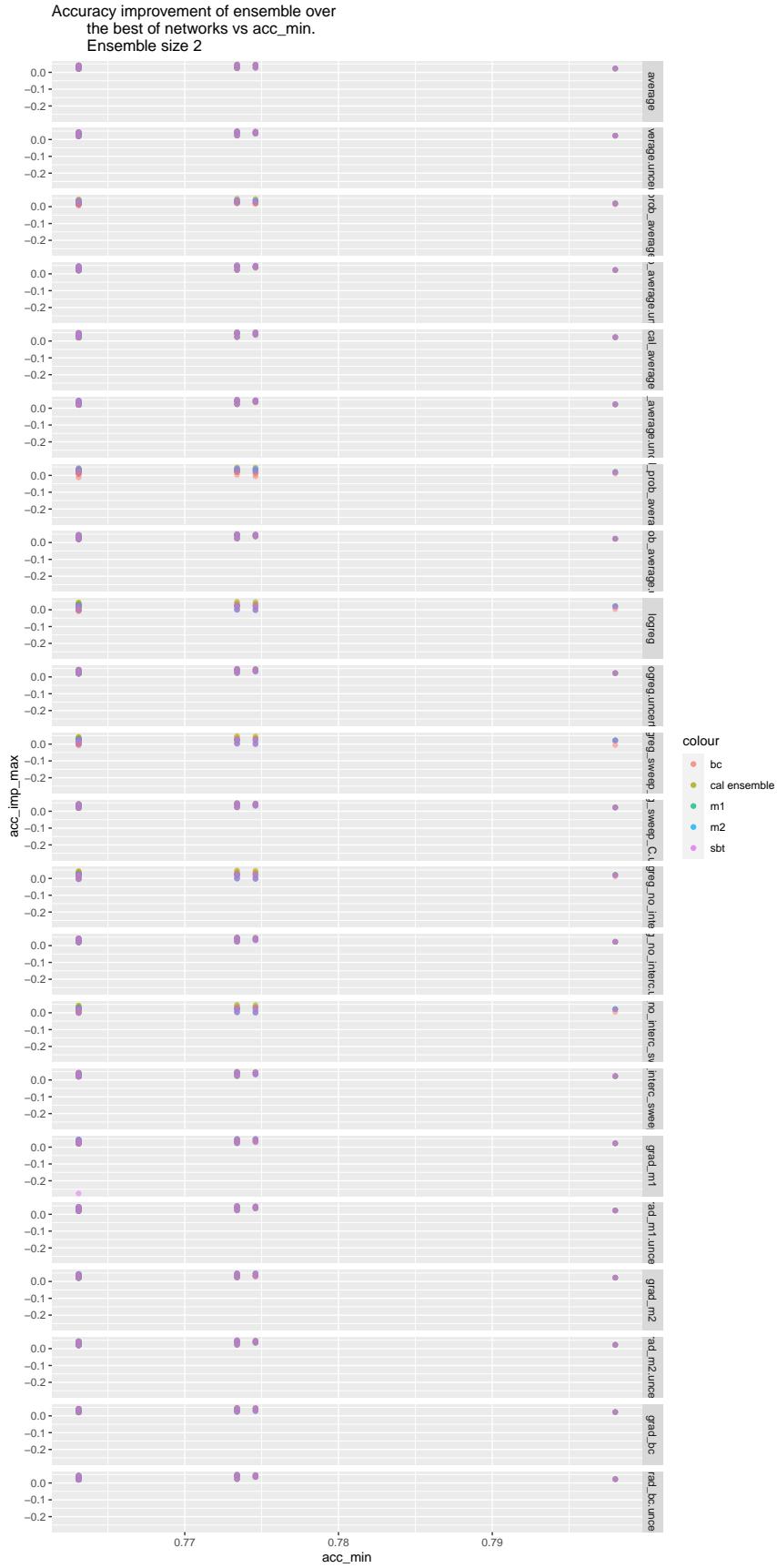
Accuracy improvement of ensemble over
the best of networks vs nll_avg.
Ensemble size 2

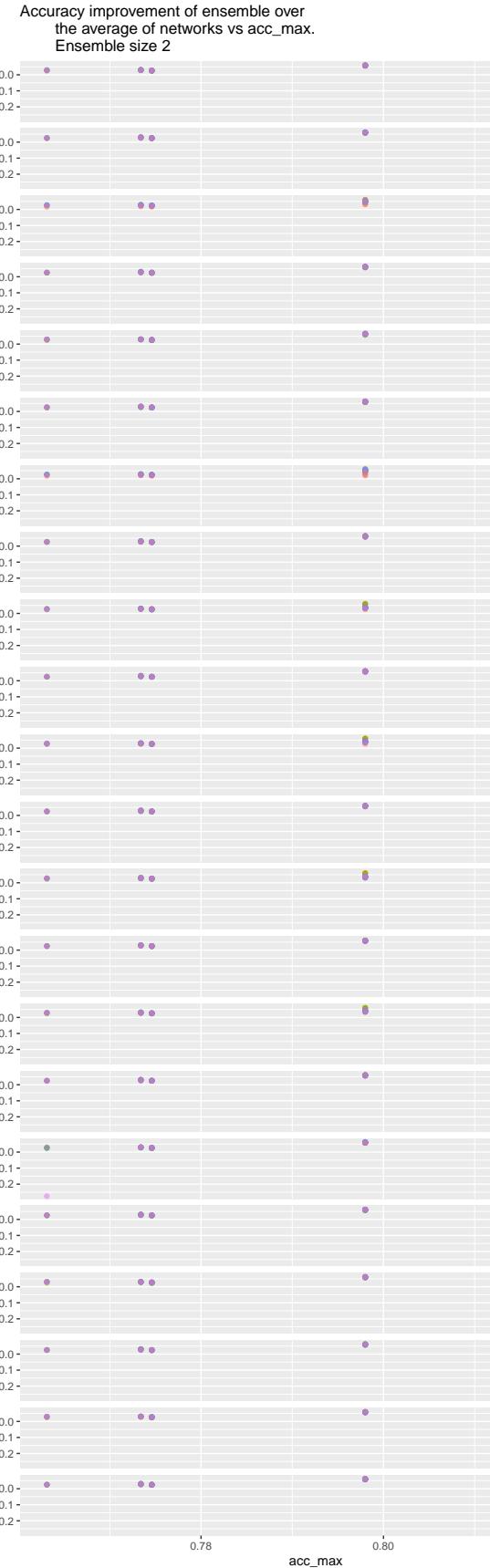


Accuracy improvement of ensemble over
the average of networks vs nll_avg.
Ensemble size 2









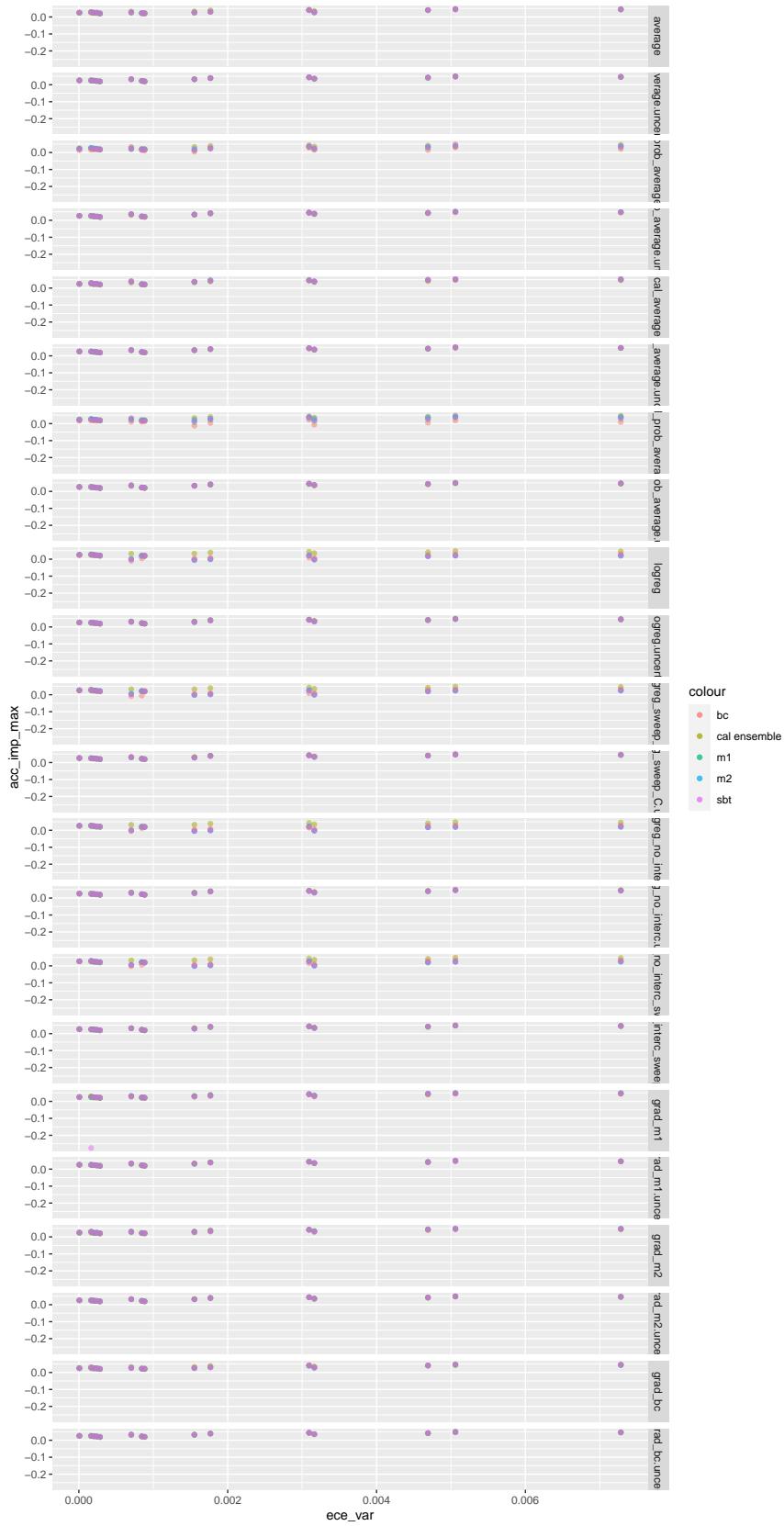
Accuracy improvement of ensemble over
the best of networks vs acc_avg.
Ensemble size 2



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



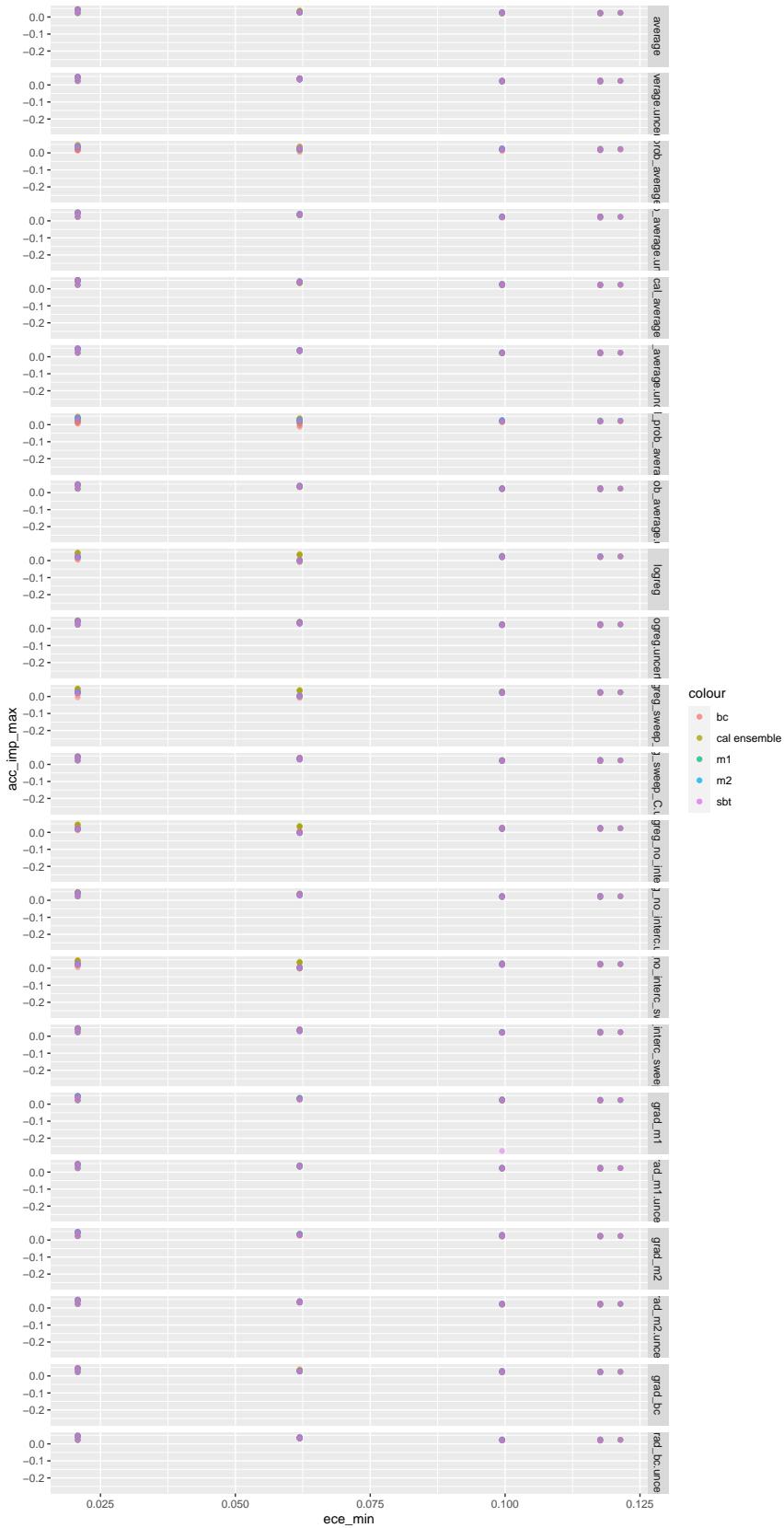
Accuracy improvement of ensemble over
the best of networks vs ece_var.
Ensemble size 2



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

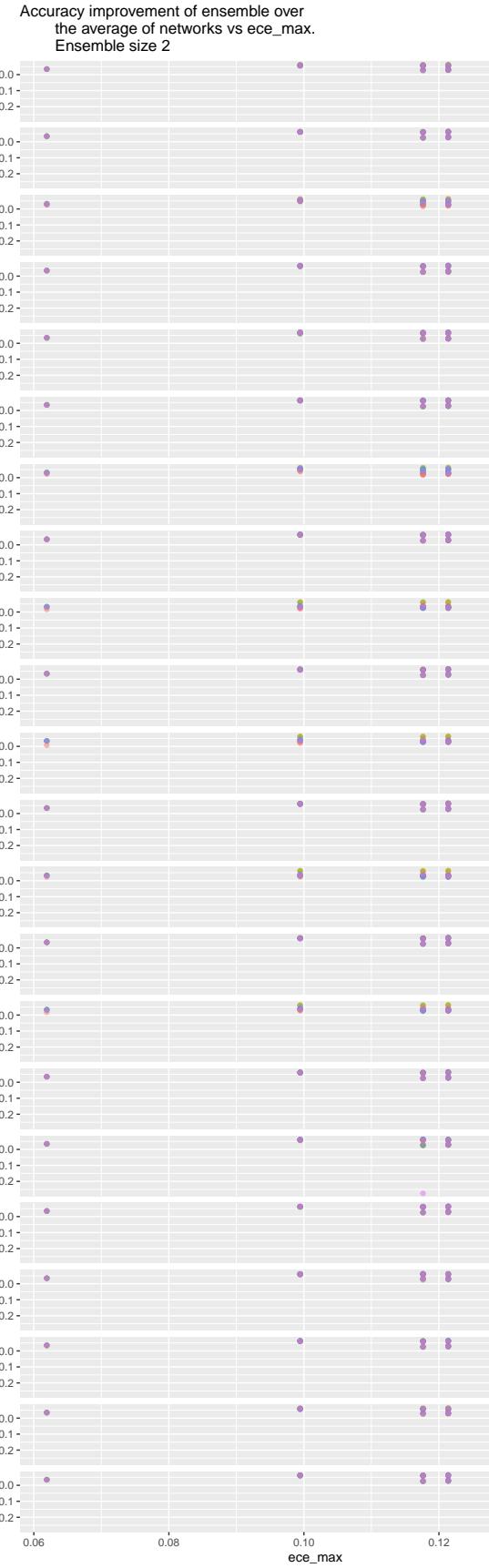


Accuracy improvement of ensemble over
the best of networks vs ece_min.
Ensemble size 2



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





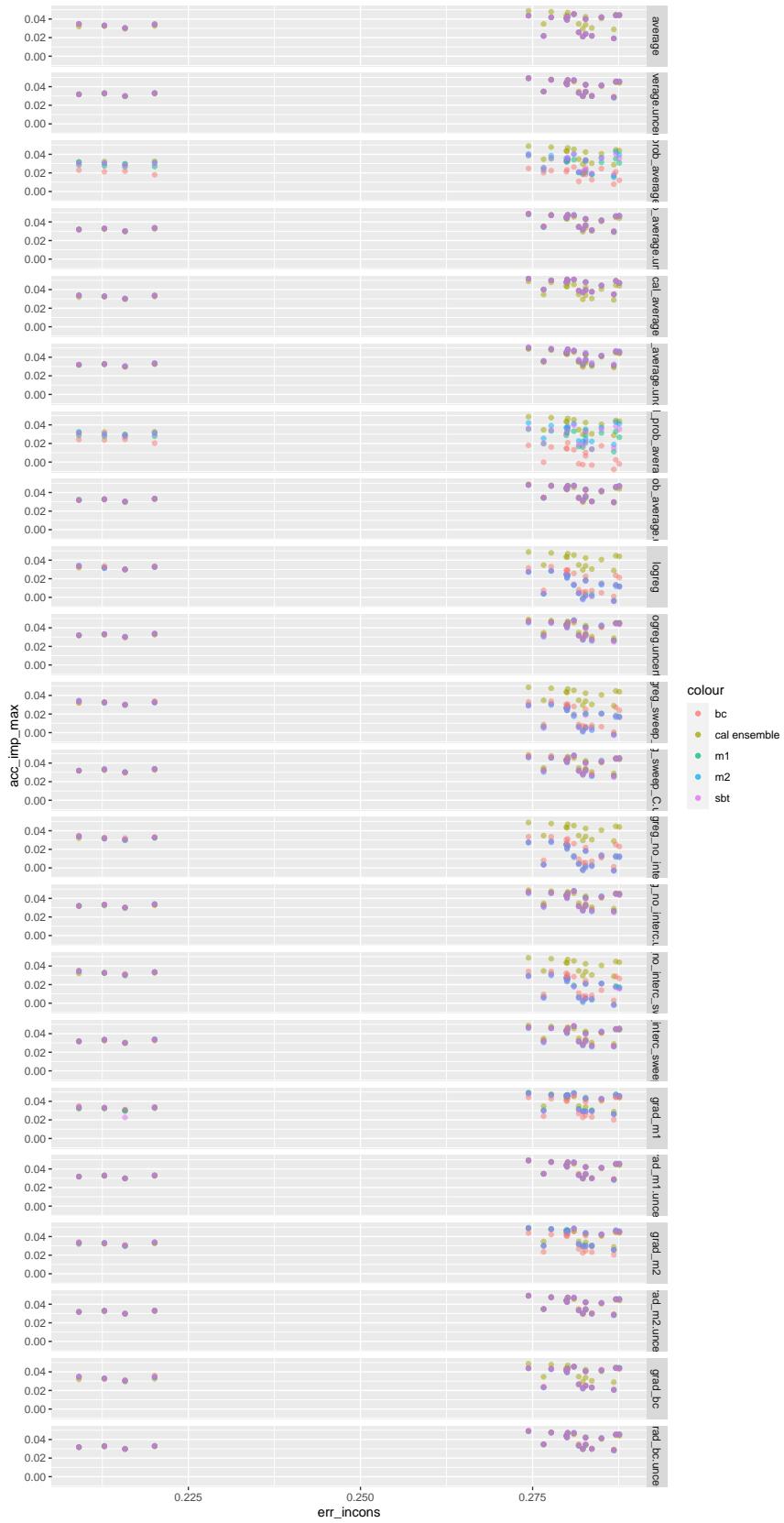
Accuracy improvement of ensemble over
the best of networks vs ece_avg.
Ensemble size 2



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



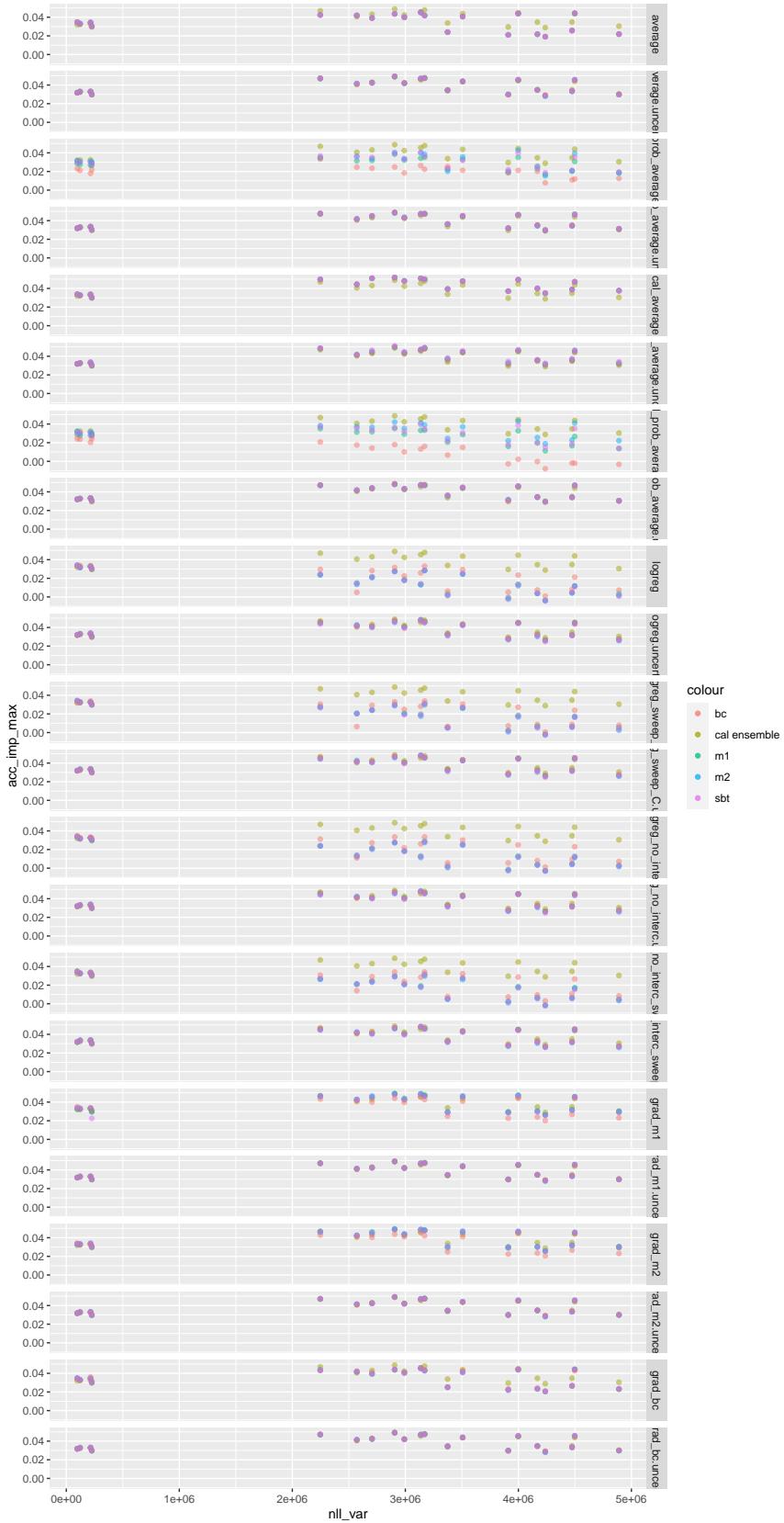
Accuracy improvement of ensemble over
the best of networks vs err_incons.
Ensemble size 3



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



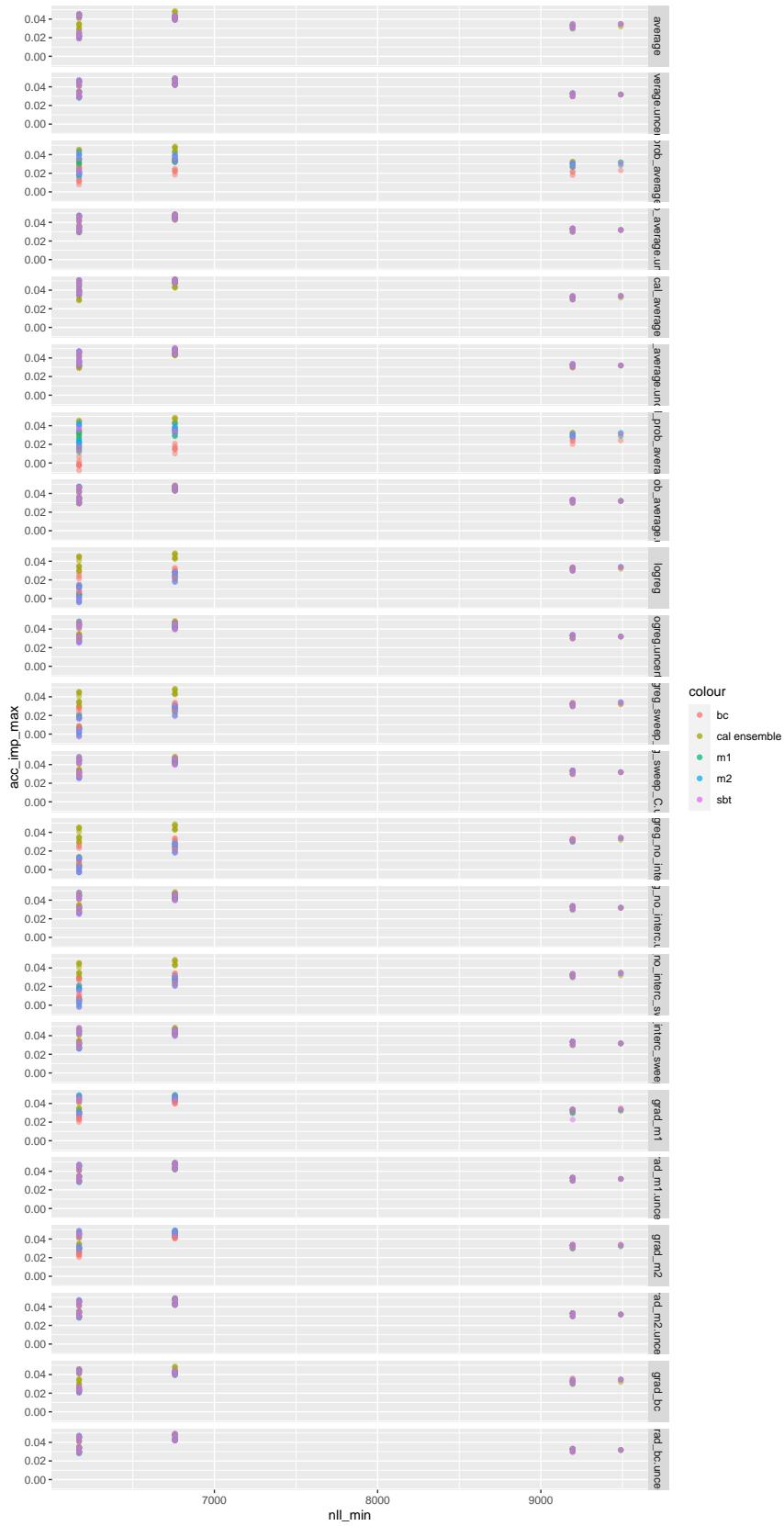
Accuracy improvement of ensemble over
the best of networks vs nll_var.
Ensemble size 3



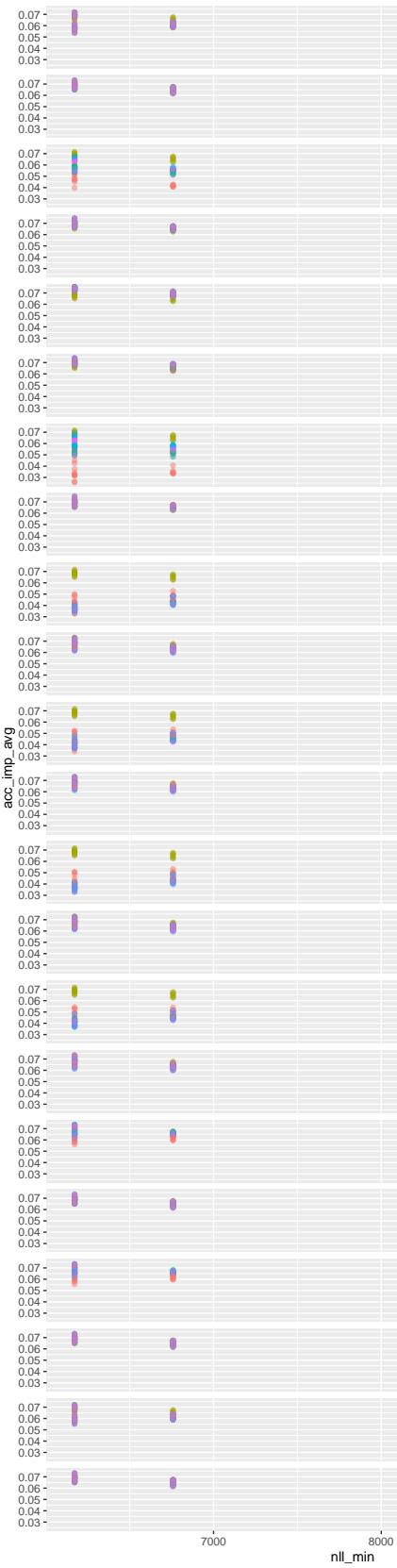
Accuracy improvement of ensemble over
the average of networks vs nll_var.
Ensemble size 3



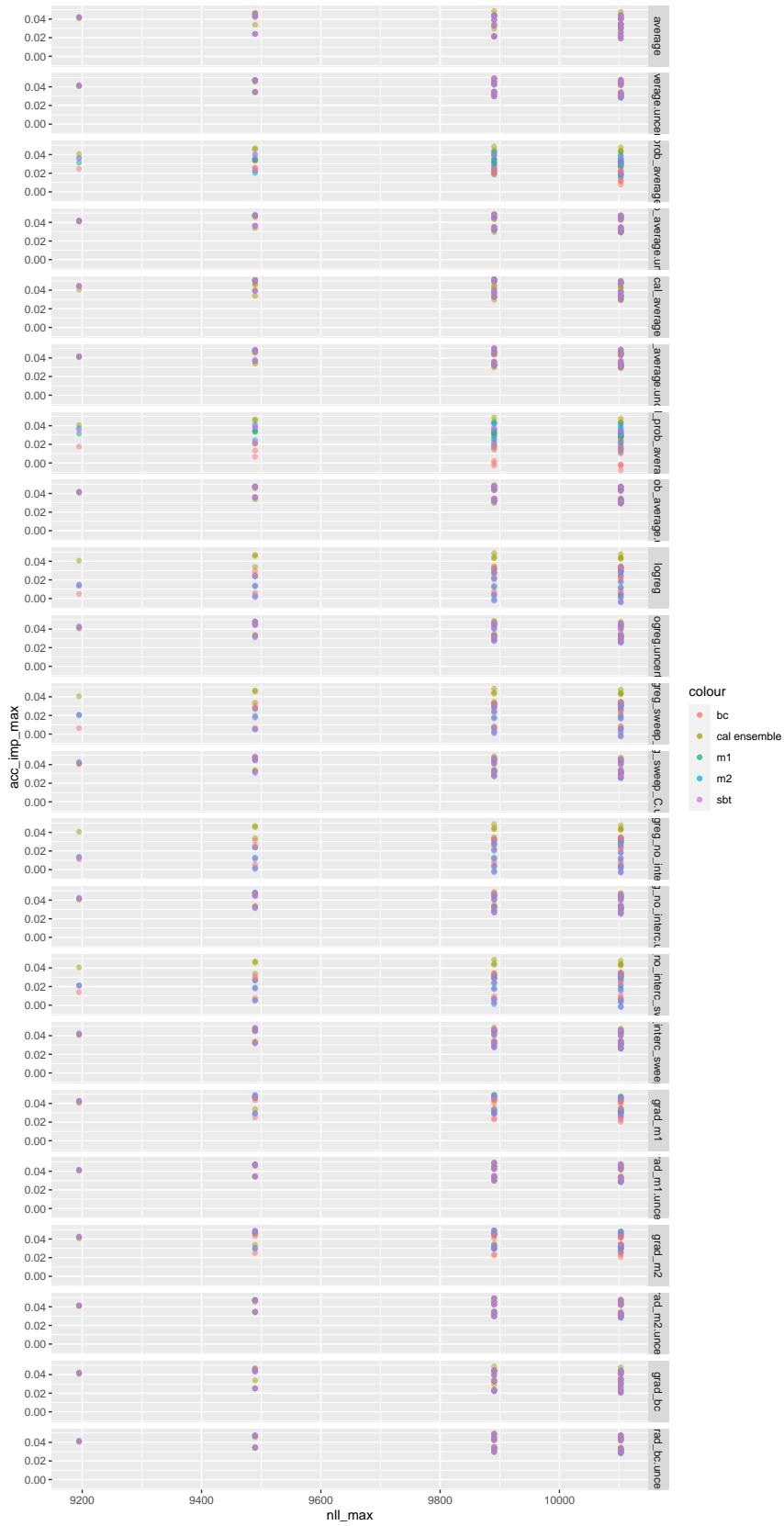
Accuracy improvement of ensemble over
the best of networks vs nll_min.
Ensemble size 3



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



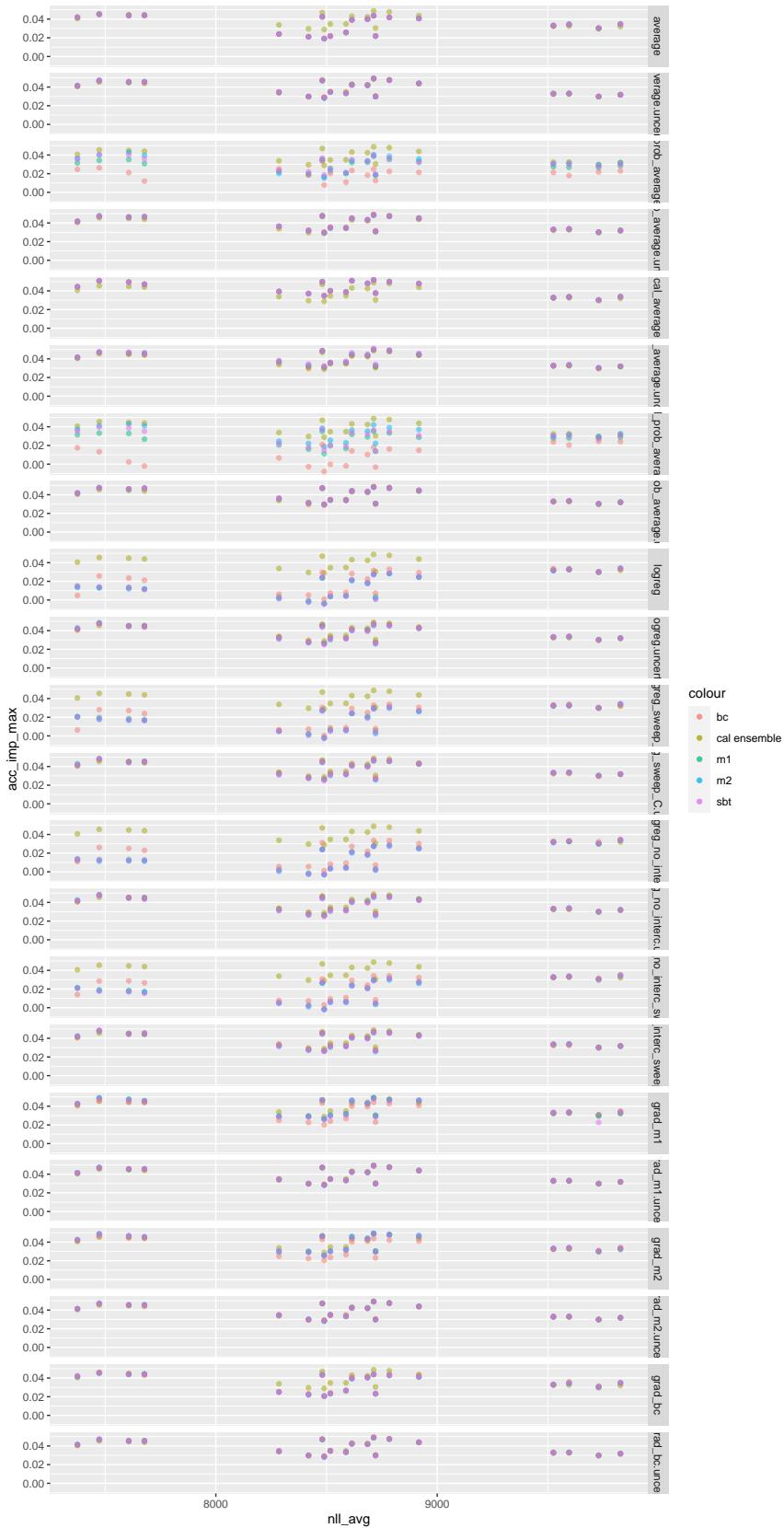
Accuracy improvement of ensemble over
the best of networks vs nll_max.
Ensemble size 3



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



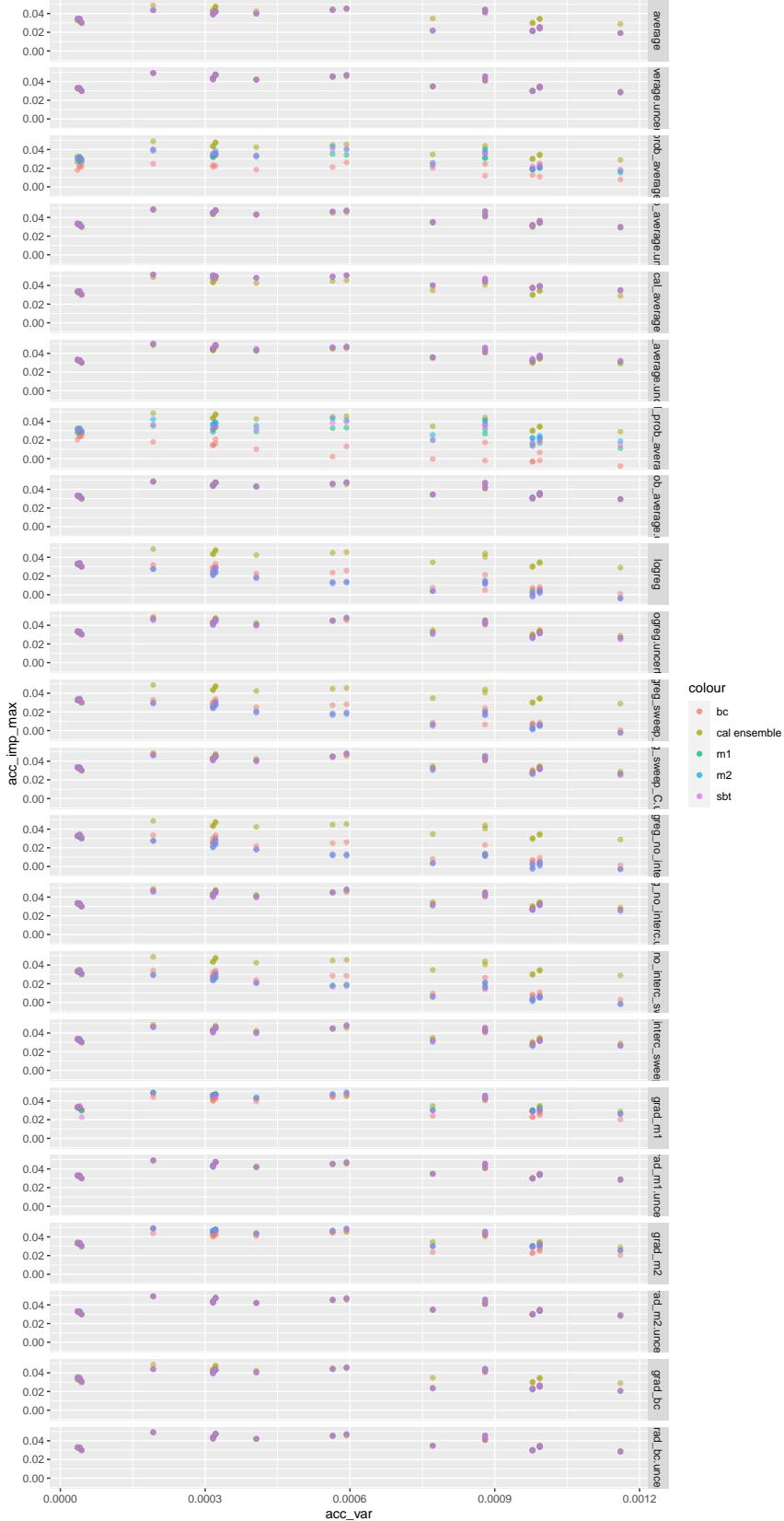
Accuracy improvement of ensemble over
the best of networks vs nll_avg.
Ensemble size 3



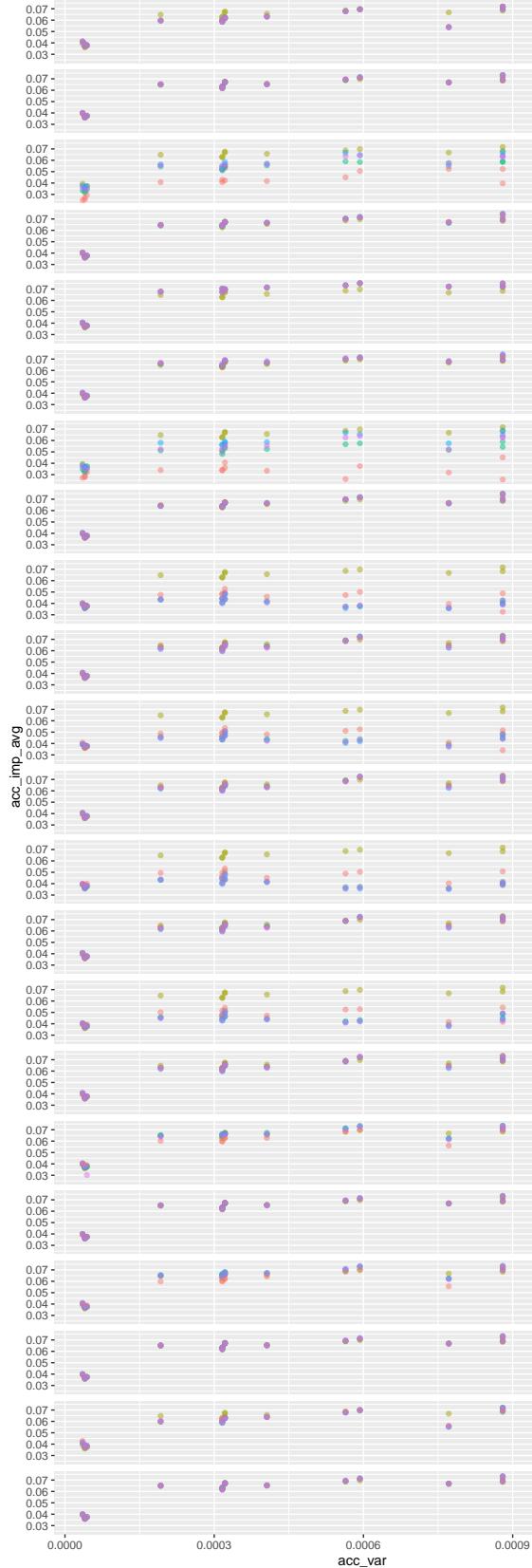
Accuracy improvement of ensemble over
the average of networks vs nll_avg.
Ensemble size 3



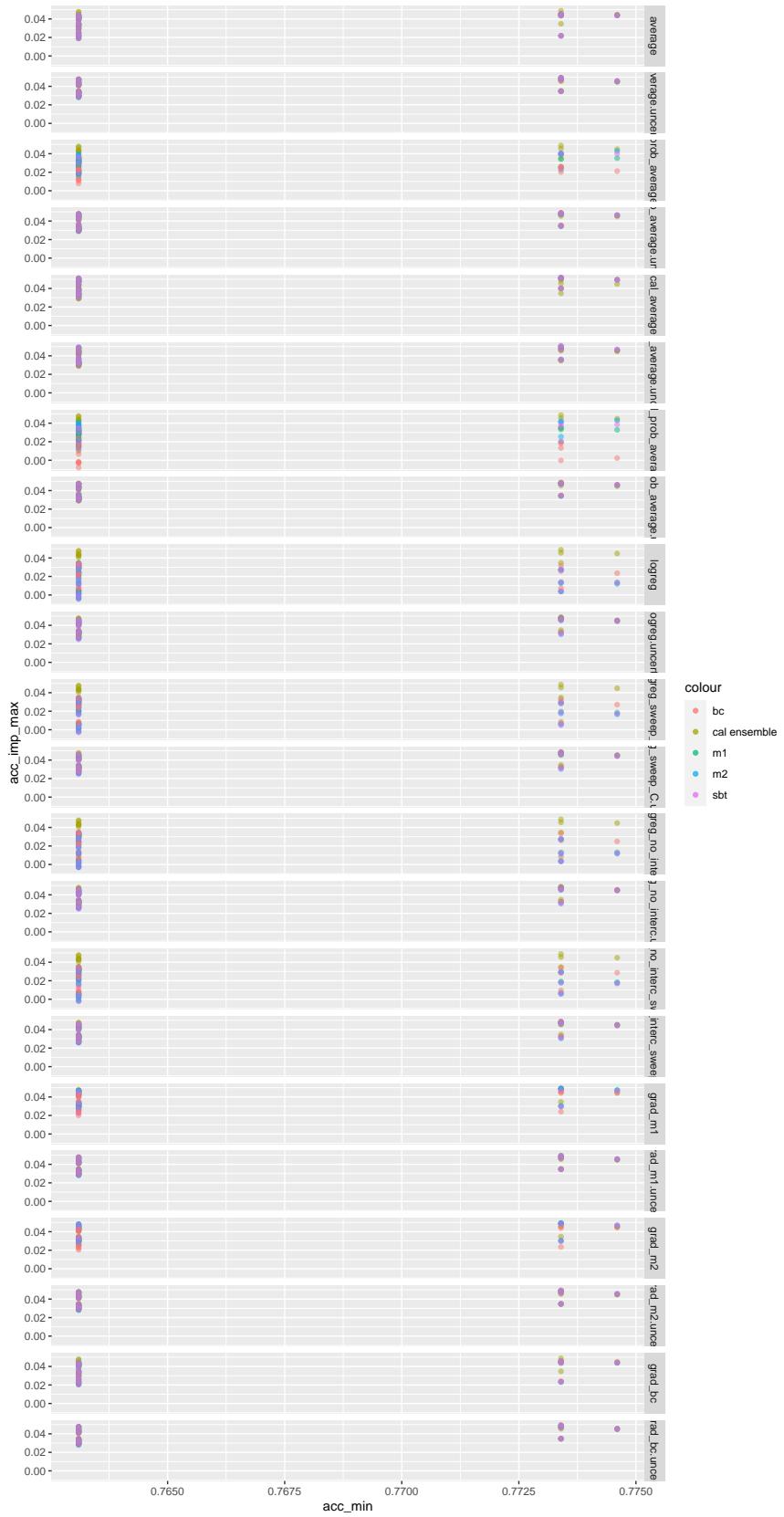
Accuracy improvement of ensemble over
the best of networks vs acc_var.
Ensemble size 3



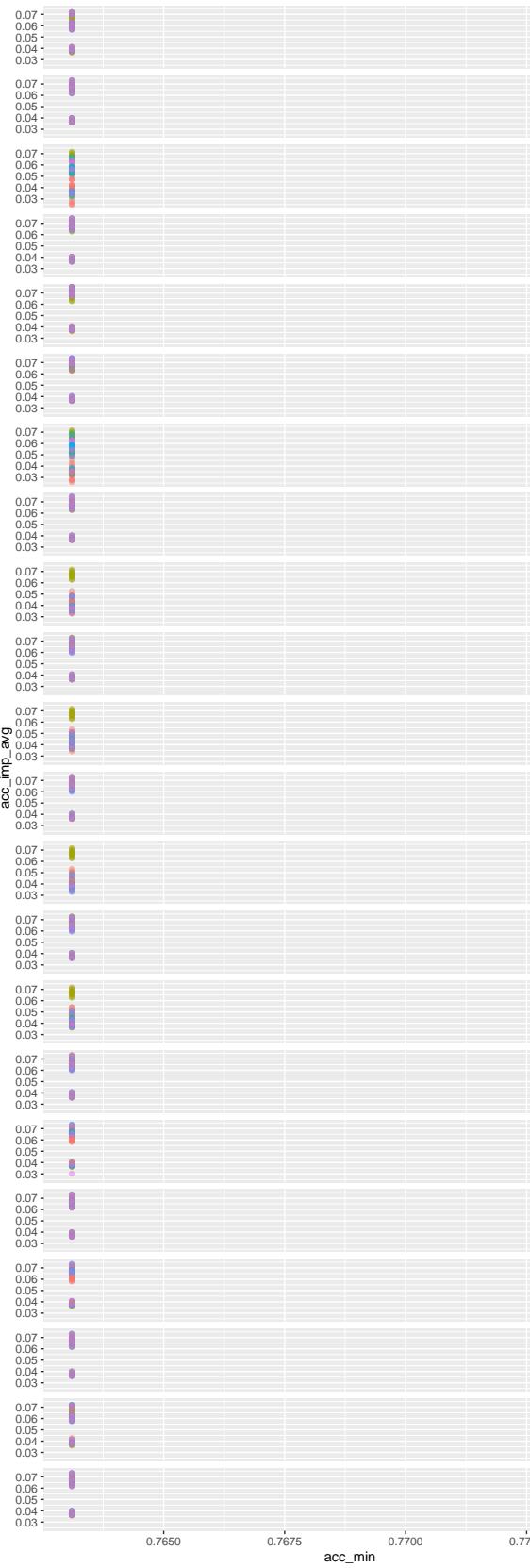
Accuracy improvement of ensemble over
the average of networks vs acc_var.
Ensemble size 3



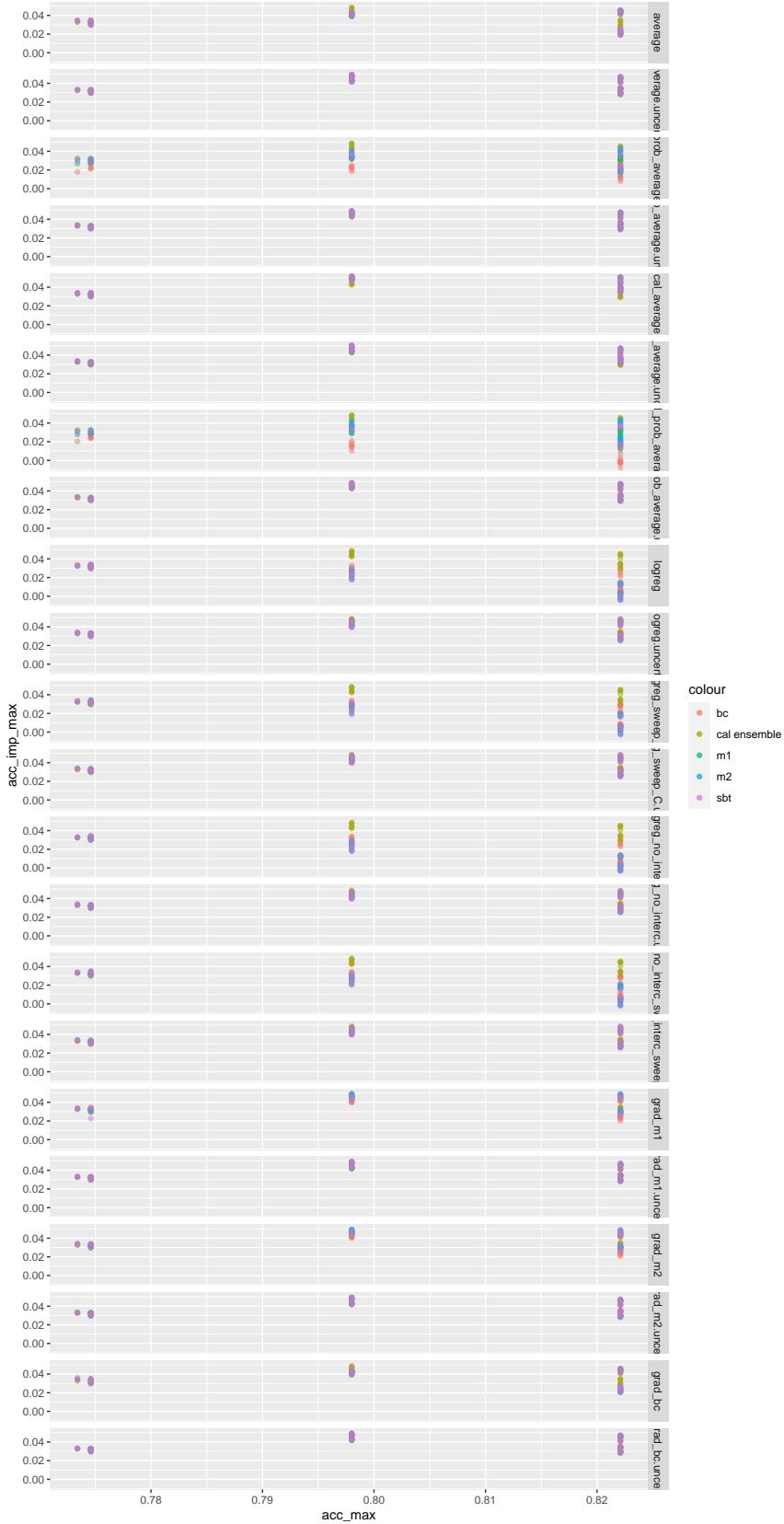
Accuracy improvement of ensemble over
the best of networks vs acc_min.
Ensemble size 3



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



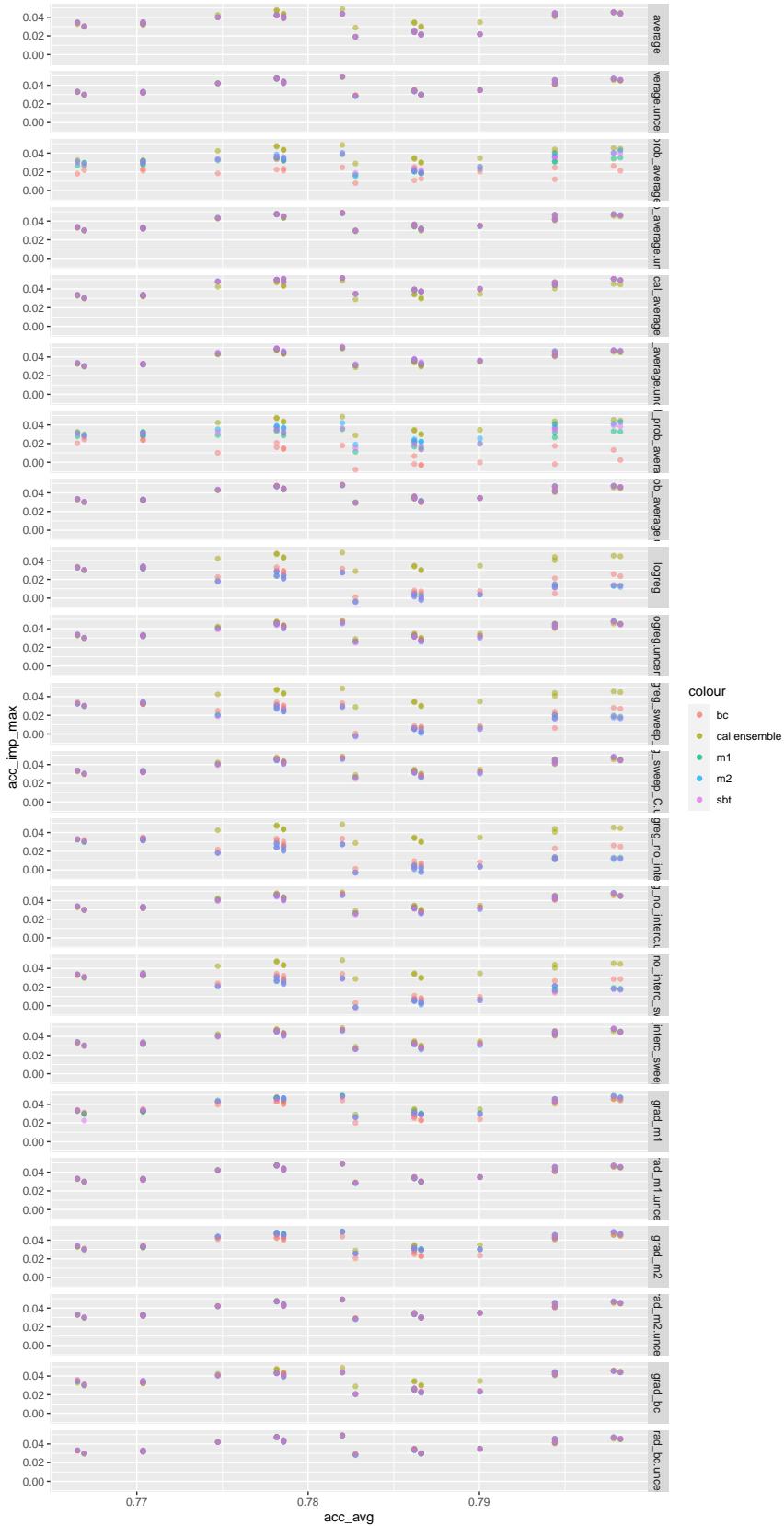
Accuracy improvement of ensemble over
the best of networks vs acc_max.
Ensemble size 3



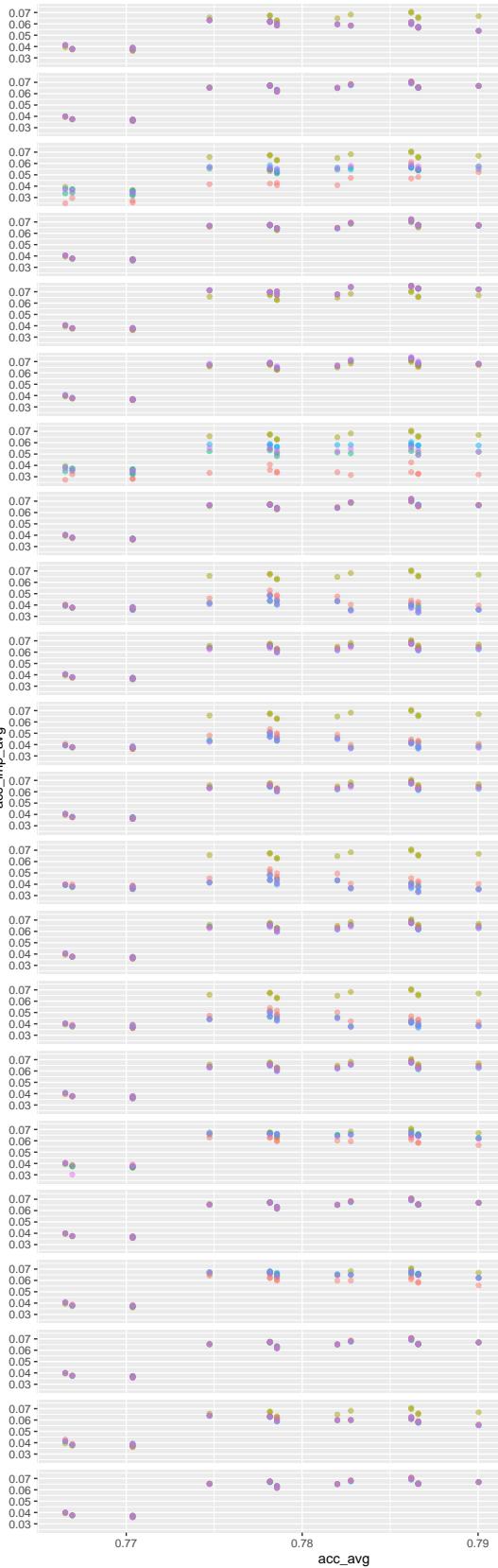
Accuracy improvement of ensemble over
the average of networks vs acc_max.
Ensemble size 3



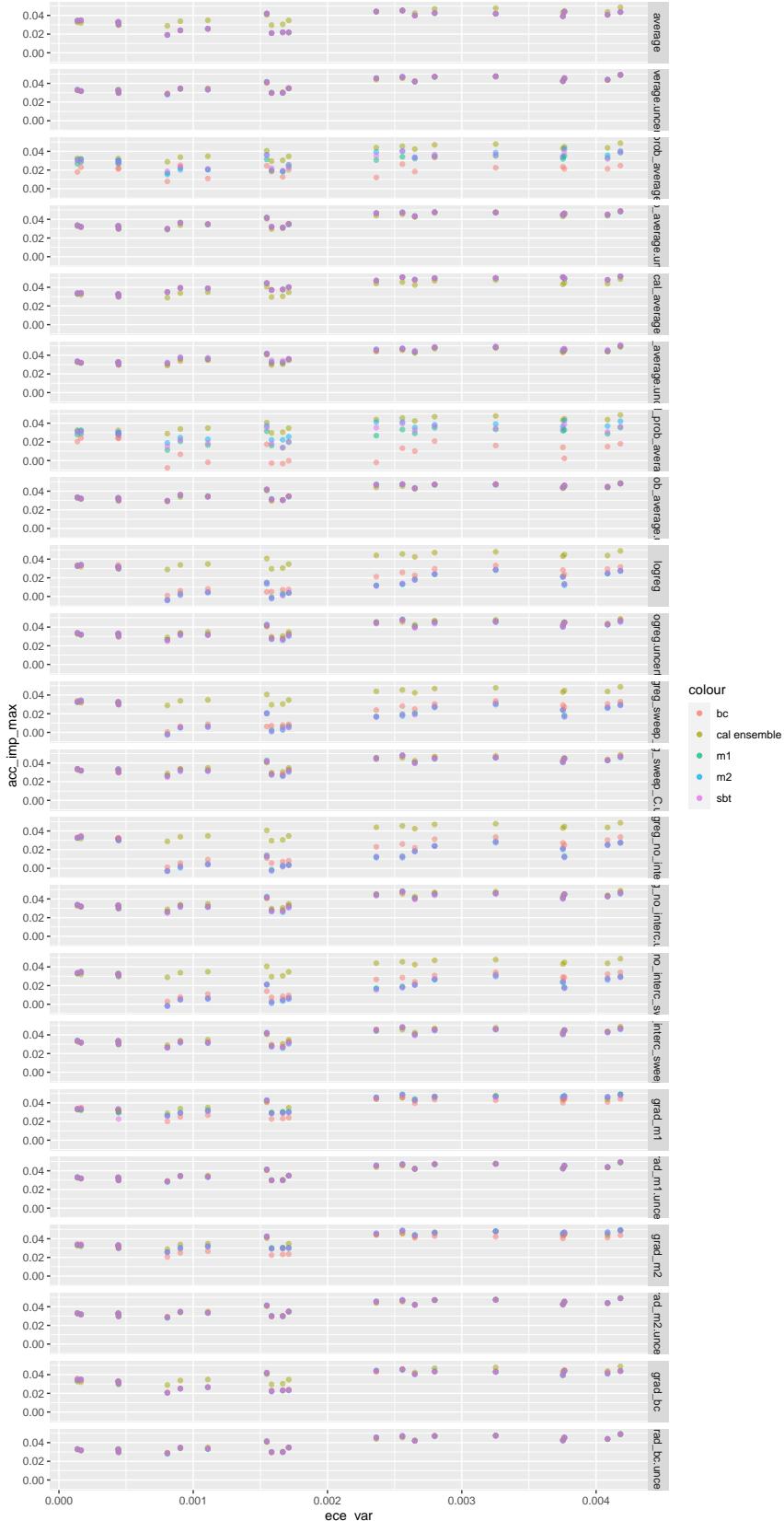
Accuracy improvement of ensemble over
the best of networks vs acc_avg.
Ensemble size 3



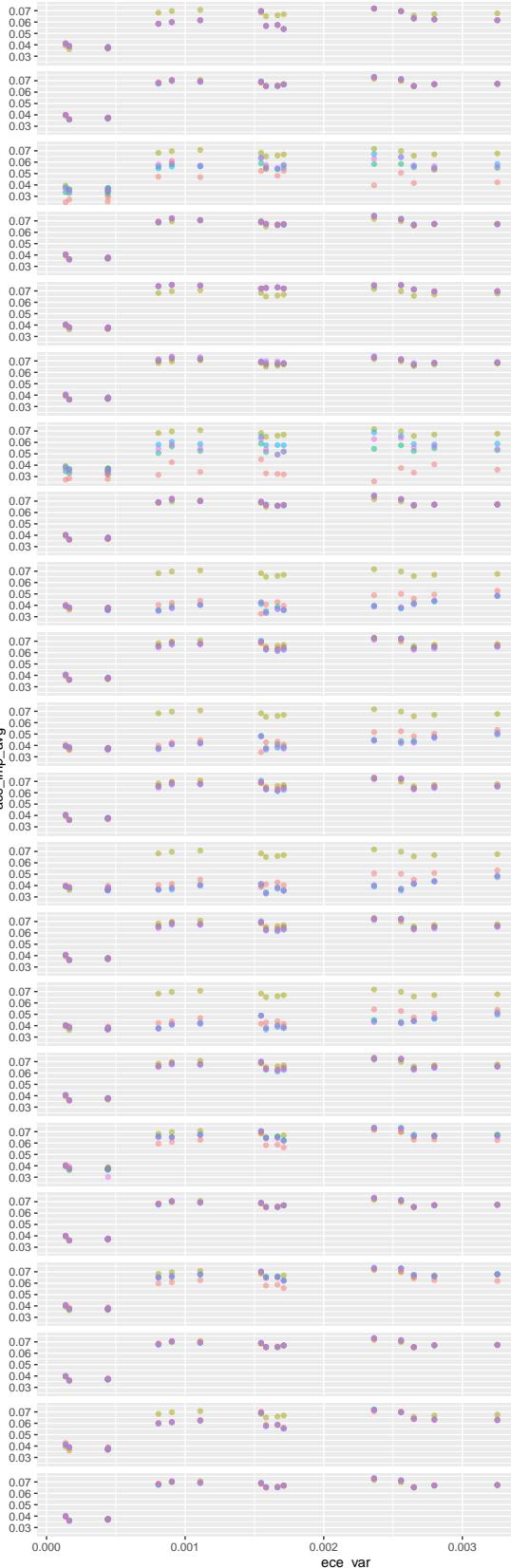
Accuracy improvement of ensemble over
the average of networks vs acc_avg.
Ensemble size 3



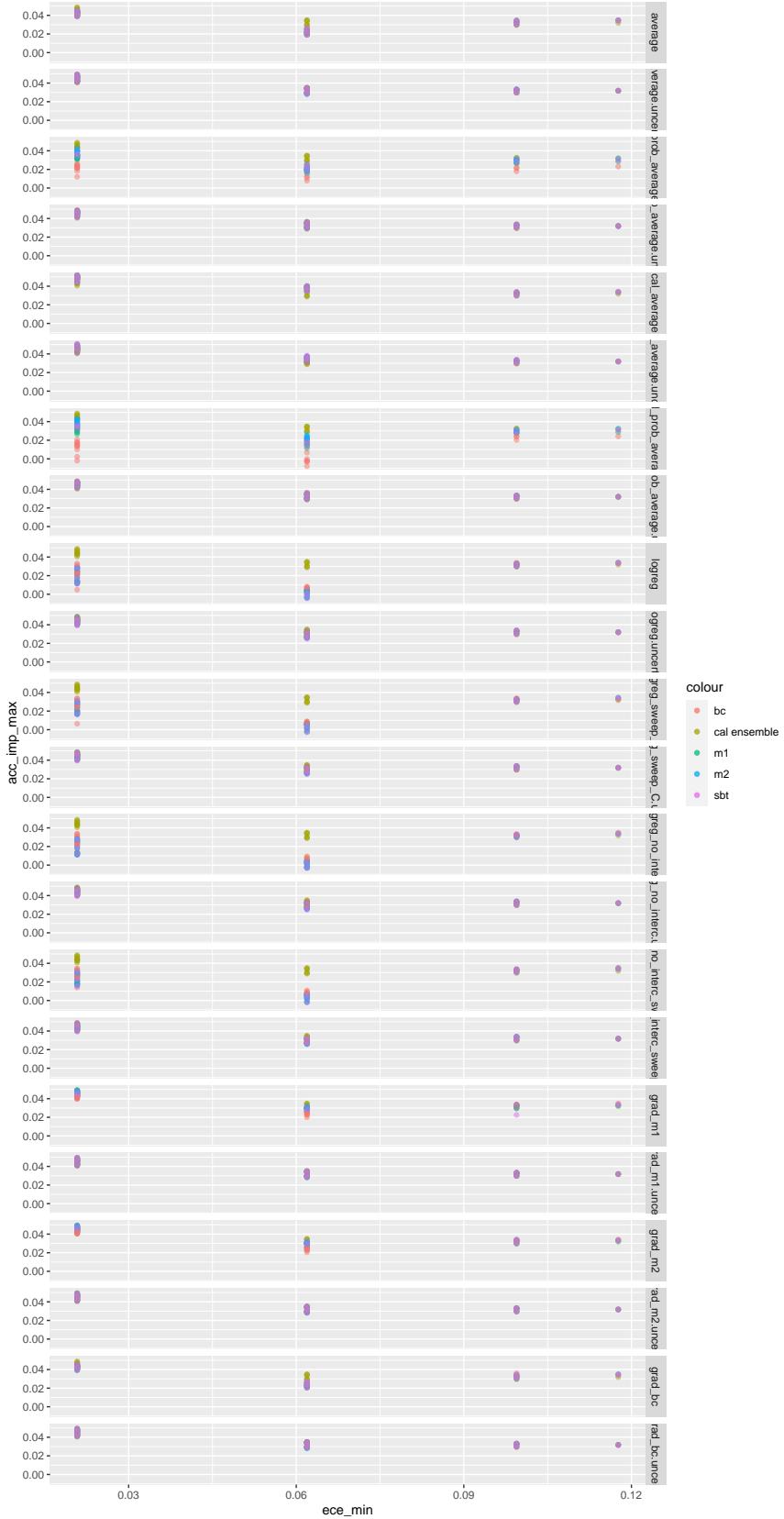
Accuracy improvement of ensemble over
the best of networks vs ece_var.
Ensemble size 3



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



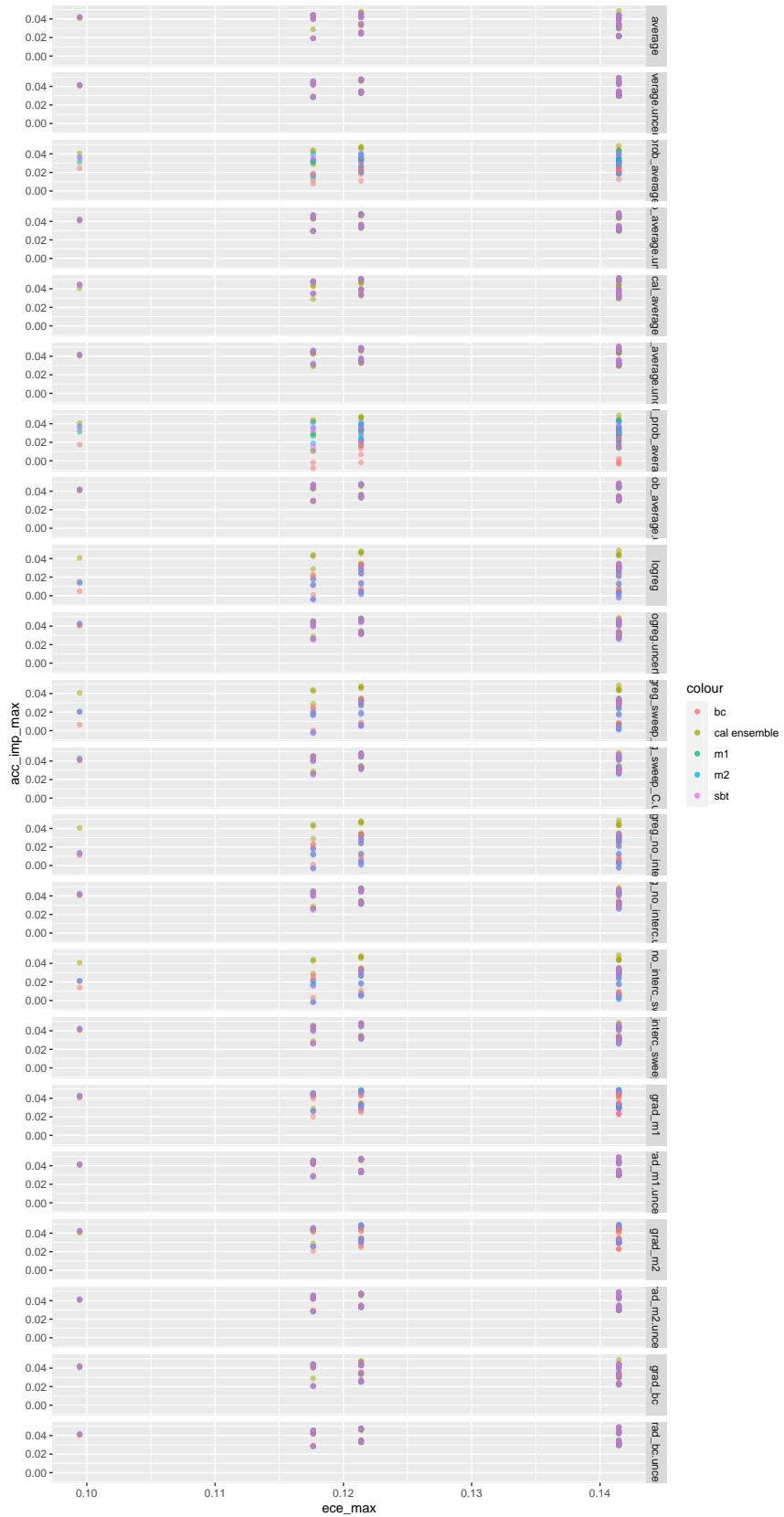
Accuracy improvement of ensemble over
the best of networks vs ece_min.
Ensemble size 3



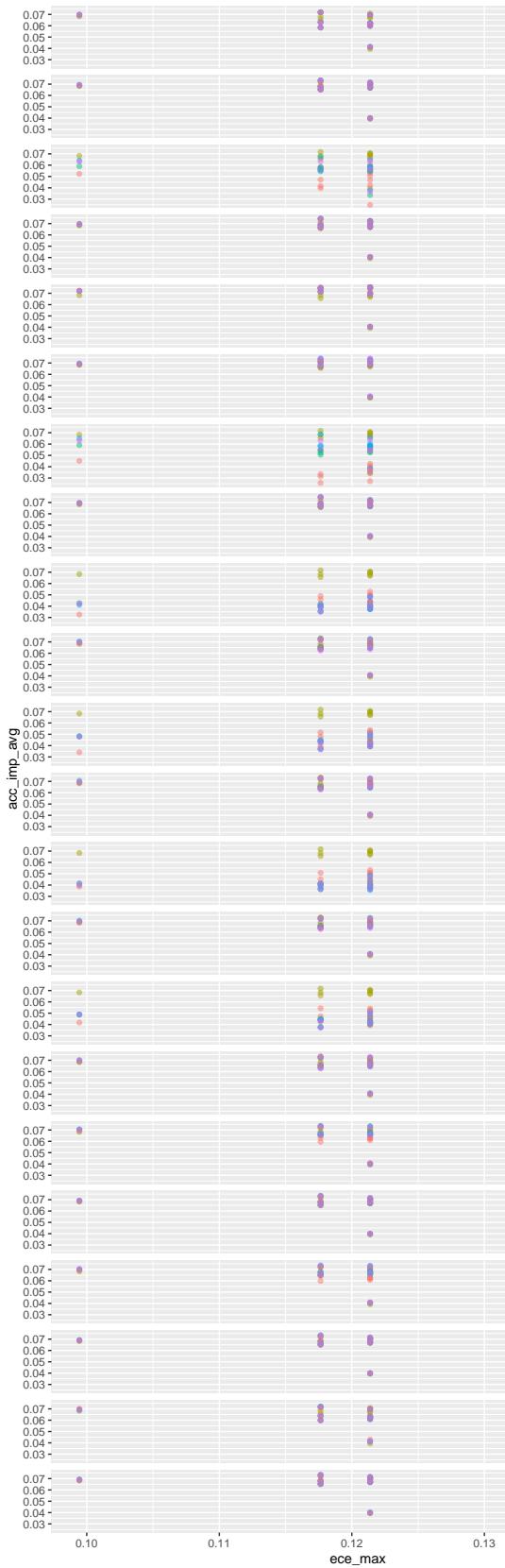
Accuracy improvement of ensemble over
the average of networks vs ece_min.
Ensemble size 3



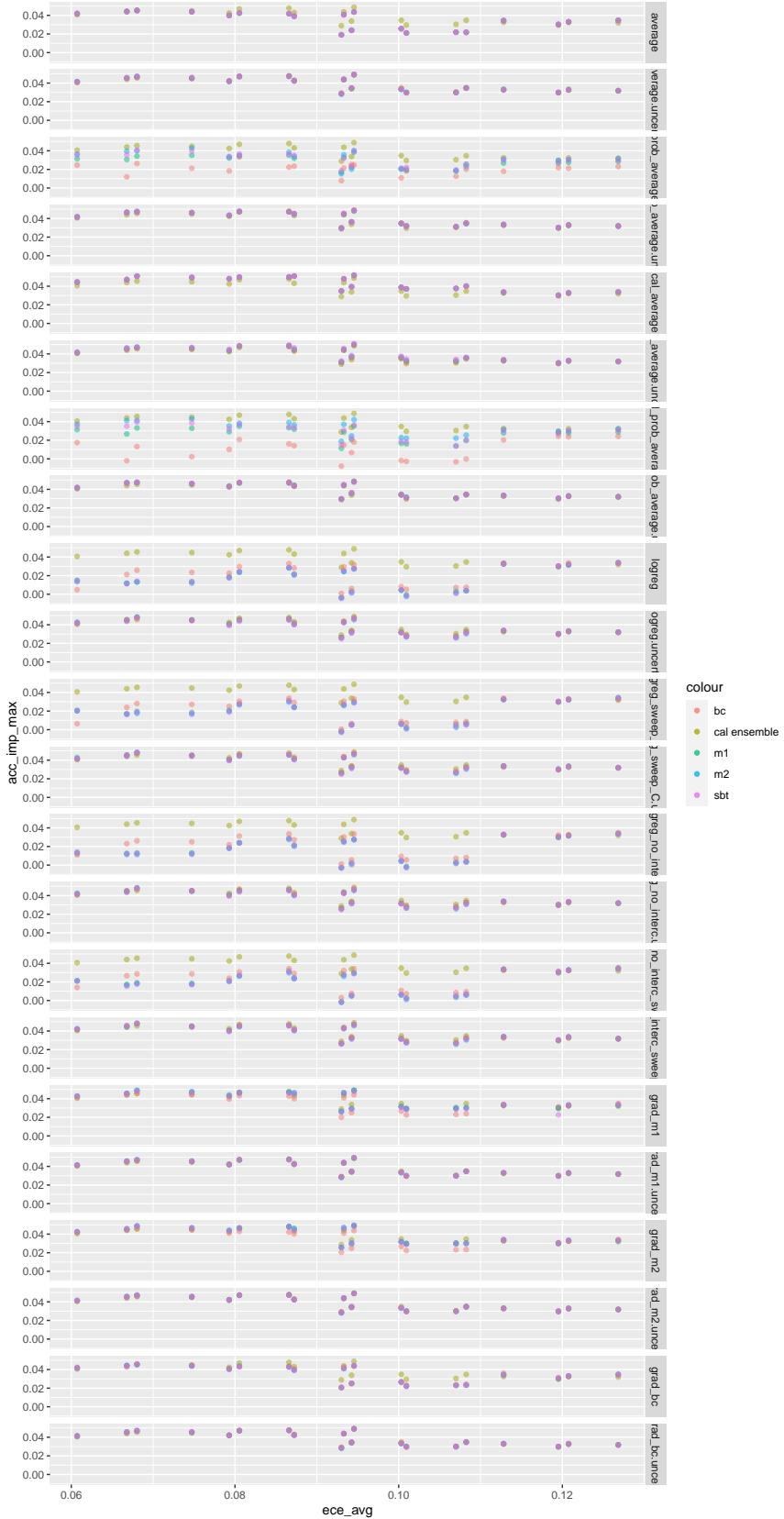
Accuracy improvement of ensemble over
the best of networks vs ece_max.
Ensemble size 3



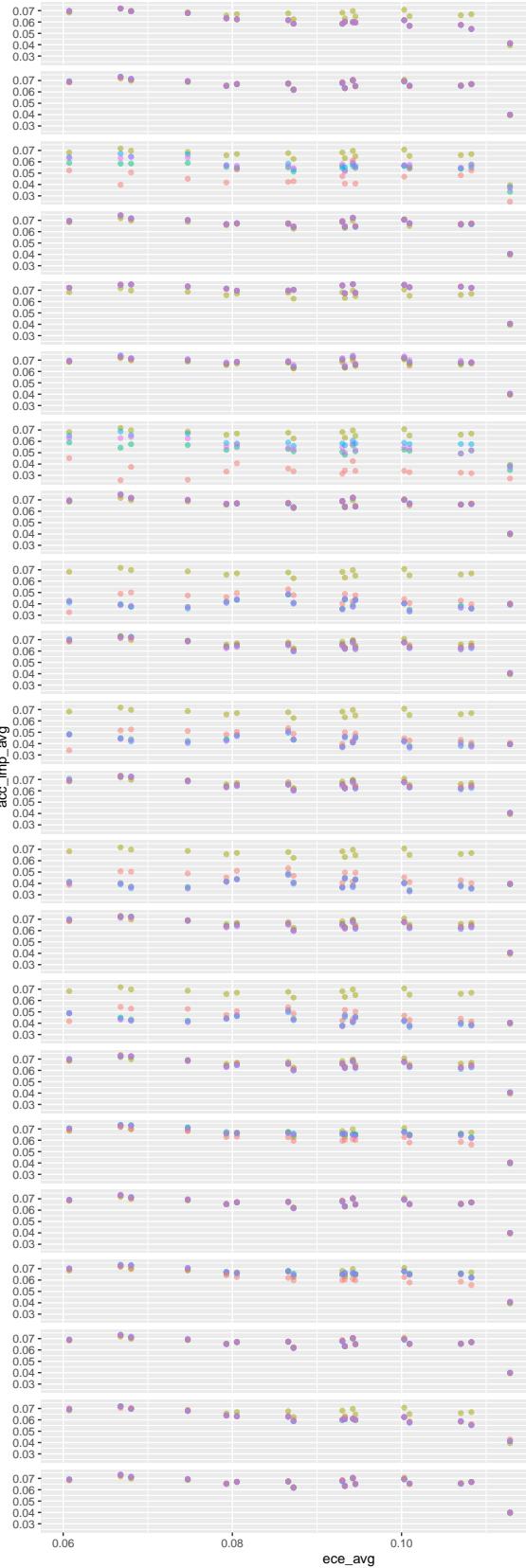
Accuracy improvement of ensemble over
the average of networks vs ece_max.
Ensemble size 3



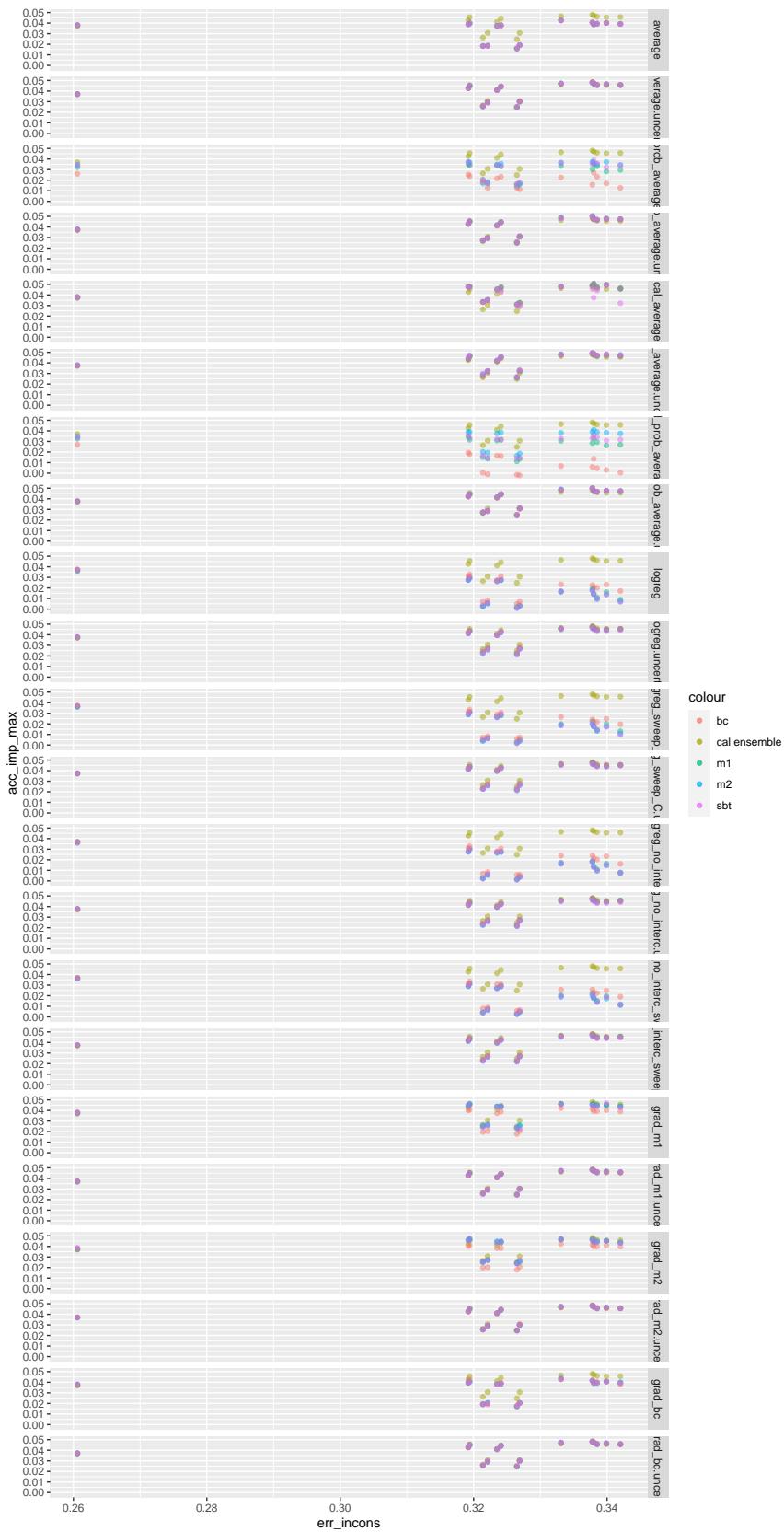
Accuracy improvement of ensemble over
the best of networks vs ece_avg.
Ensemble size 3



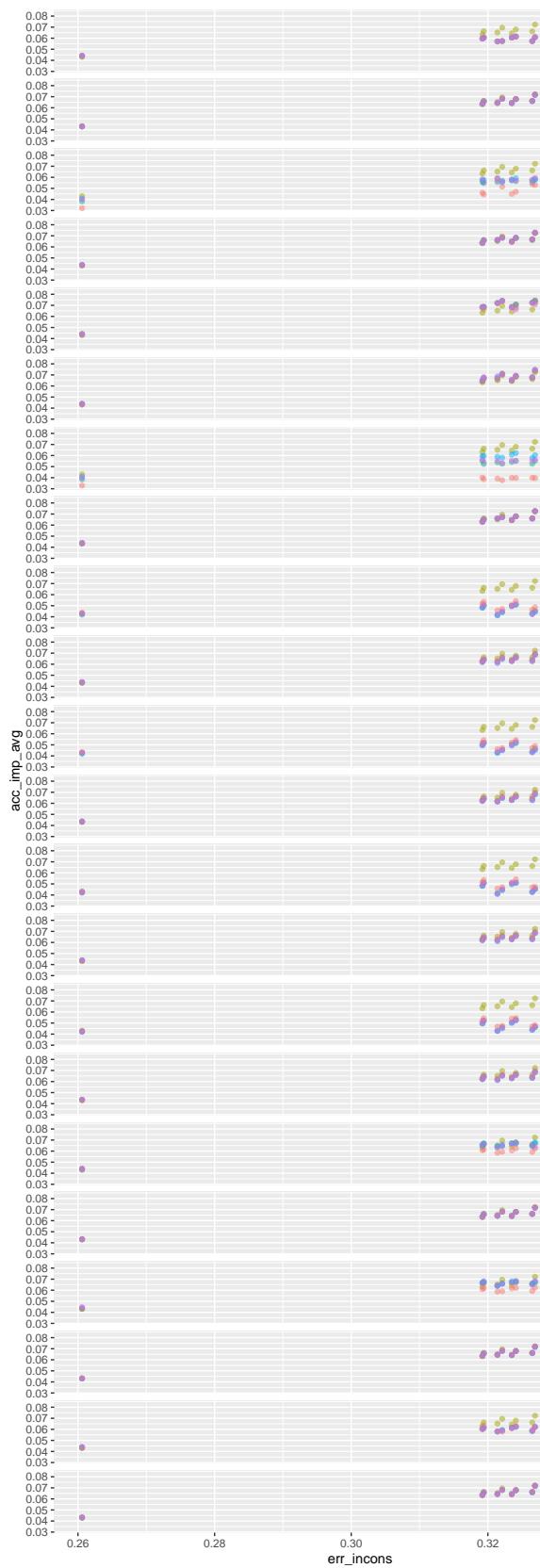
Accuracy improvement of ensemble over
the average of networks vs ece_avg.
Ensemble size 3



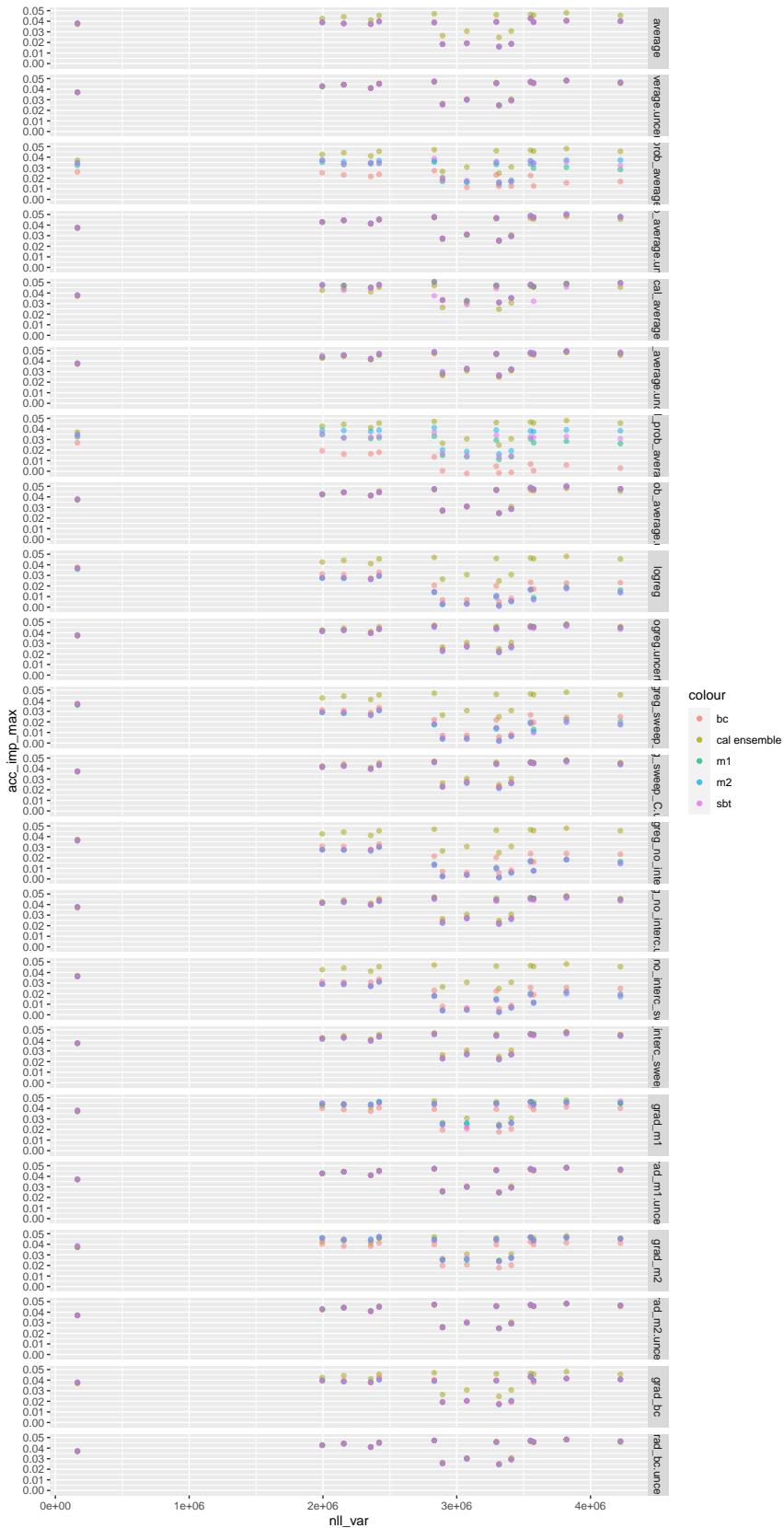
Accuracy improvement of ensemble over
the best of networks vs err_incons.
Ensemble size 4



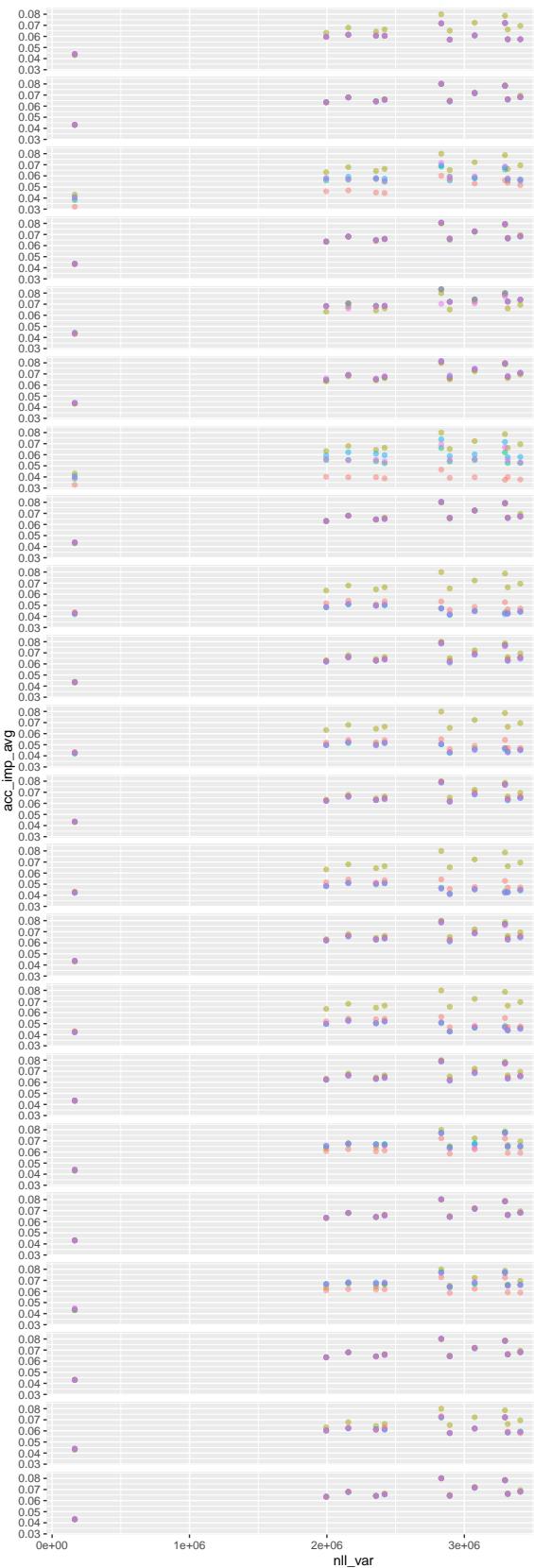
Accuracy improvement of ensemble over
the average of networks vs err_incons.
Ensemble size 4



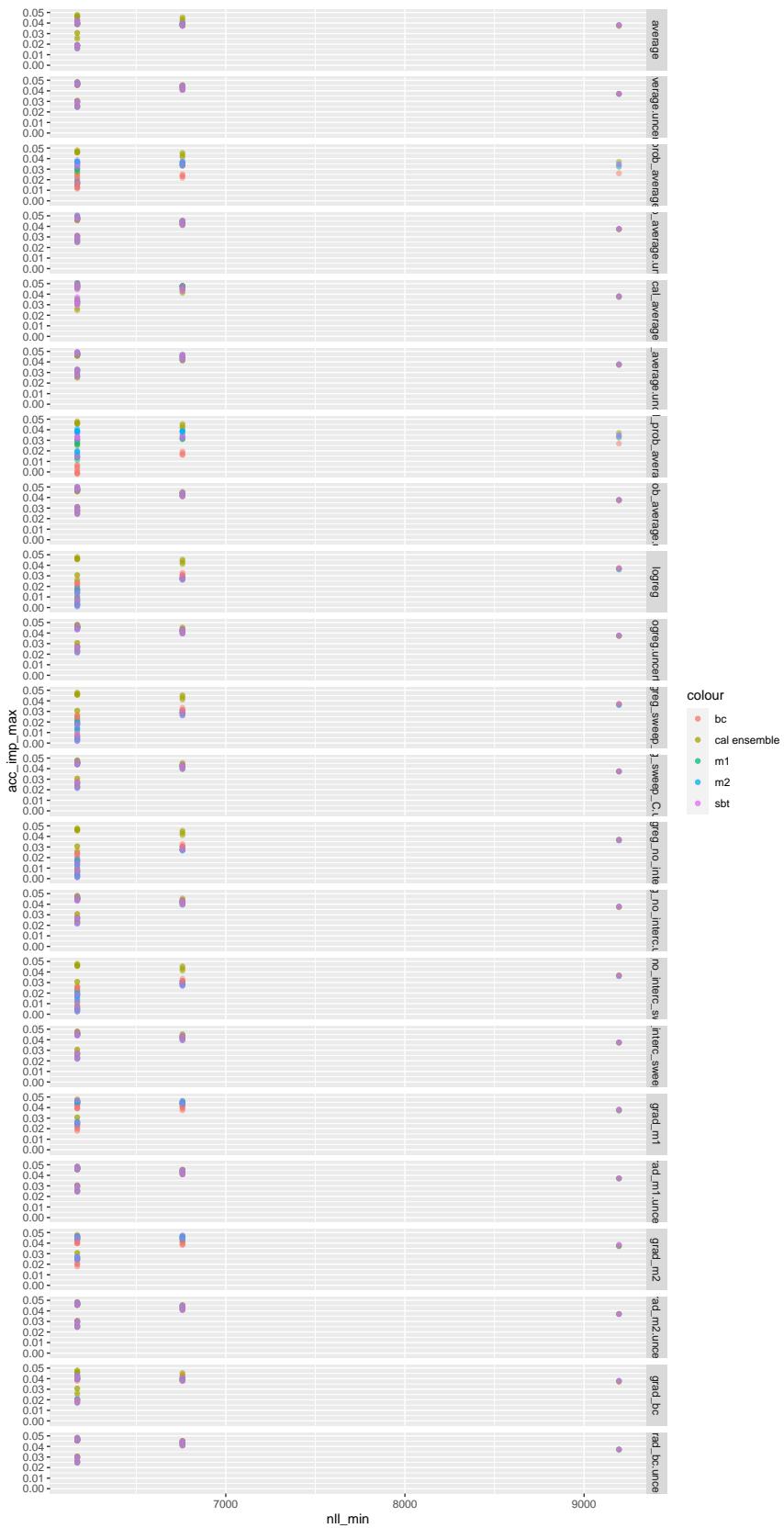
Accuracy improvement of ensemble over
the best of networks vs nll_var.
Ensemble size 4



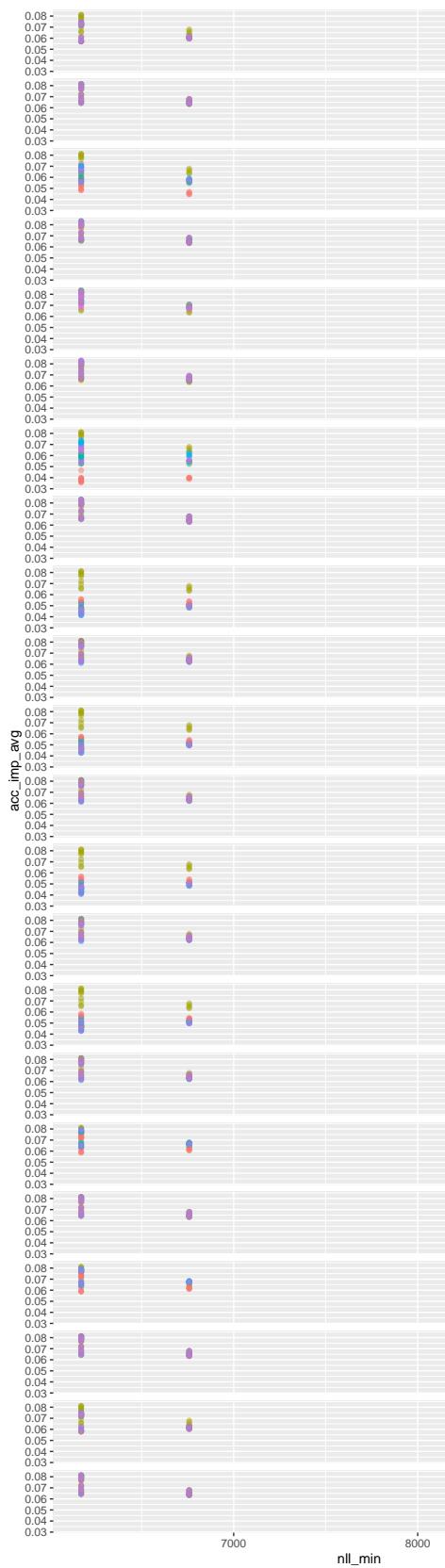
Accuracy improvement of ensemble over
the average of networks vs nll_var.
Ensemble size 4



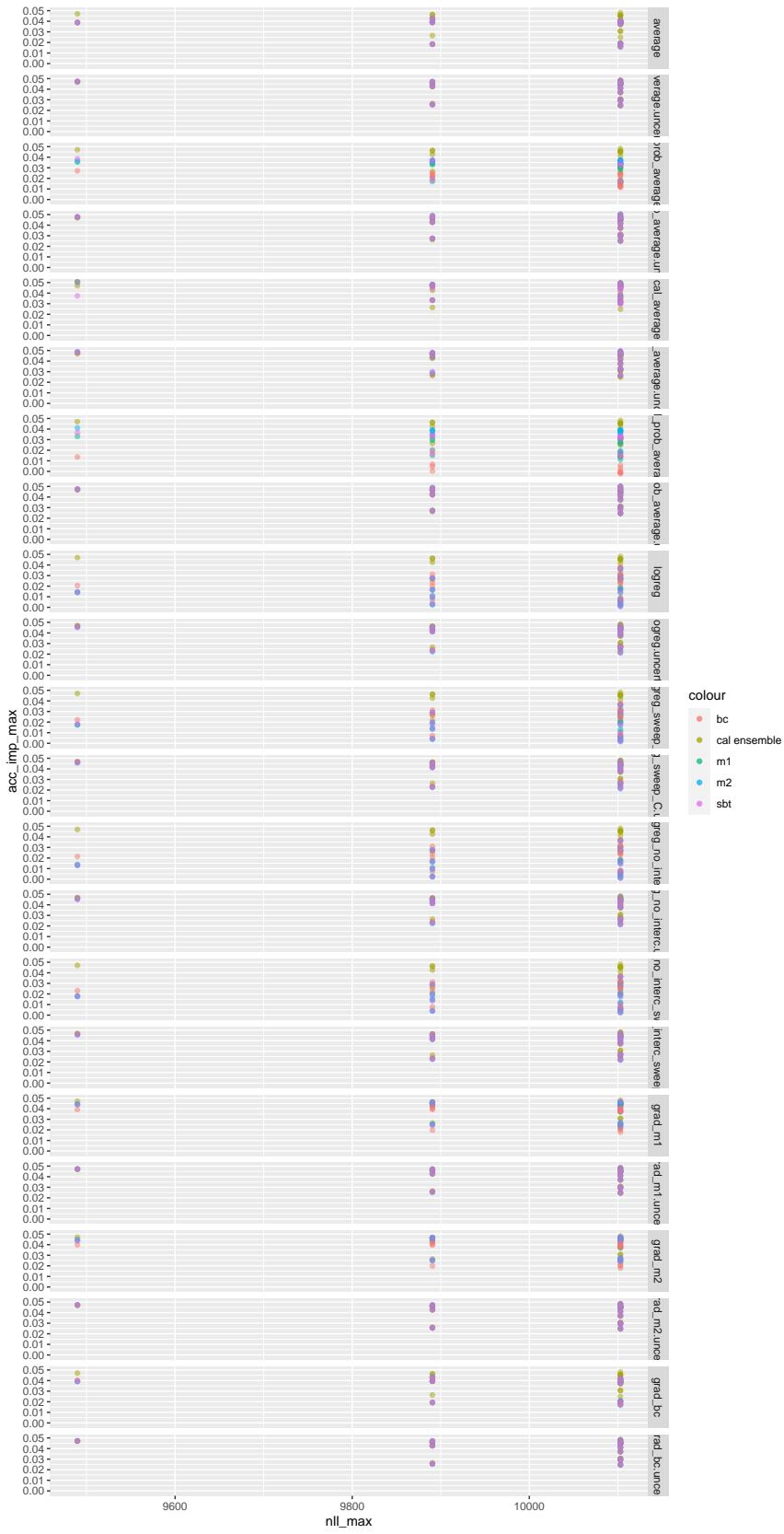
Accuracy improvement of ensemble over
the best of networks vs nll_min.
Ensemble size 4



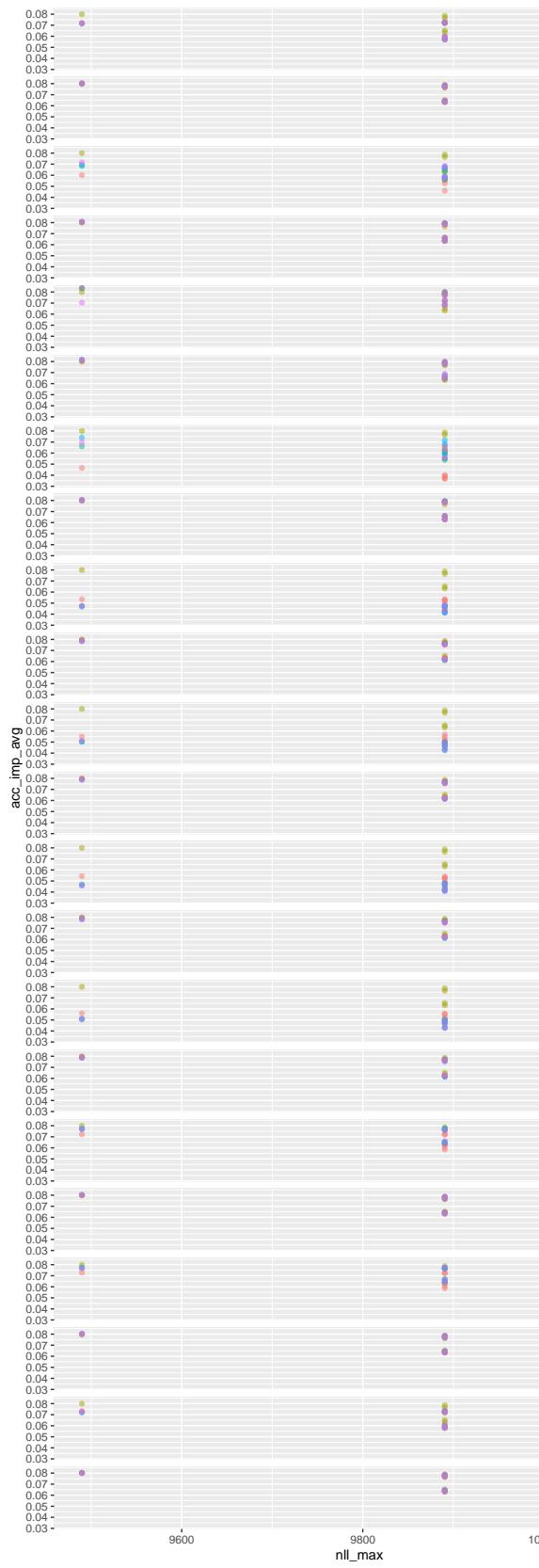
Accuracy improvement of ensemble over
the average of networks vs nll_min.
Ensemble size 4



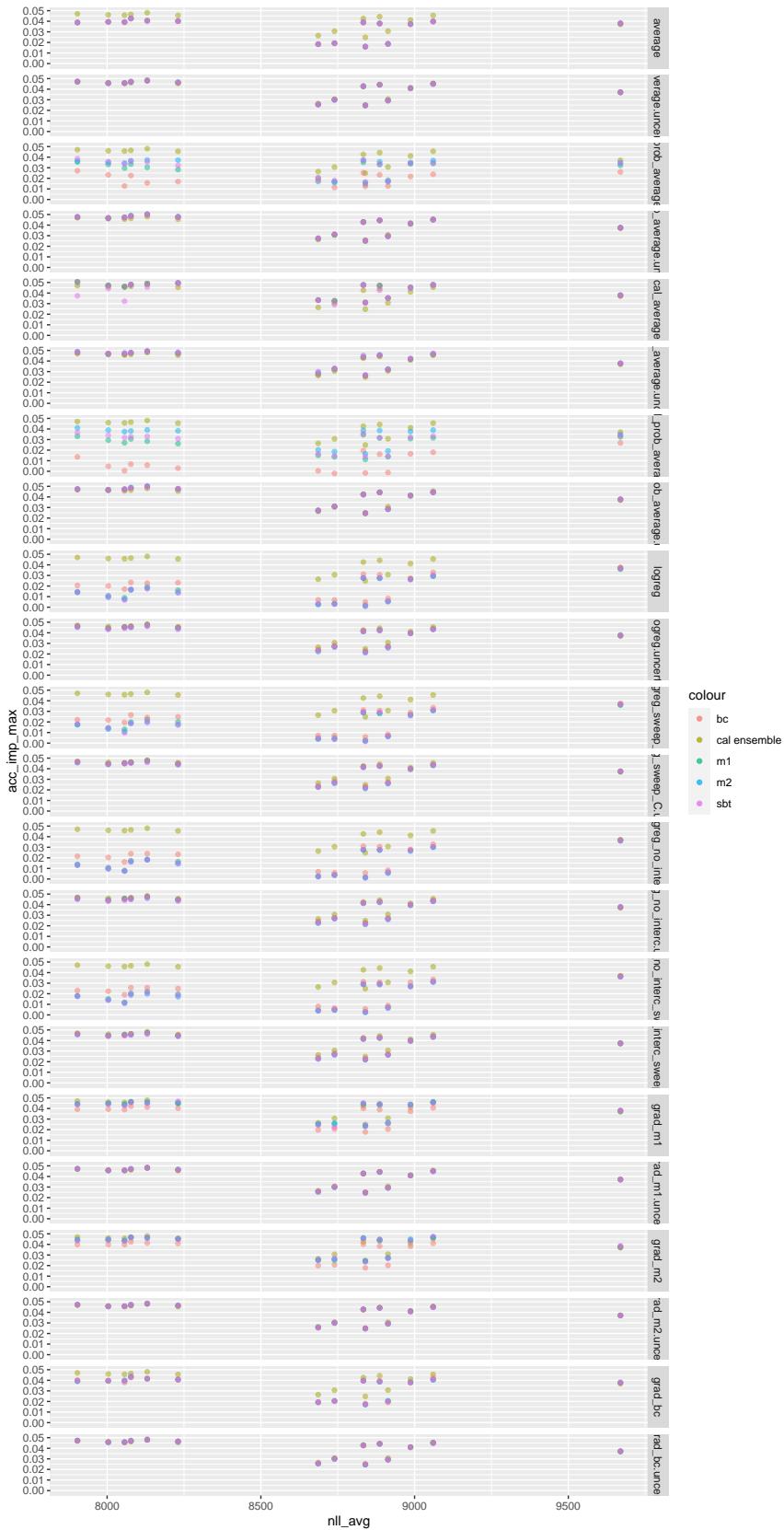
Accuracy improvement of ensemble over
the best of networks vs nll_max.
Ensemble size 4



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



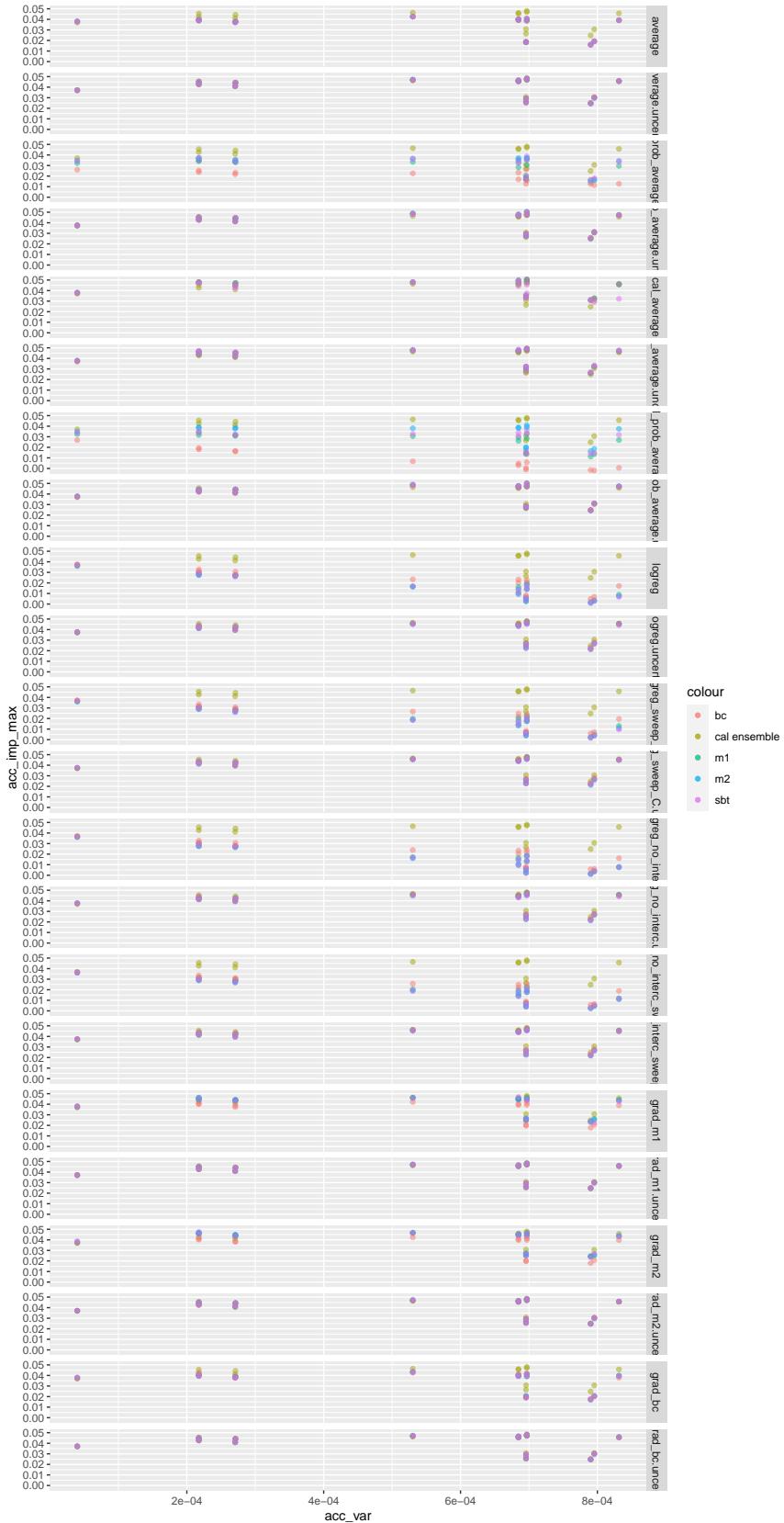
Accuracy improvement of ensemble over
the best of networks vs nll_avg.
Ensemble size 4



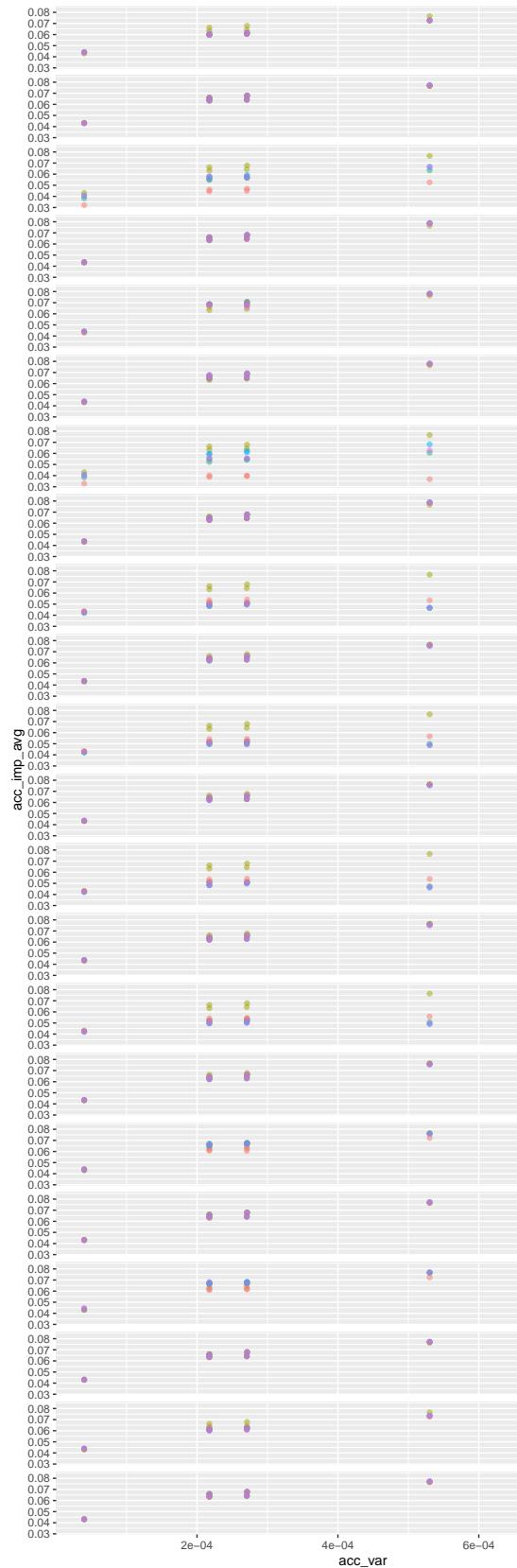
Accuracy improvement of ensemble over
the average of networks vs nll_avg.
Ensemble size 4



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



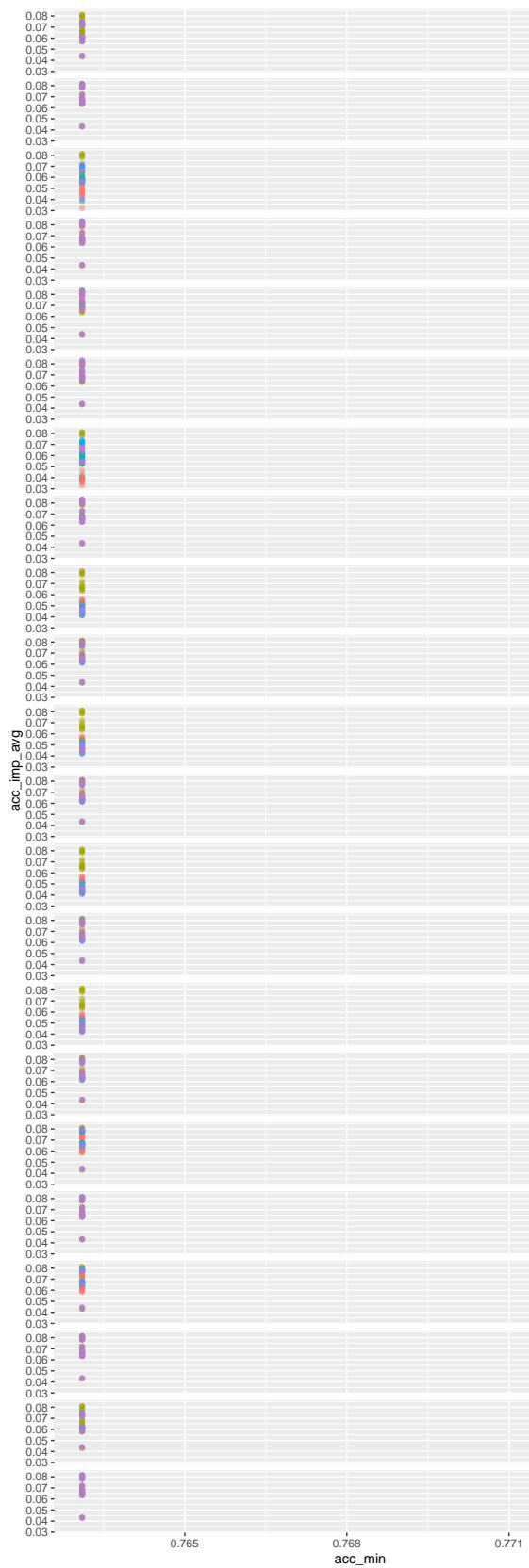
Accuracy improvement of ensemble over
the average of networks vs acc_var.
Ensemble size 4



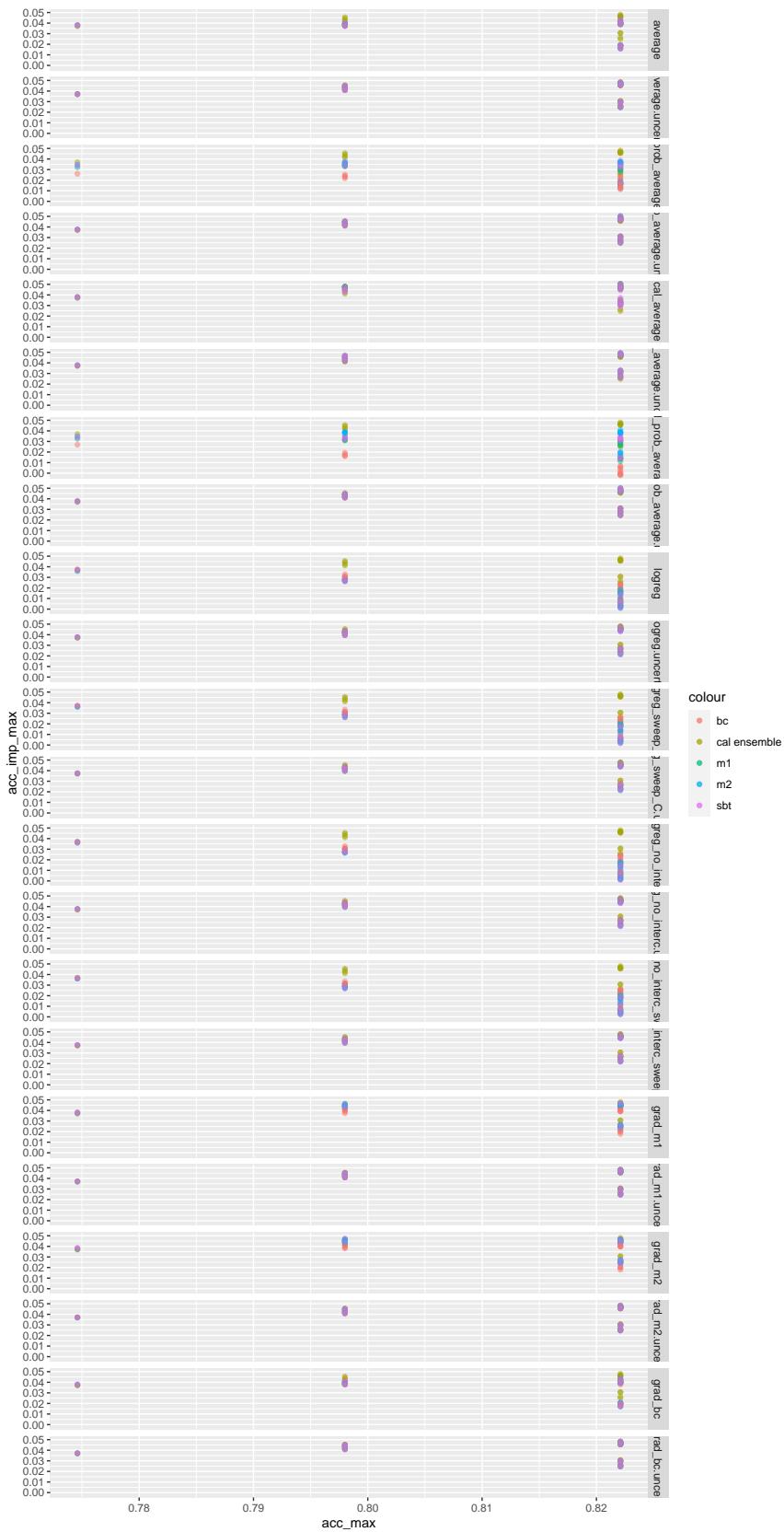
Accuracy improvement of ensemble over
the best of networks vs acc_min.
Ensemble size 4



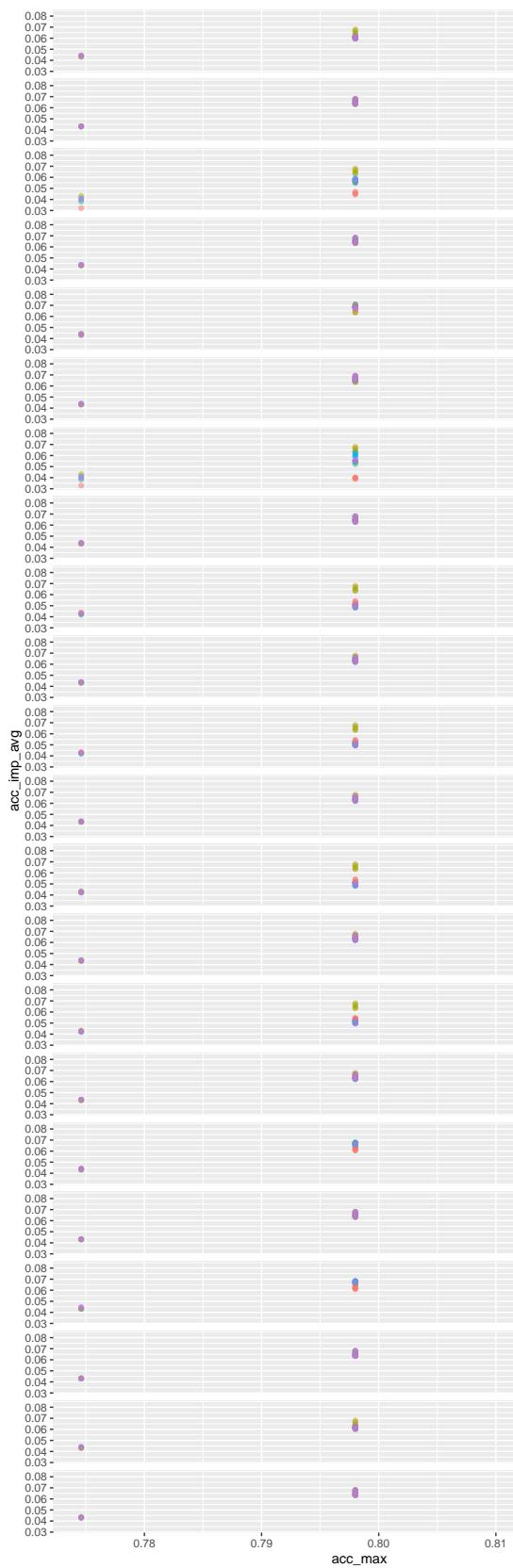
Accuracy improvement of ensemble over
the average of networks vs acc_min.
Ensemble size 4



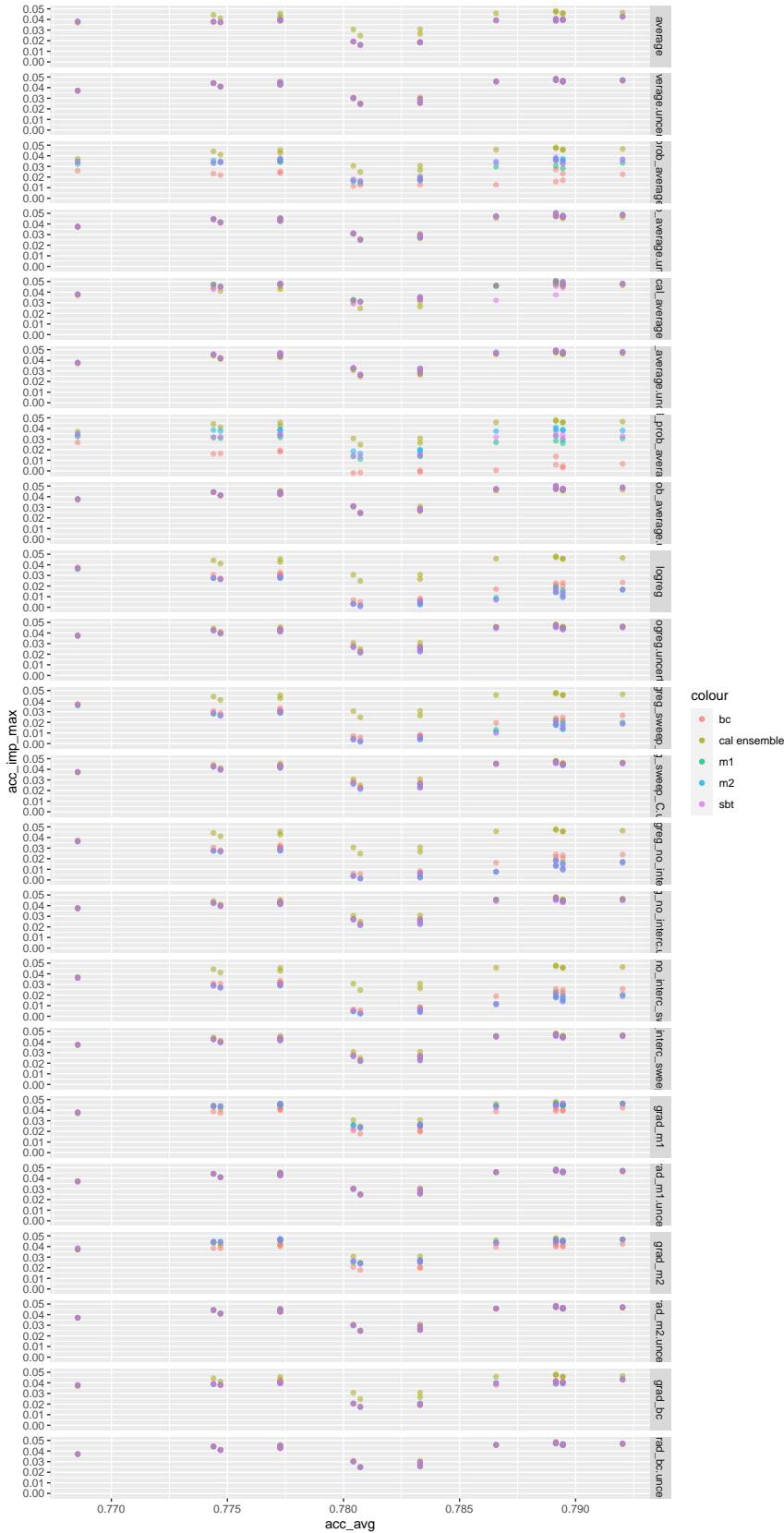
Accuracy improvement of ensemble over
the best of networks vs acc_max.
Ensemble size 4



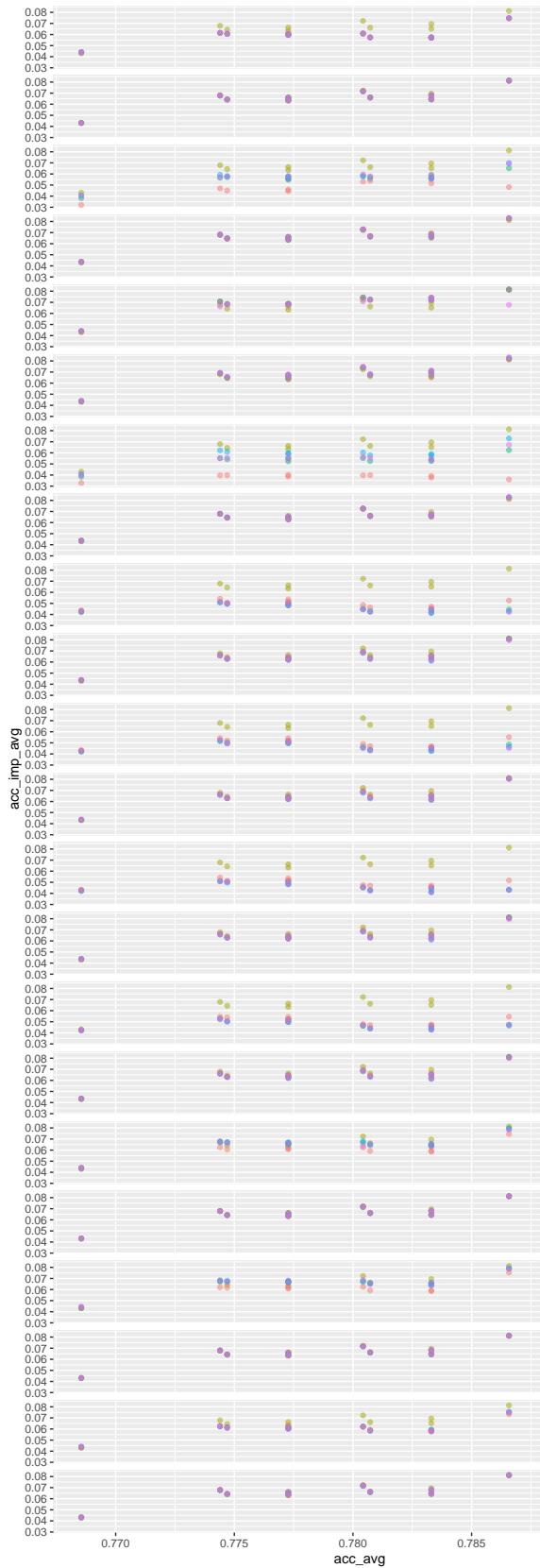
Accuracy improvement of ensemble over
the average of networks vs acc_max.
Ensemble size 4



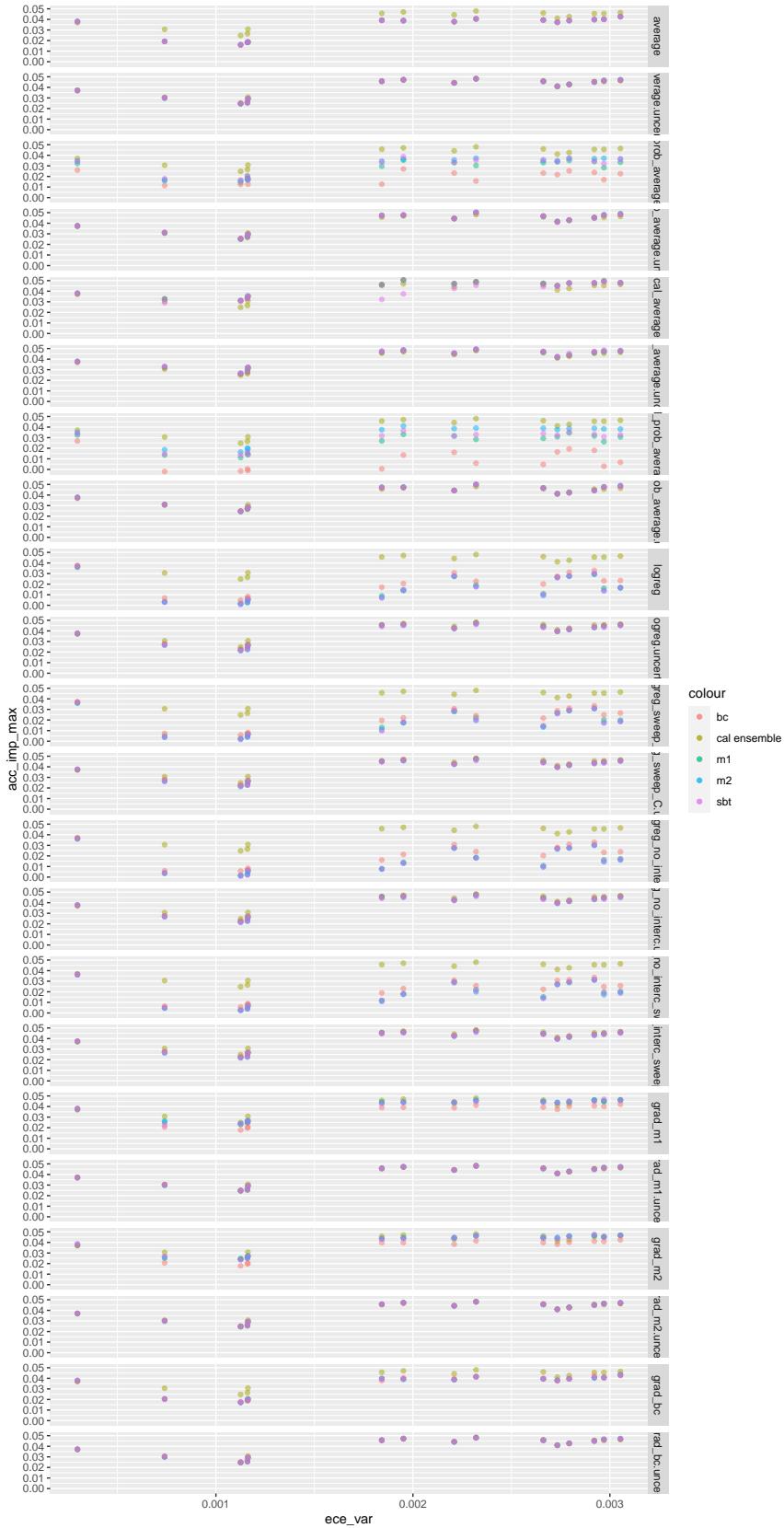
Accuracy improvement of ensemble over
the best of networks vs acc_avg.
Ensemble size 4



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



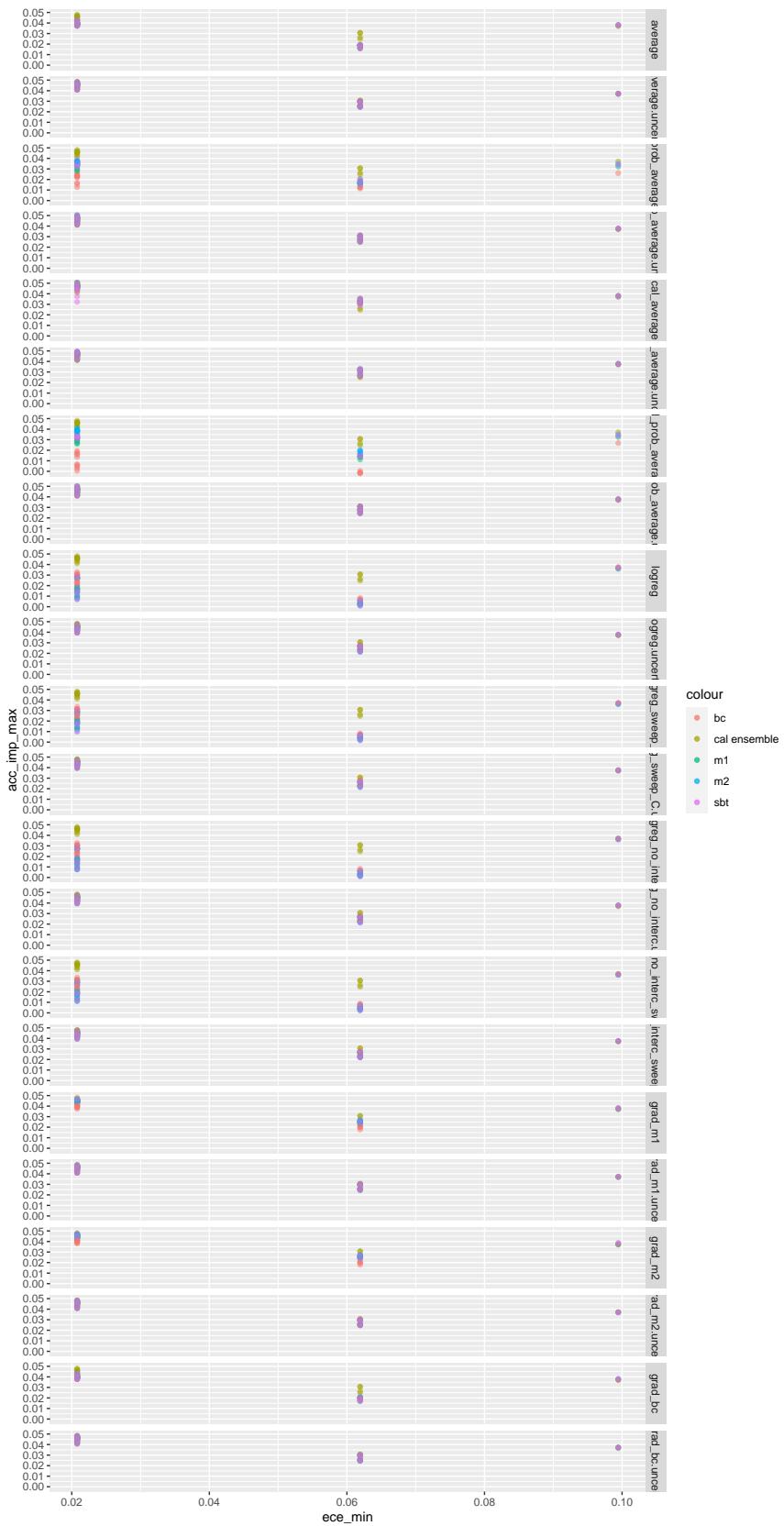
Accuracy improvement of ensemble over
the best of networks vs ece_var.
Ensemble size 4



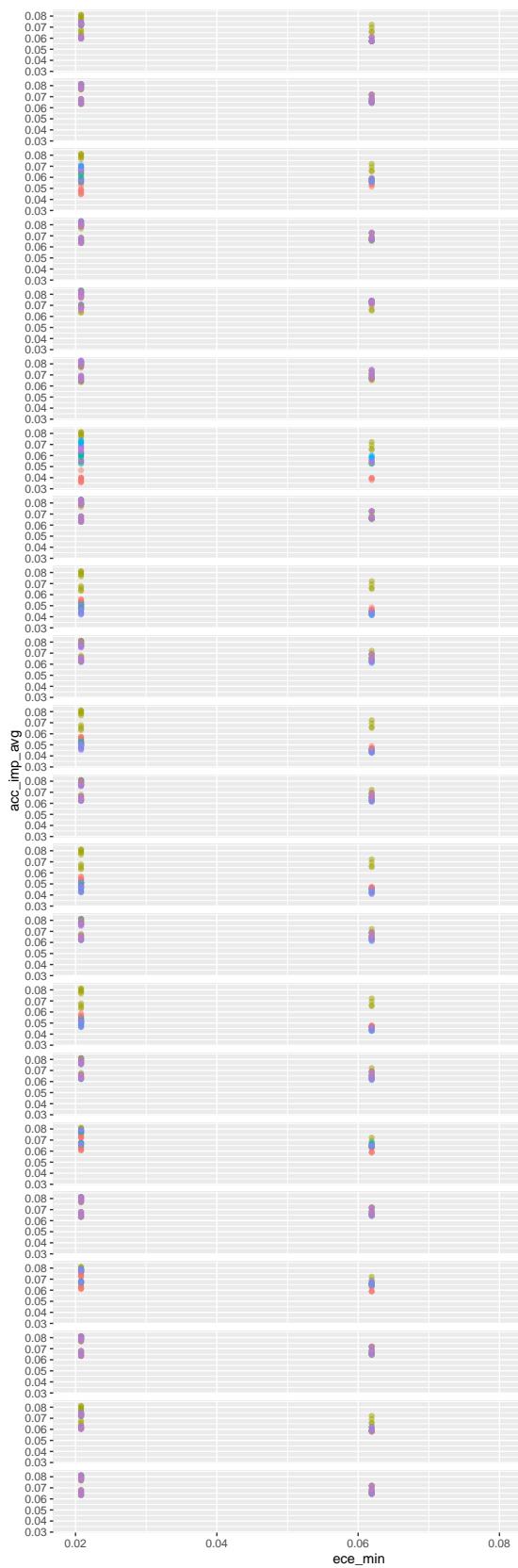
Accuracy improvement of ensemble over
the average of networks vs ece_var.
Ensemble size 4



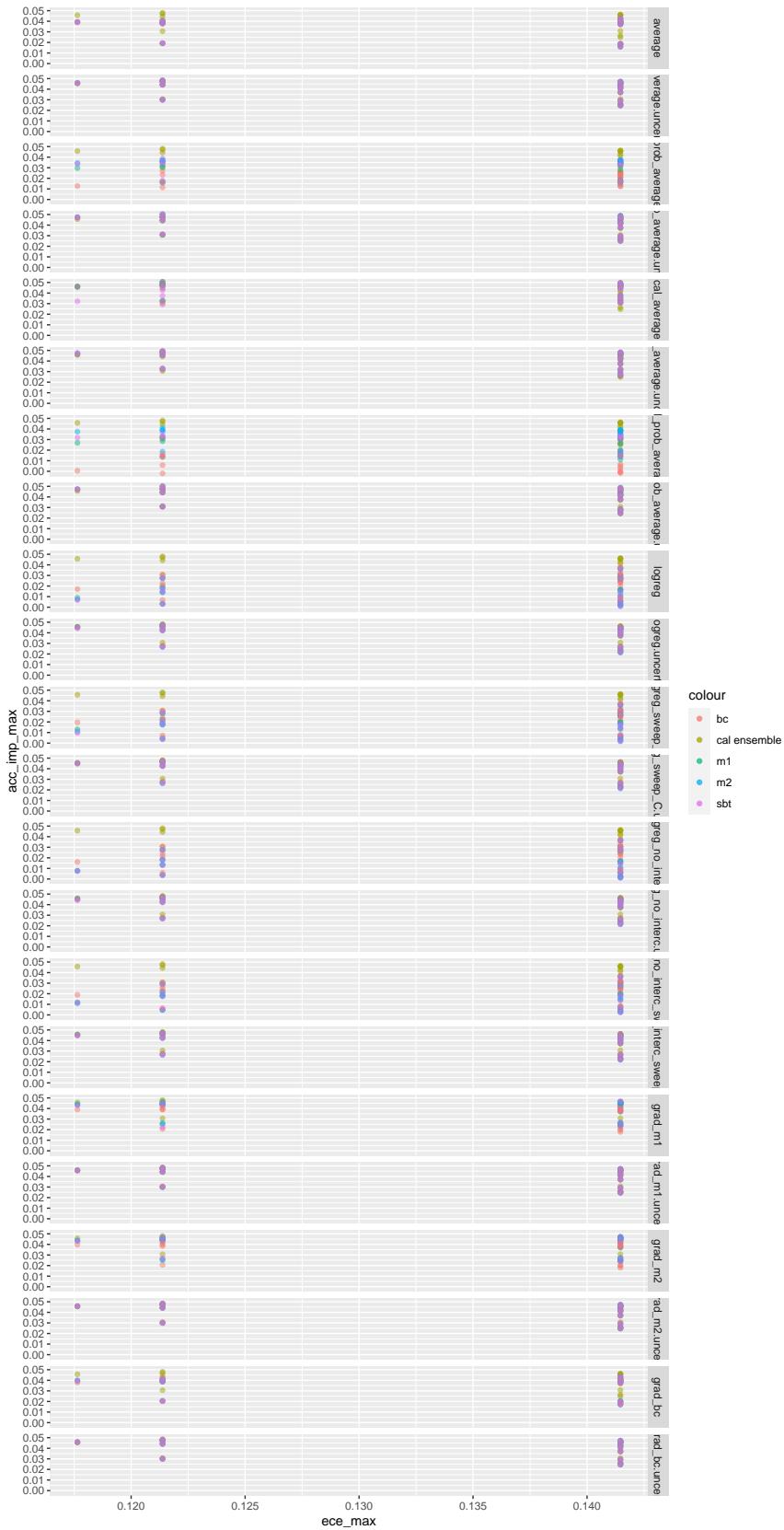
Accuracy improvement of ensemble over
the best of networks vs ece_min.
Ensemble size 4



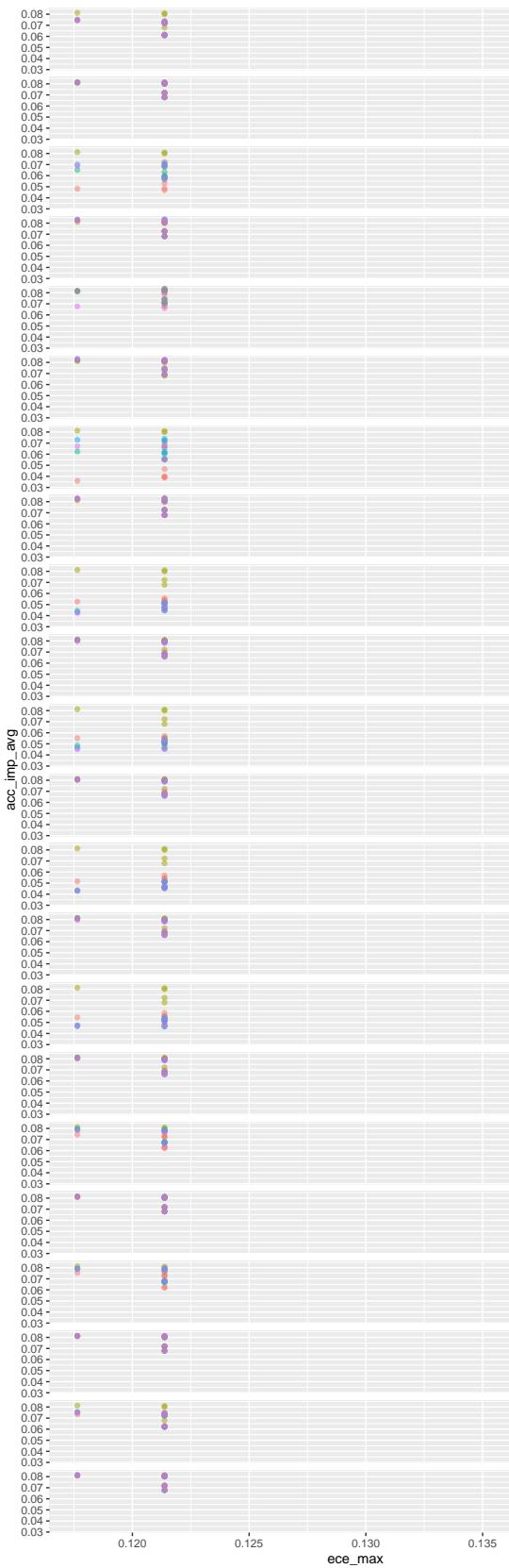
Accuracy improvement of ensemble over
the average of networks vs ece_min.
Ensemble size 4



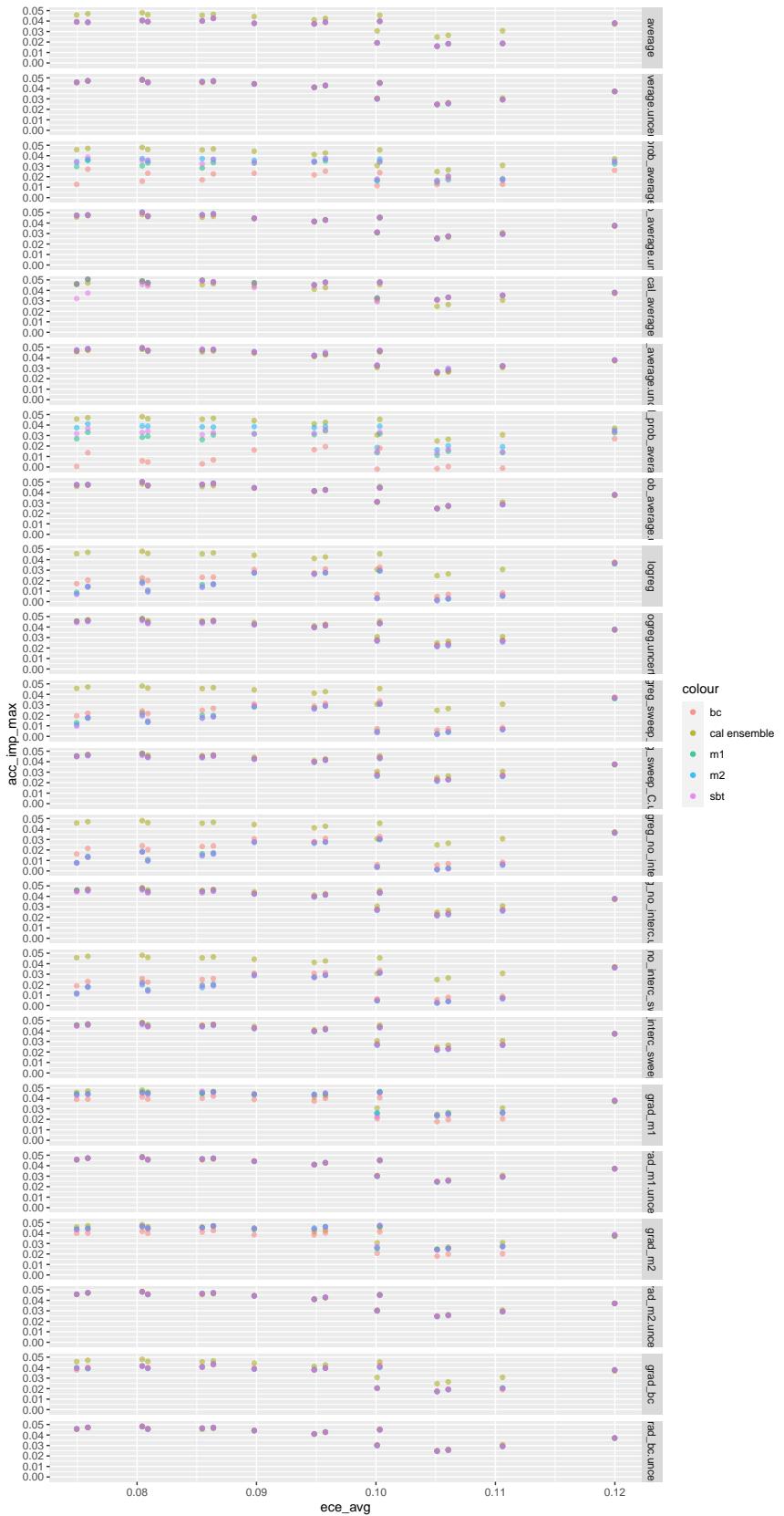
Accuracy improvement of ensemble over
the best of networks vs ece_max.
Ensemble size 4



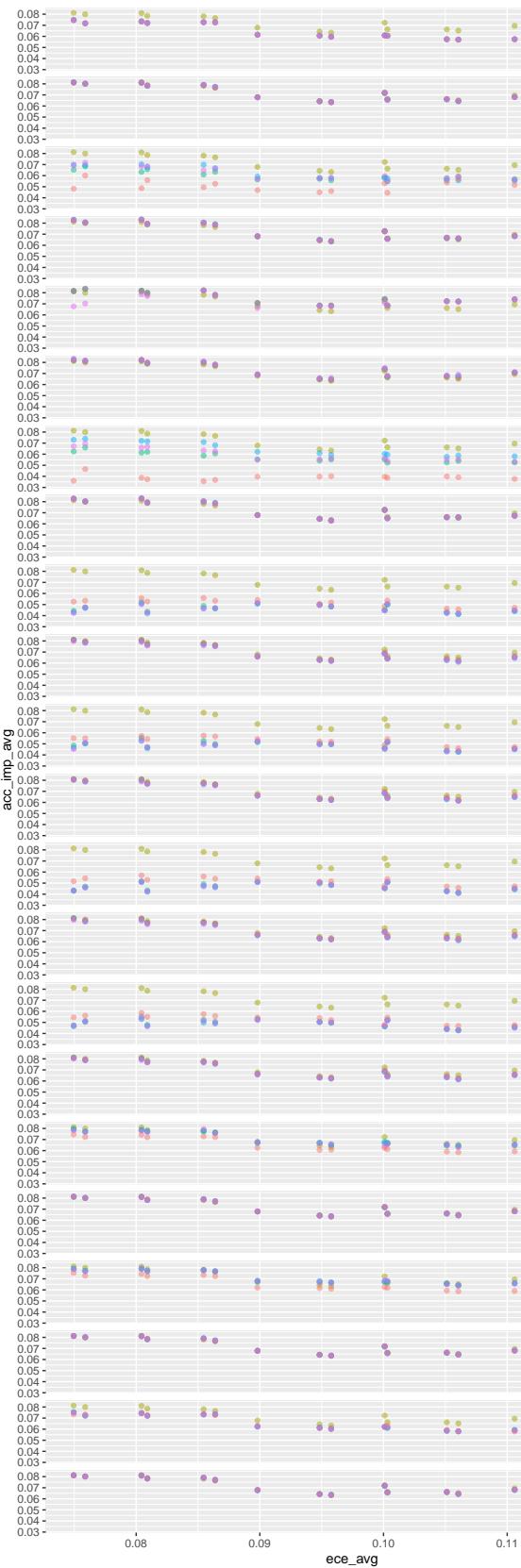
Accuracy improvement of ensemble over
the average of networks vs ece_max.
Ensemble size 4

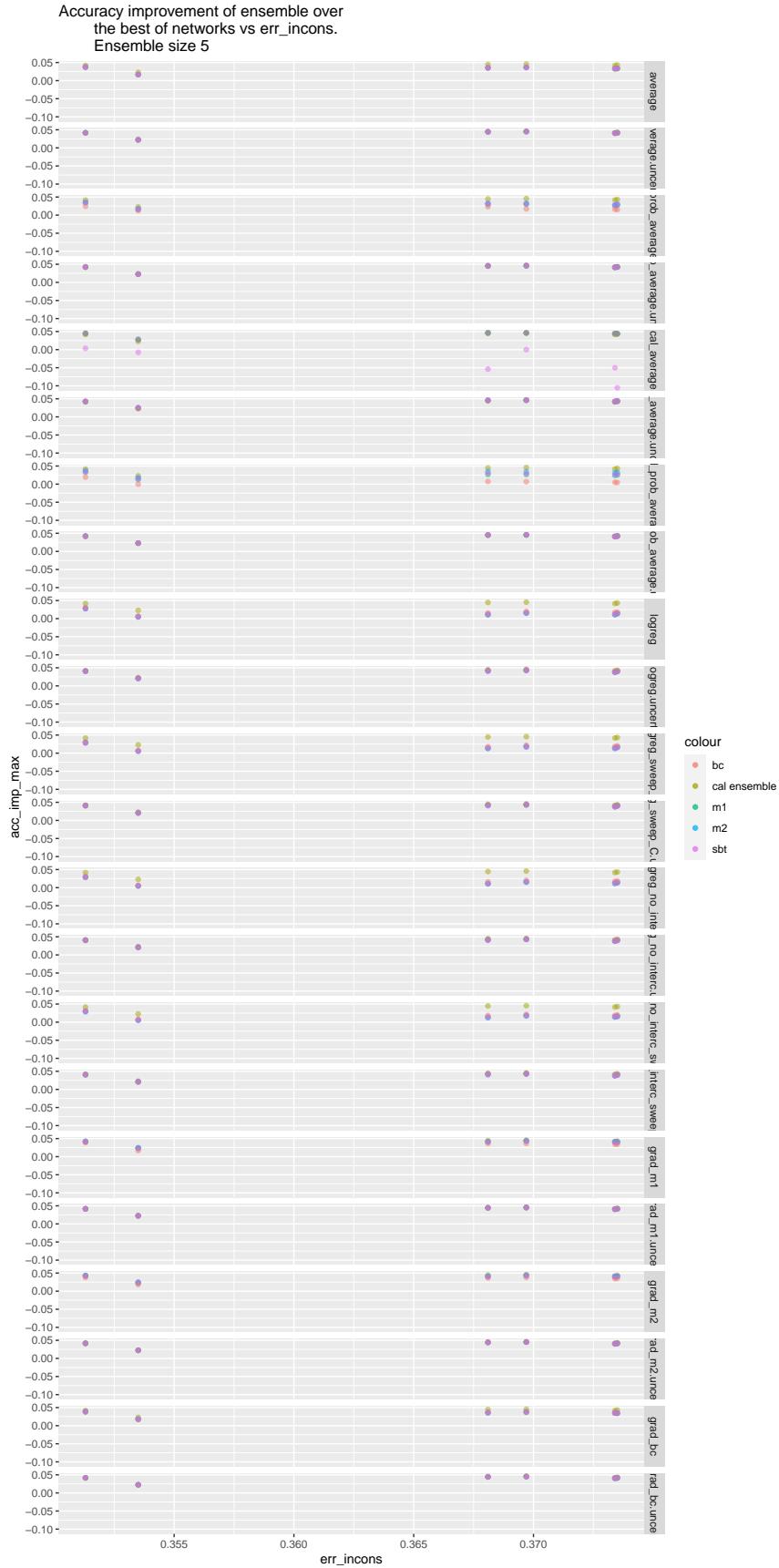


Accuracy improvement of ensemble over
the best of networks vs ece_avg.
Ensemble size 4

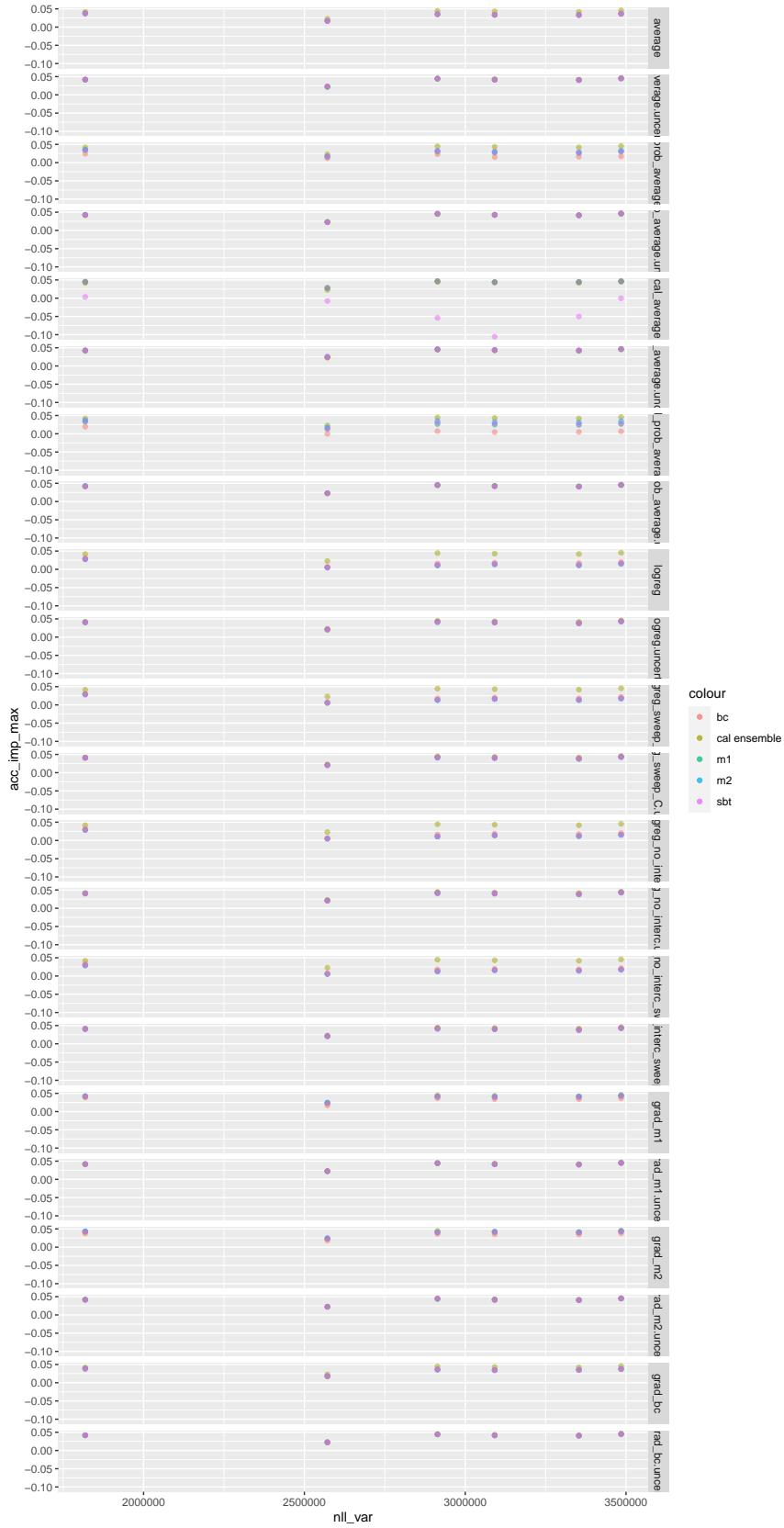


Accuracy improvement of ensemble over
the average of networks vs ece_avg.
Ensemble size 4

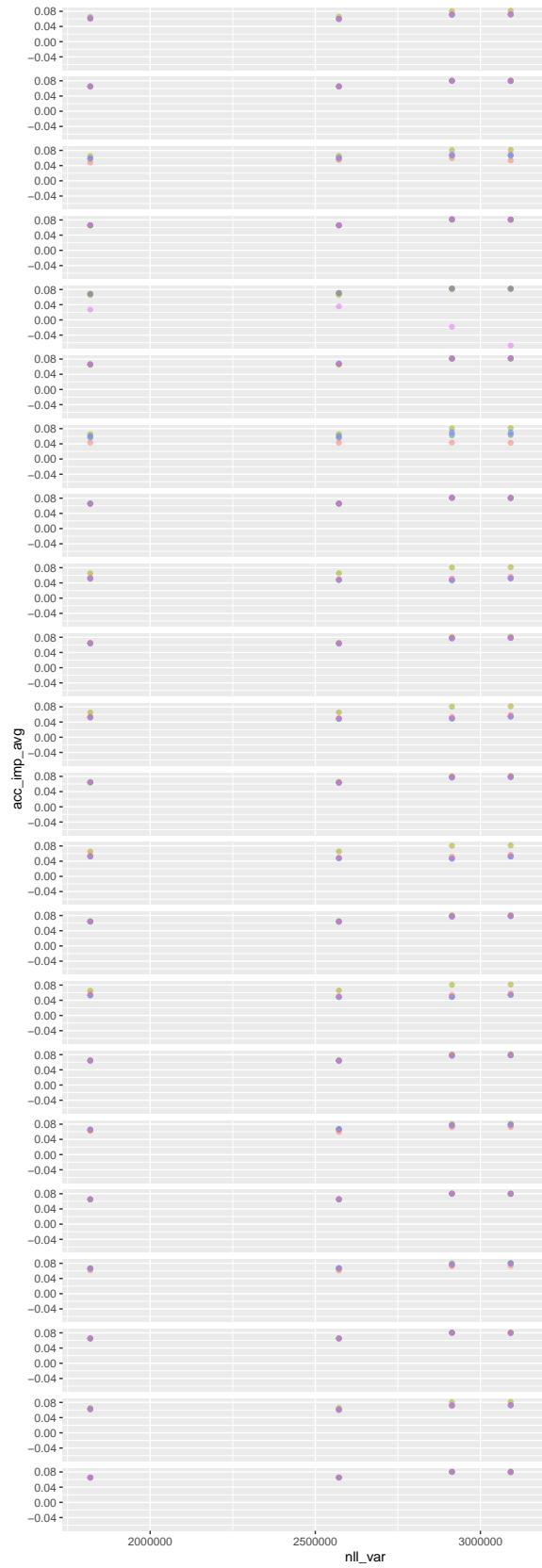




Accuracy improvement of ensemble over
the best of networks vs nll_var.
Ensemble size 5



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



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



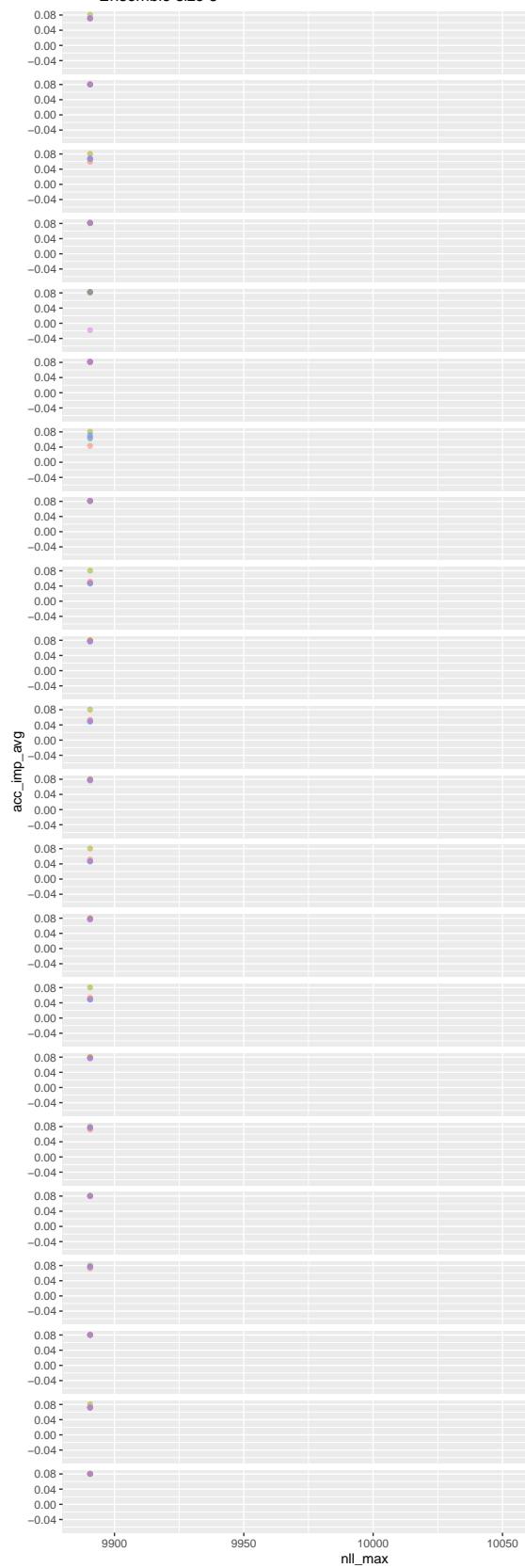
Accuracy improvement of ensemble over
the average of networks vs nll_min.
Ensemble size 5

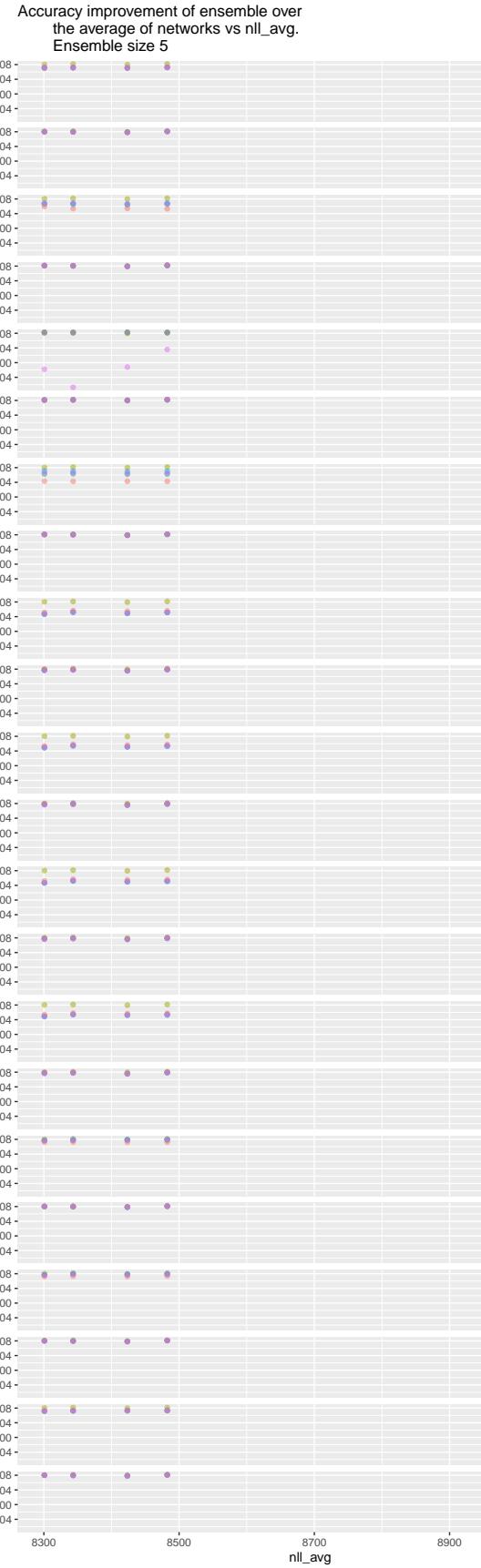
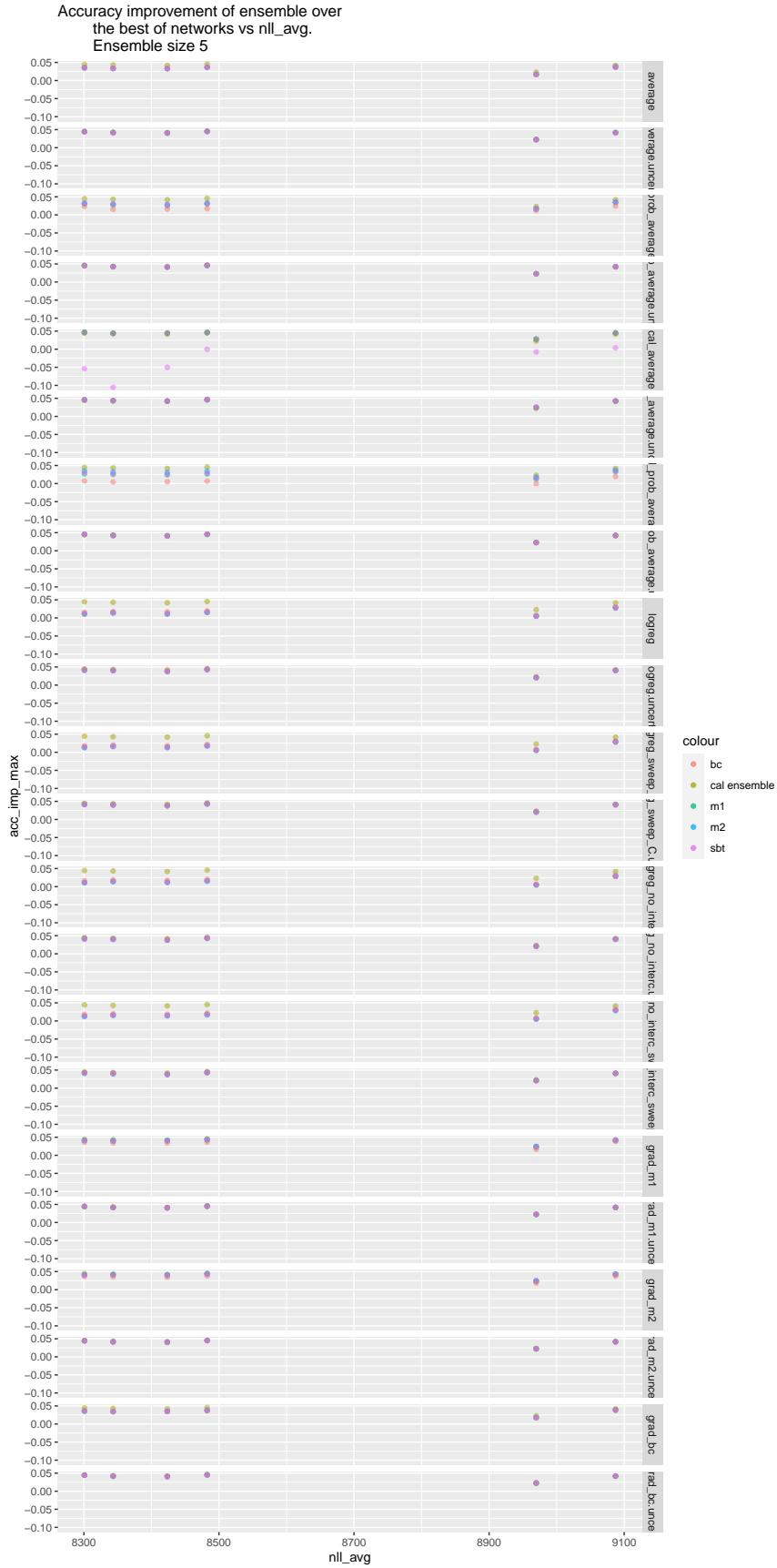


Accuracy improvement of ensemble over
the best of networks vs nll_max.
Ensemble size 5



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

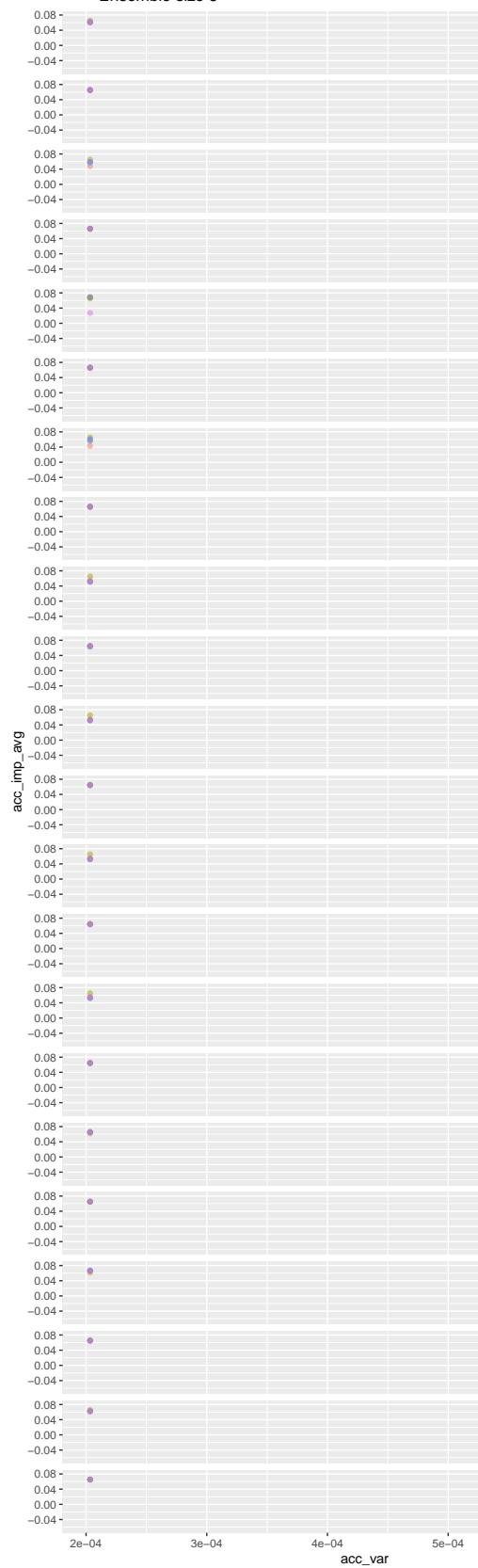


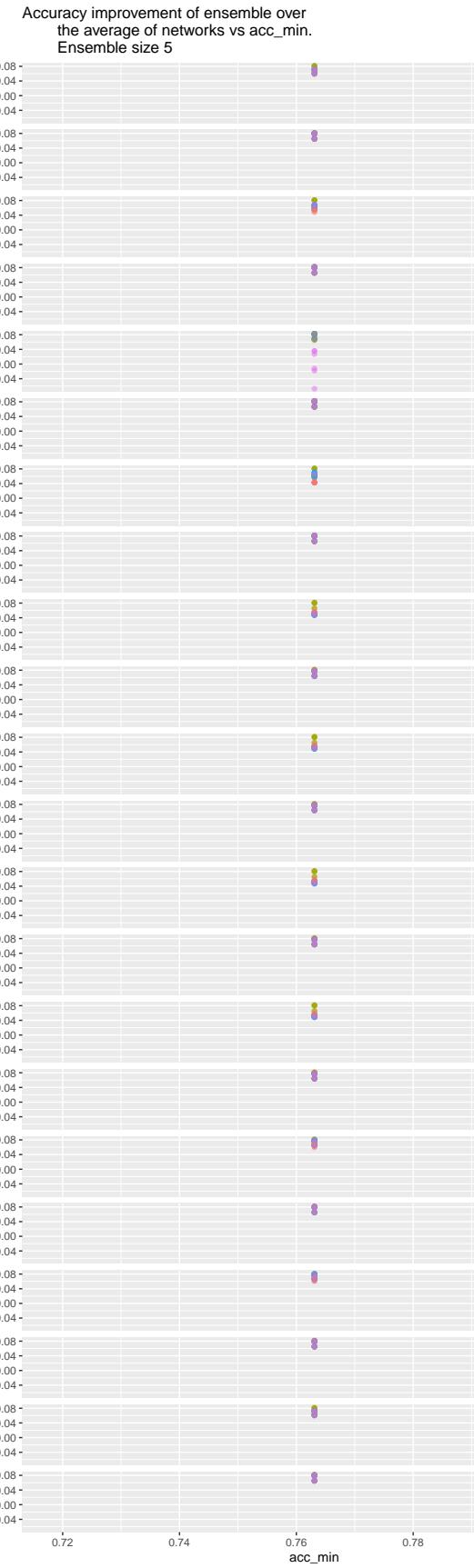
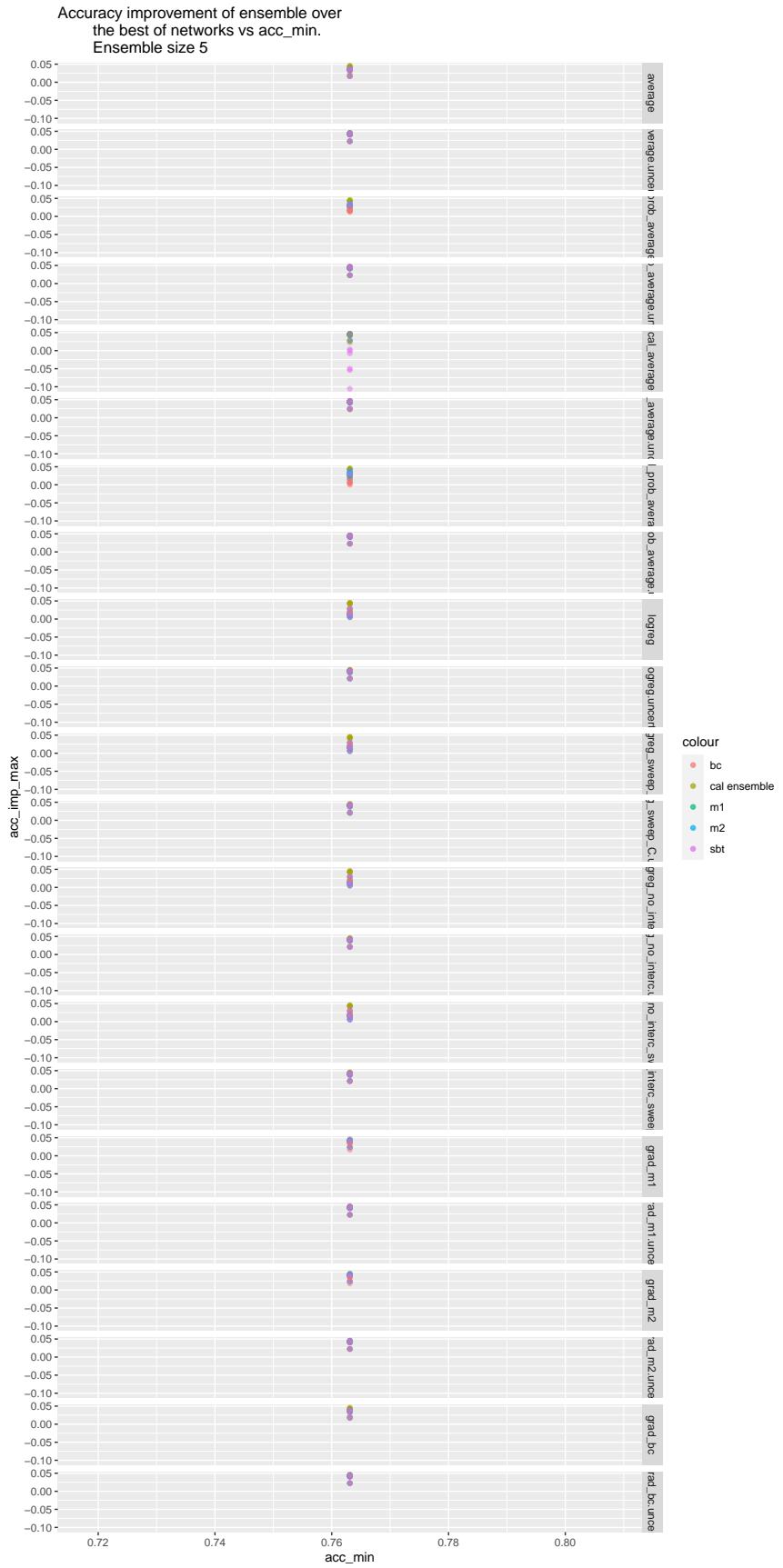


Accuracy improvement of ensemble over
the best of networks vs acc_var.
Ensemble size 5



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

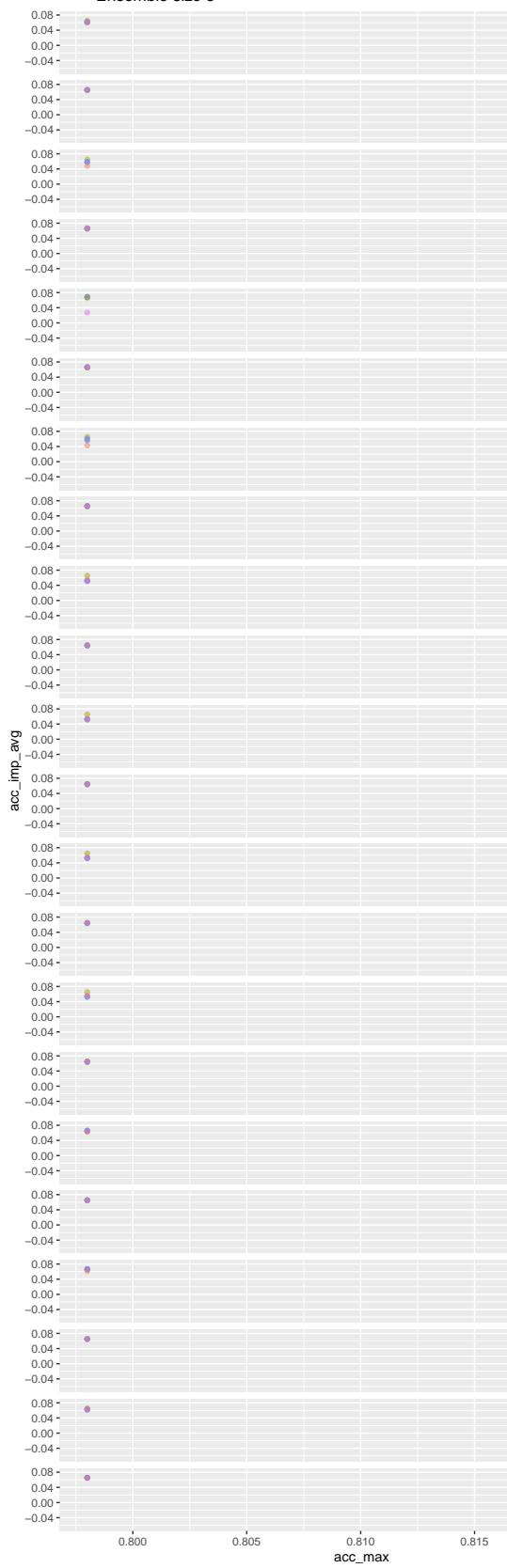




Accuracy improvement of ensemble over
the best of networks vs acc_max.
Ensemble size 5



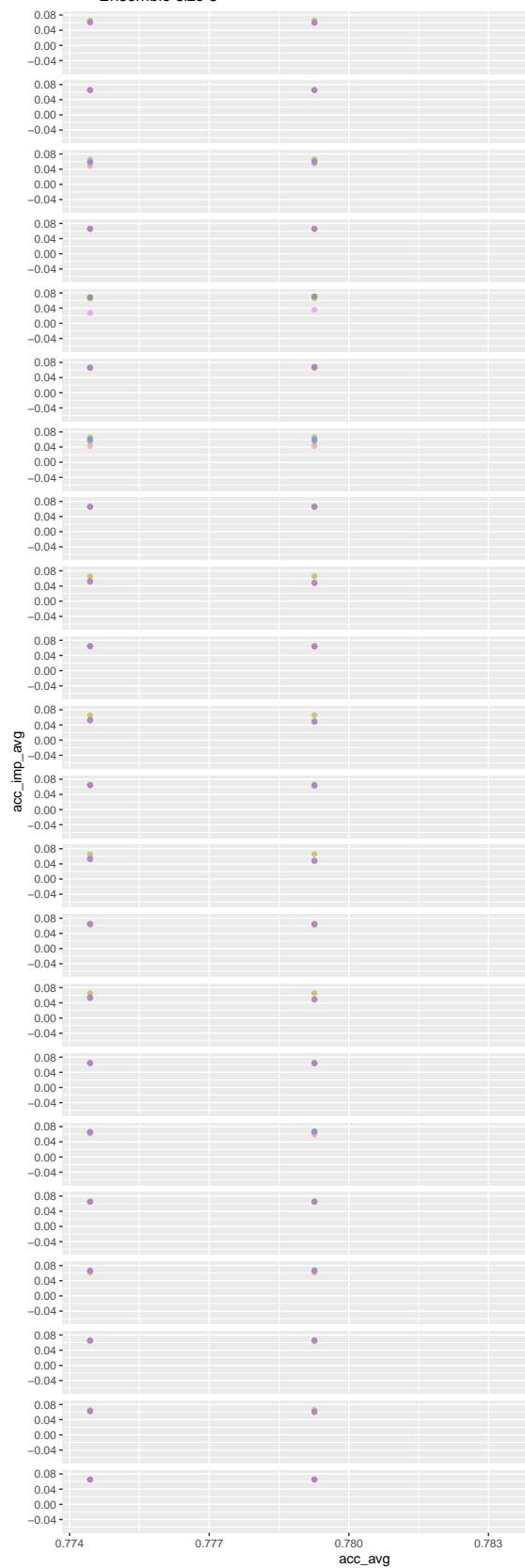
Accuracy improvement of ensemble over
the average of networks vs acc_max.
Ensemble size 5

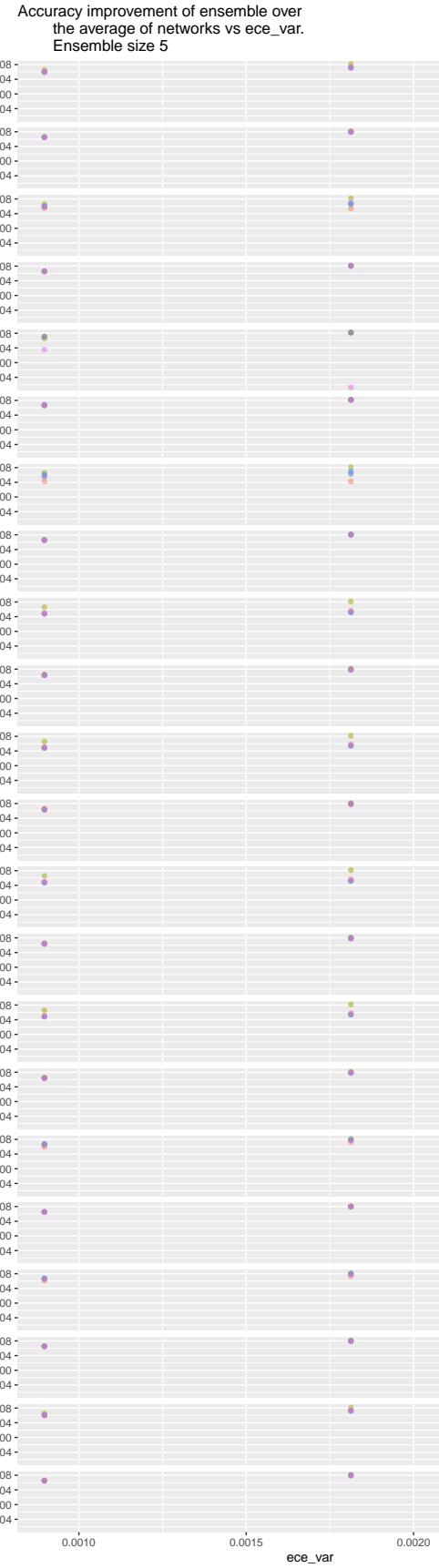
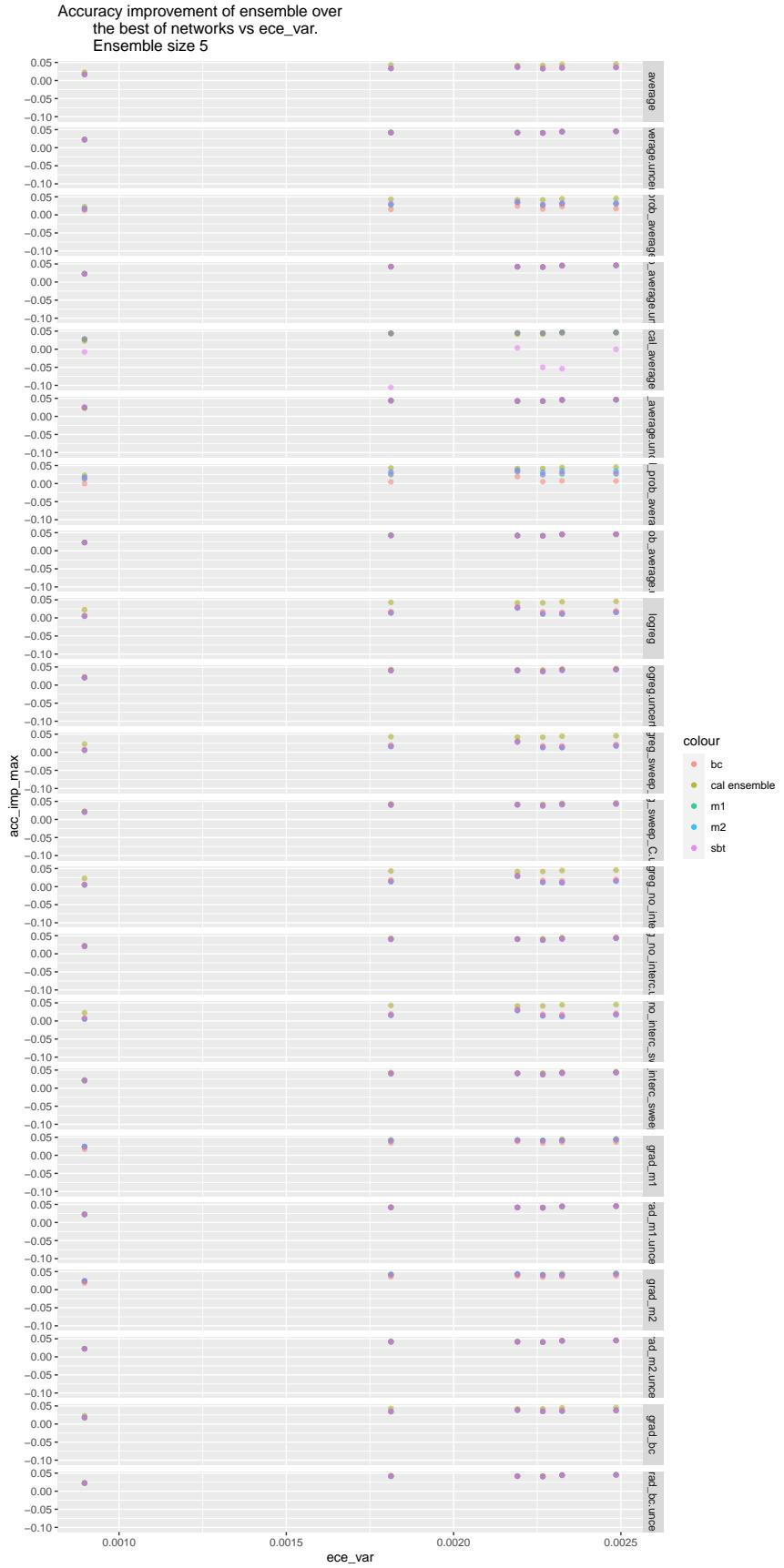


Accuracy improvement of ensemble over
the best of networks vs acc_avg.
Ensemble size 5

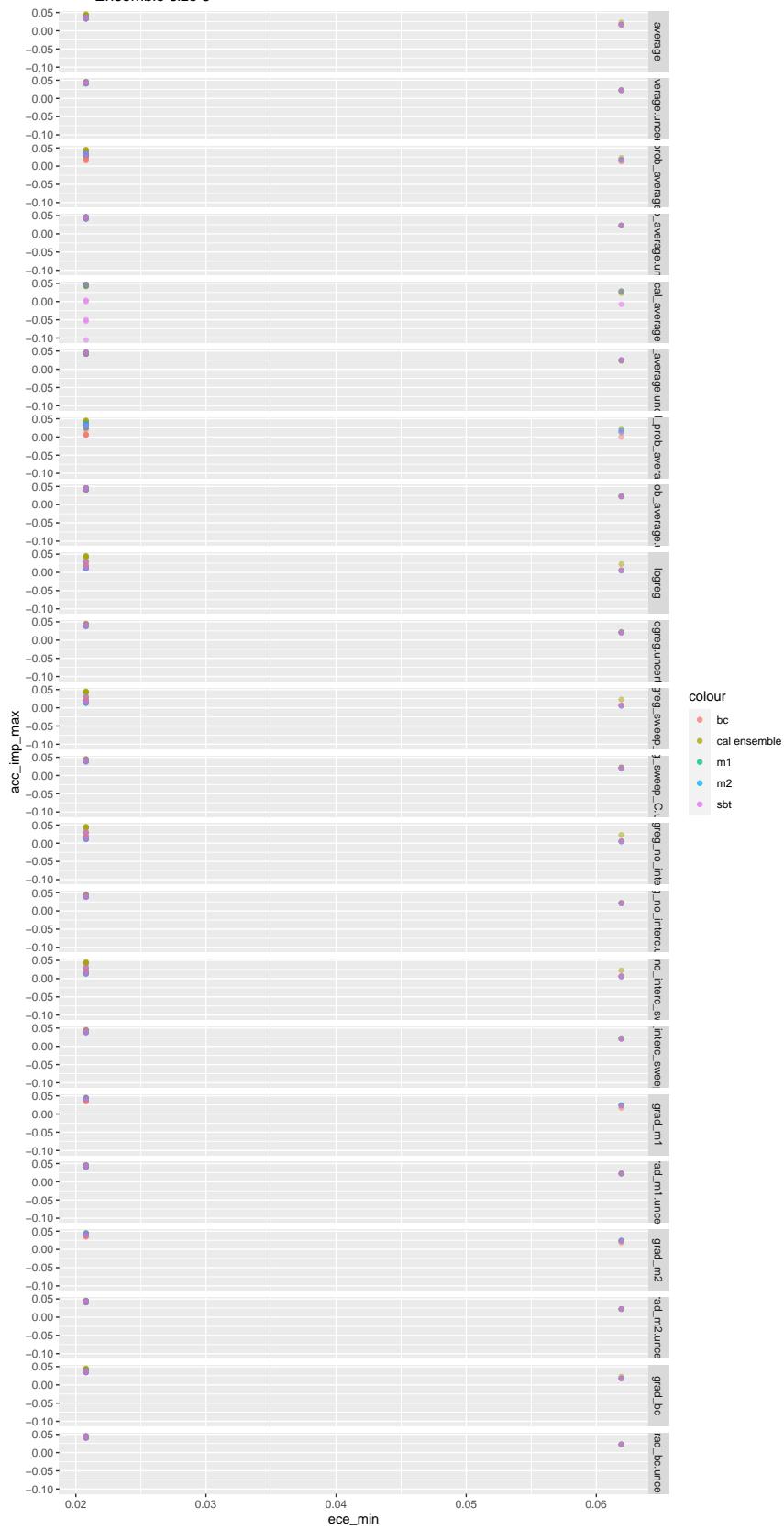


Accuracy improvement of ensemble over
the average of networks vs acc_avg.
Ensemble size 5

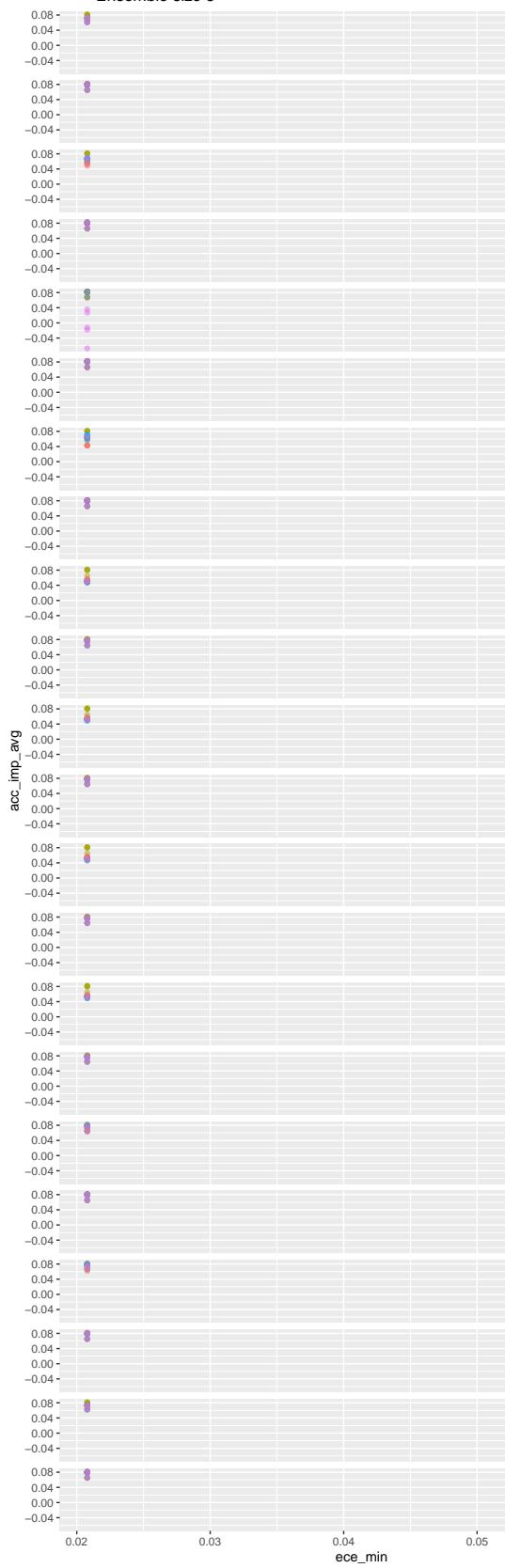




Accuracy improvement of ensemble over
the best of networks vs ece_min.
Ensemble size 5



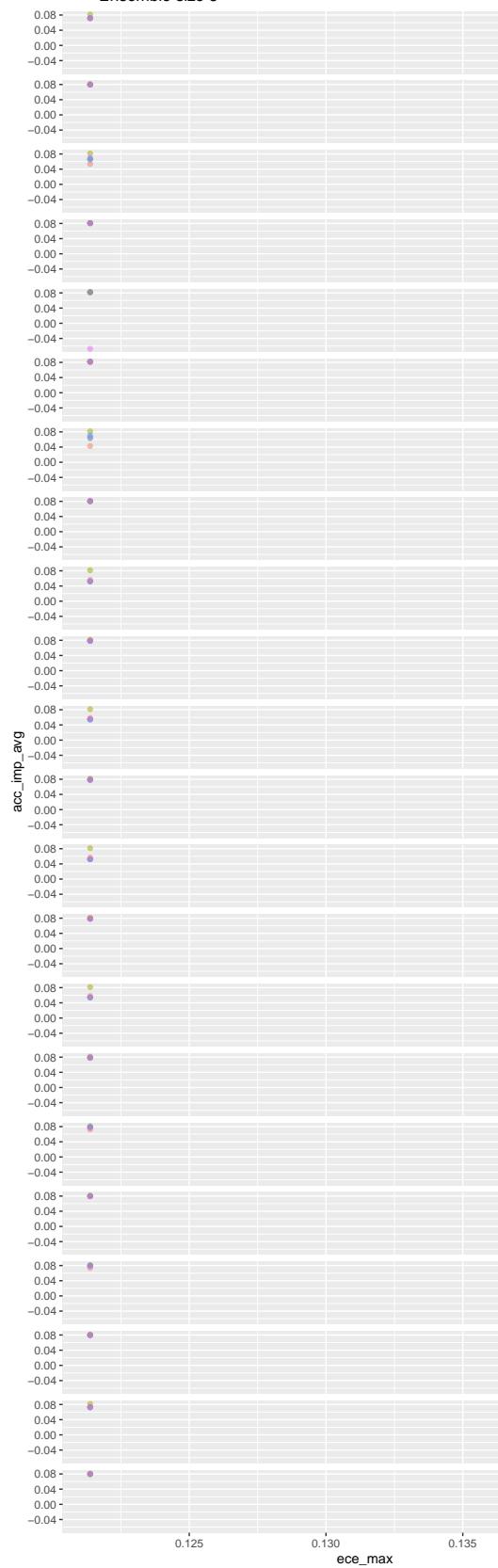
Accuracy improvement of ensemble over
the average of networks vs ece_min.
Ensemble size 5

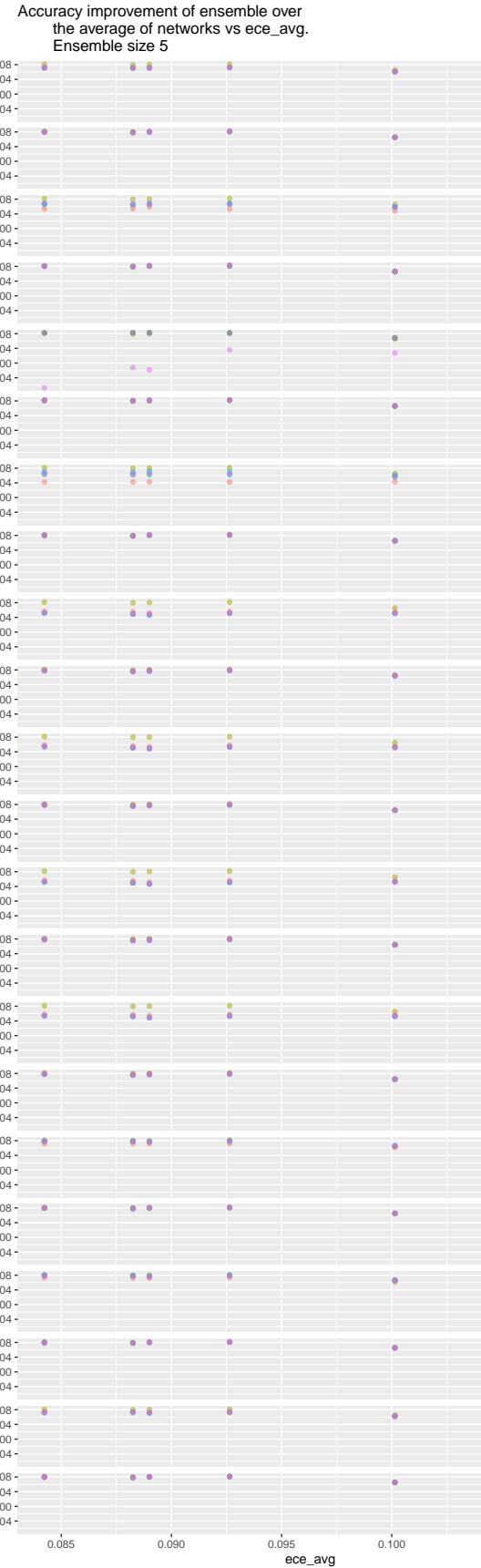
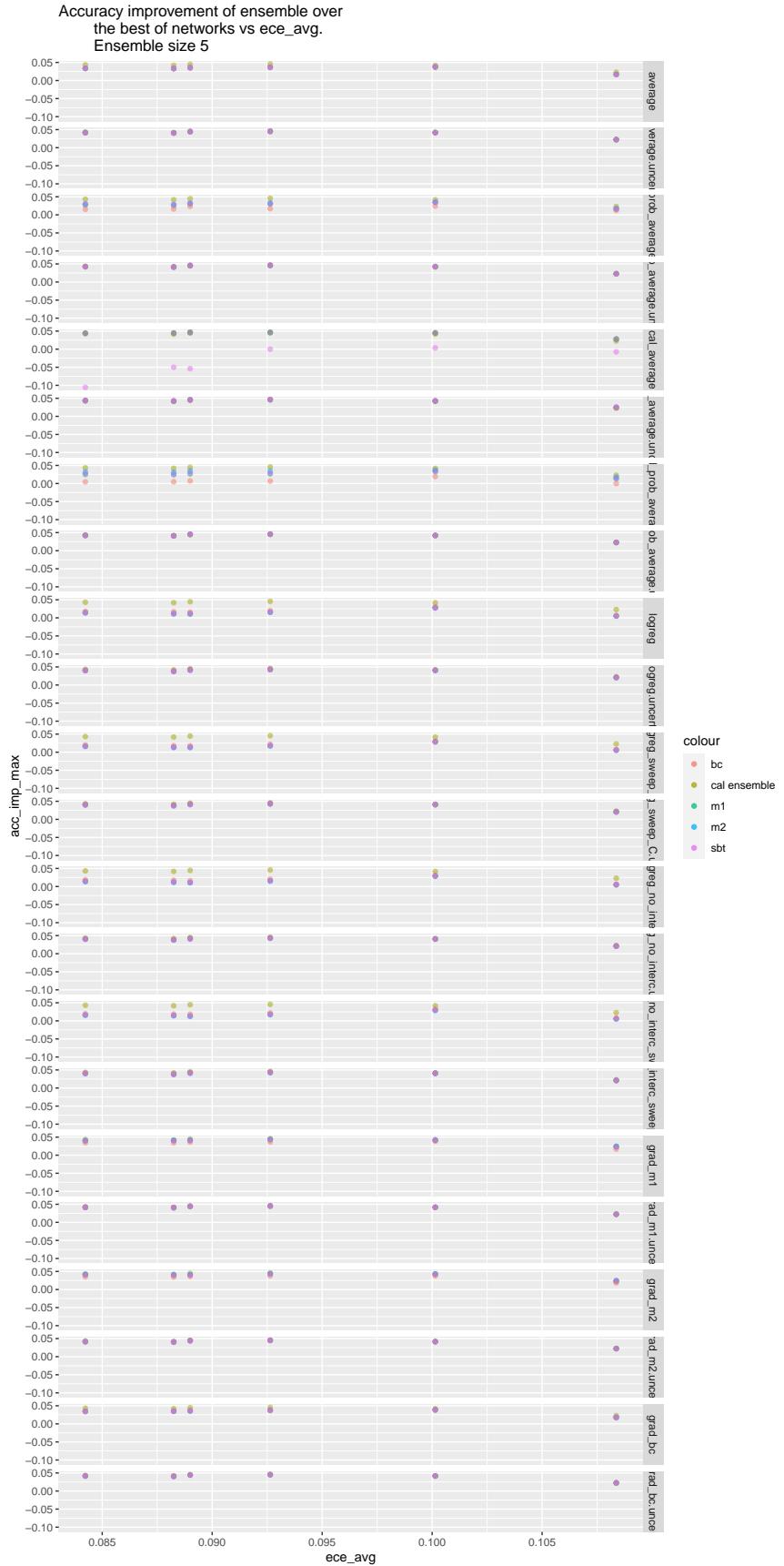


Accuracy improvement of ensemble over
the best of networks vs ece_max.
Ensemble size 5

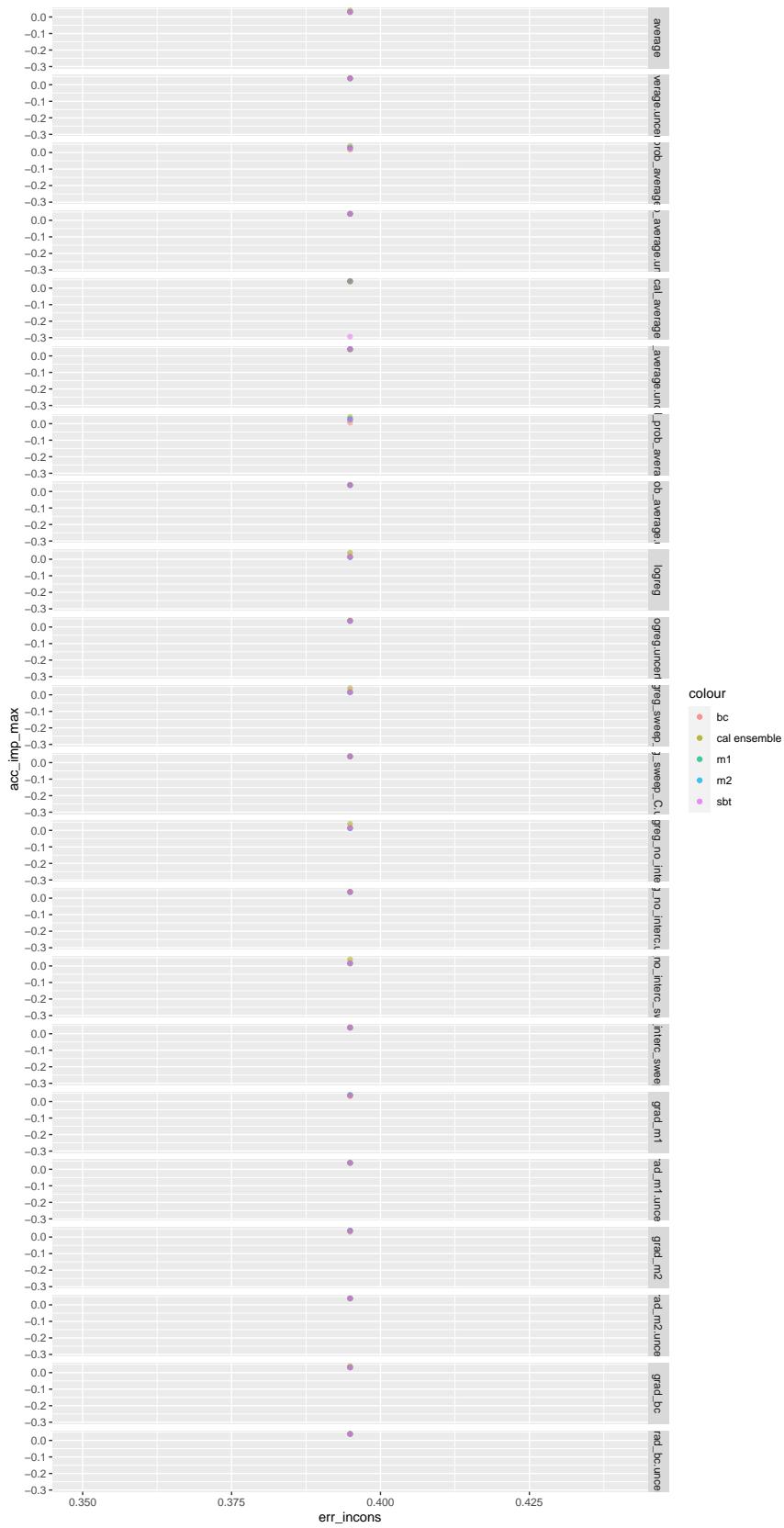


Accuracy improvement of ensemble over
the average of networks vs ece_max.
Ensemble size 5

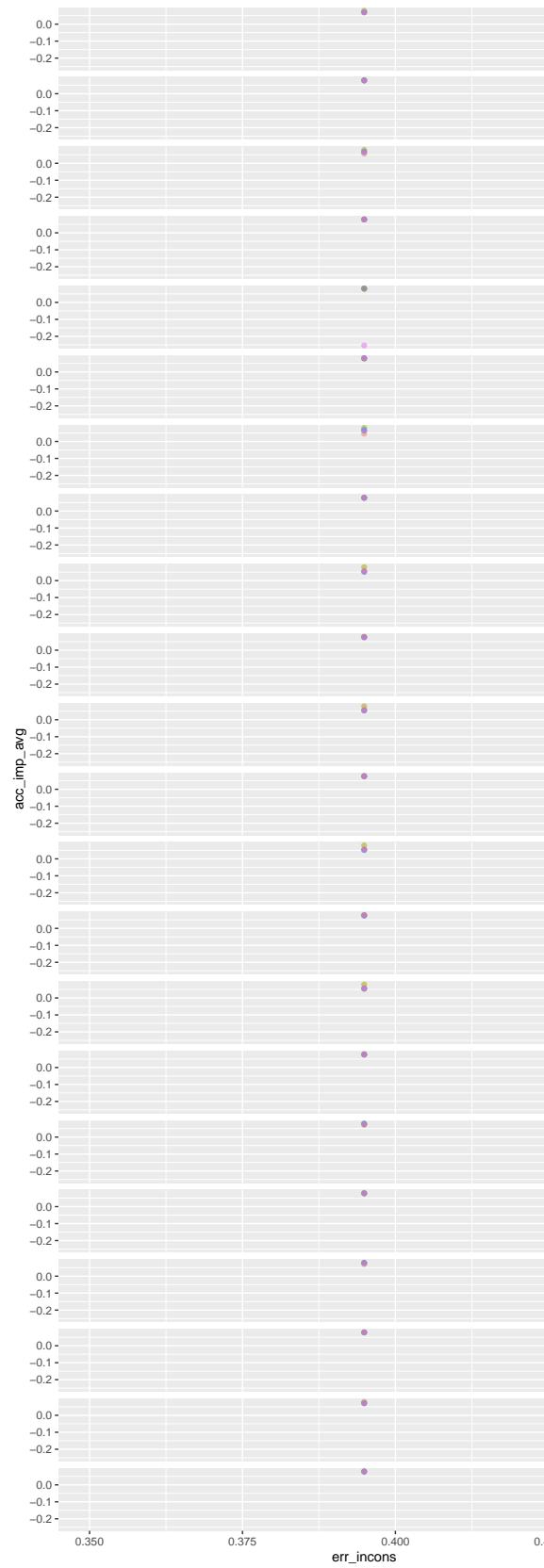




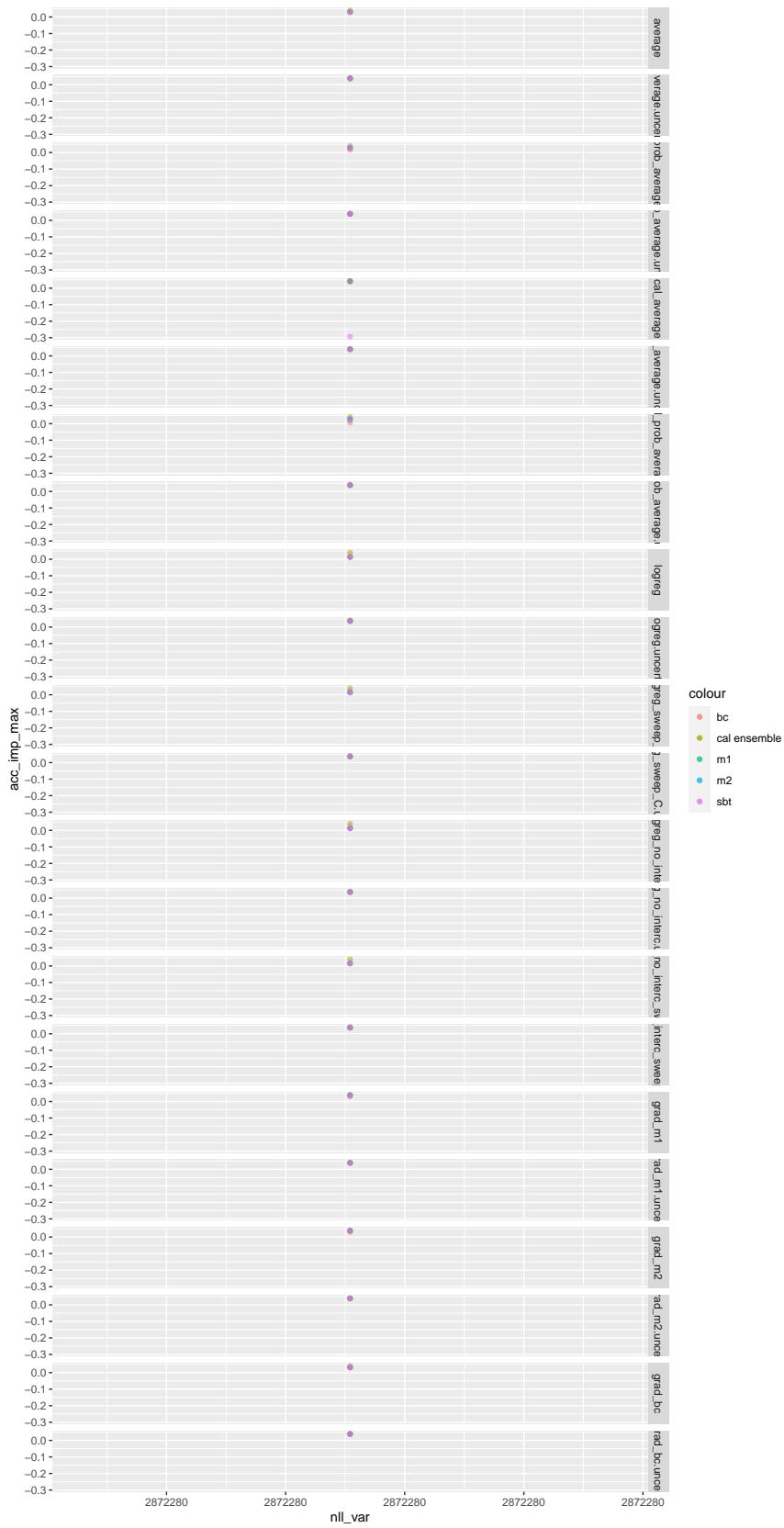
Accuracy improvement of ensemble over
the best of networks vs err_incons.
Ensemble size 6



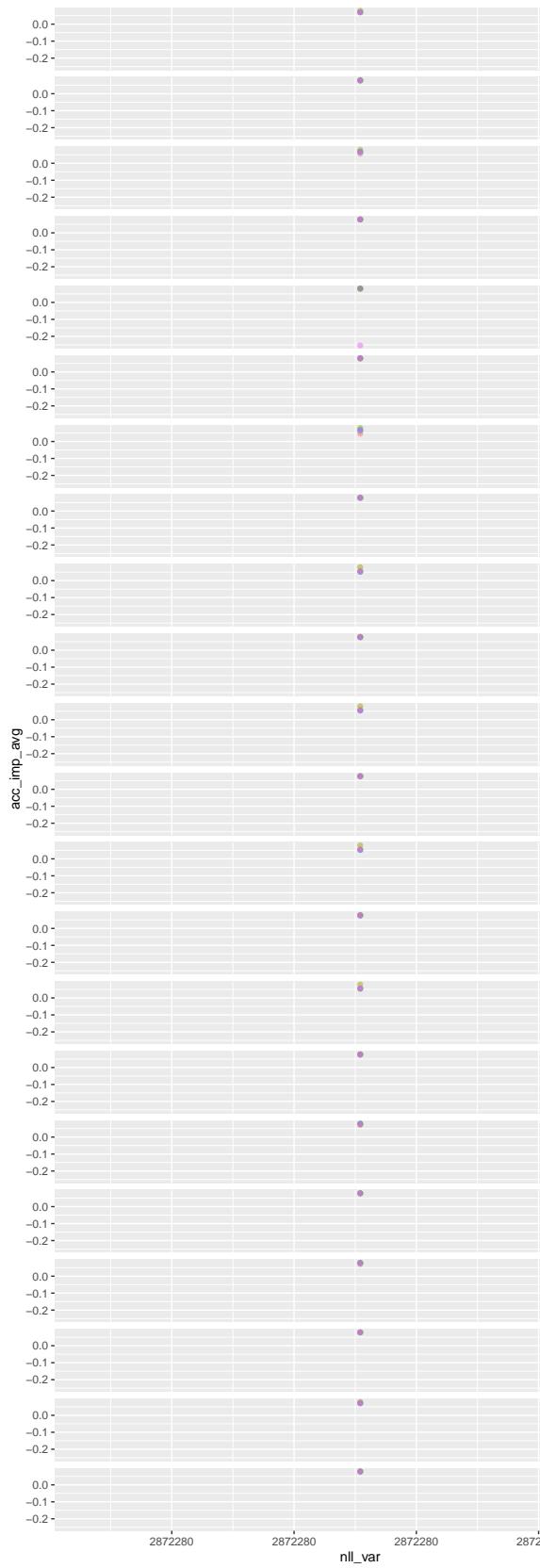
Accuracy improvement of ensemble over
the average of networks vs err_incons.
Ensemble size 6



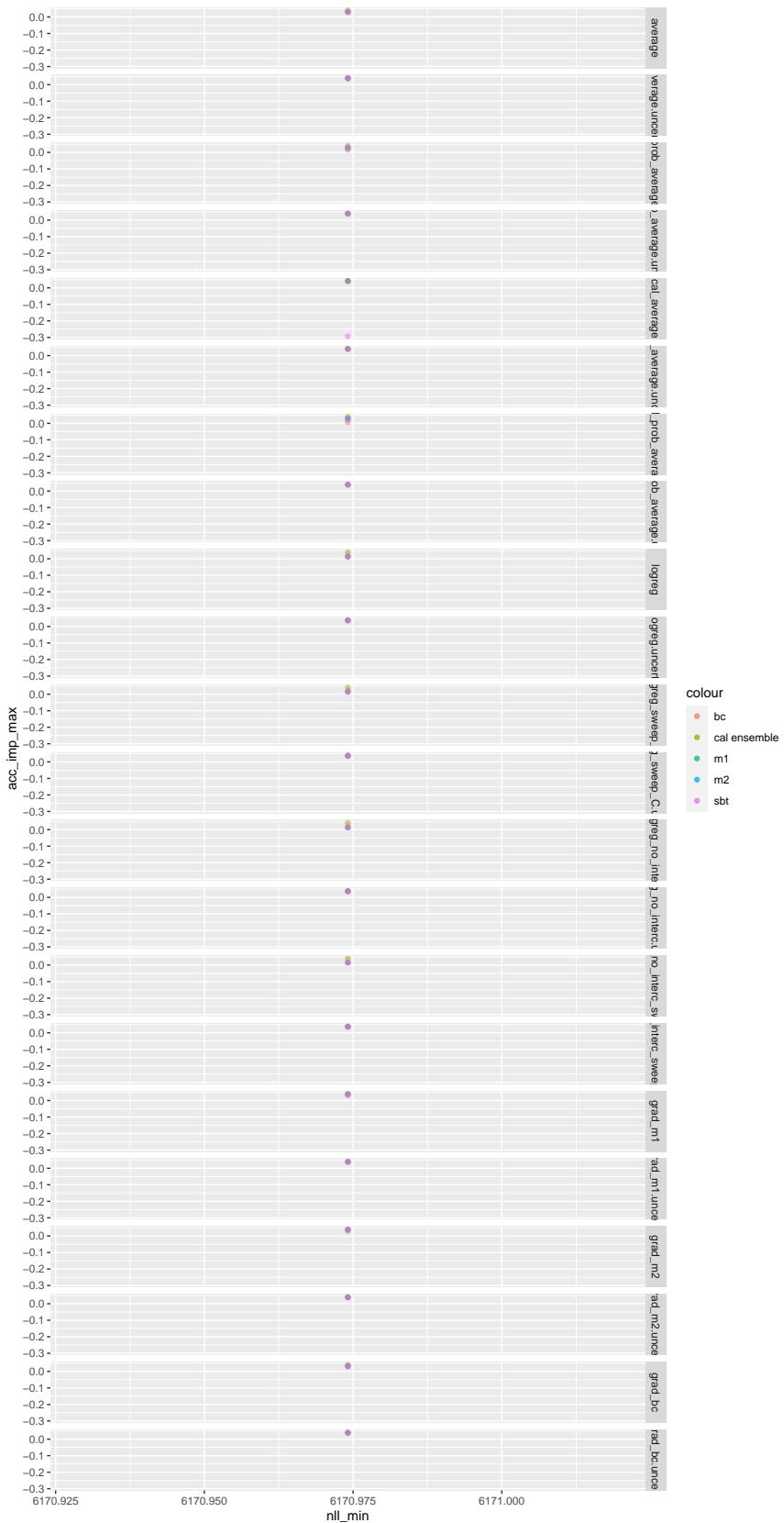
Accuracy improvement of ensemble over
the best of networks vs nll_var.
Ensemble size 6



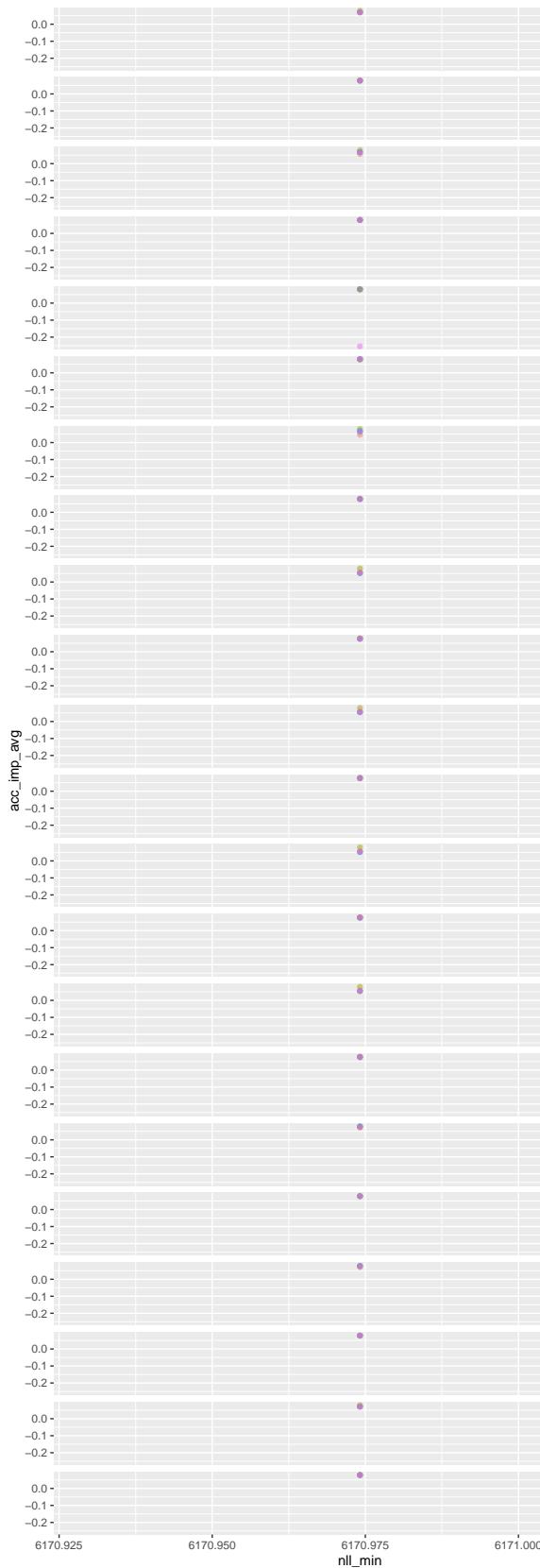
Accuracy improvement of ensemble over
the average of networks vs nll_var.
Ensemble size 6



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



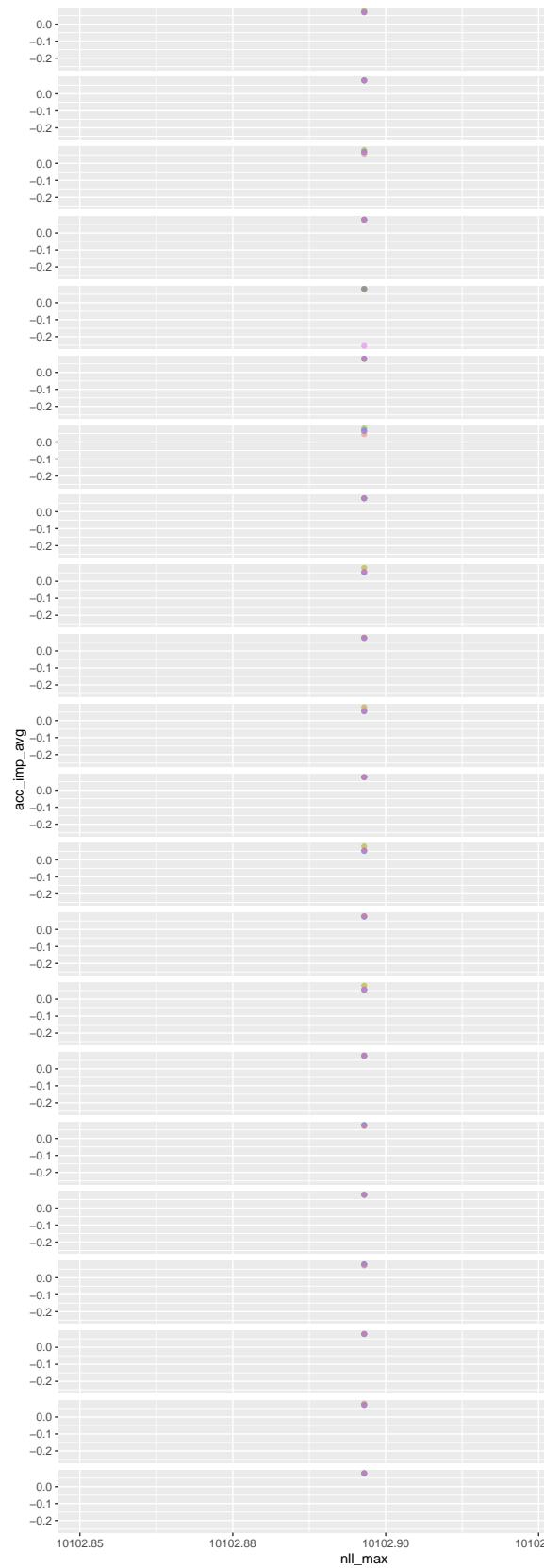
Accuracy improvement of ensemble over
the average of networks vs nll_min.
Ensemble size 6



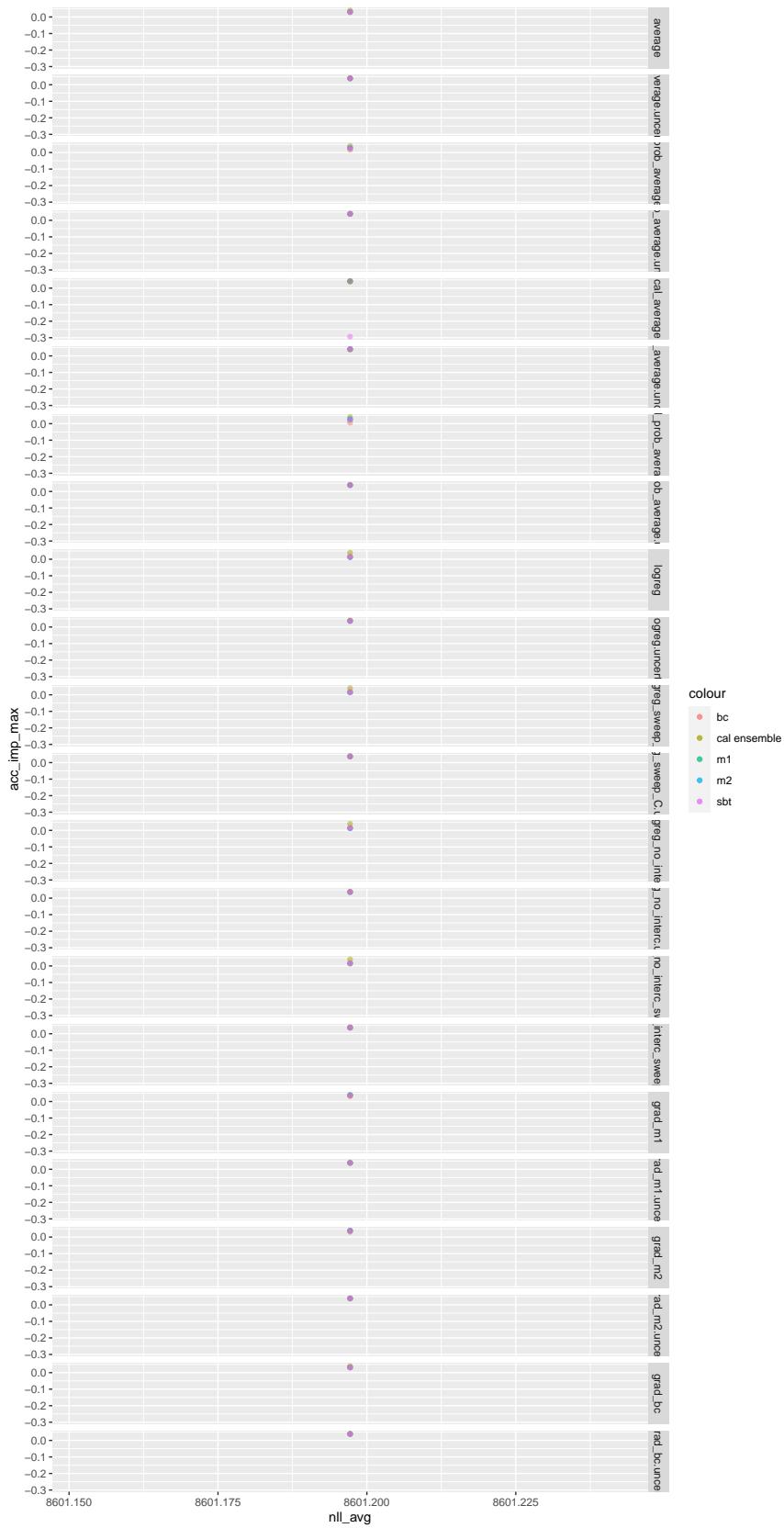
Accuracy improvement of ensemble over
the best of networks vs nll_max.
Ensemble size 6



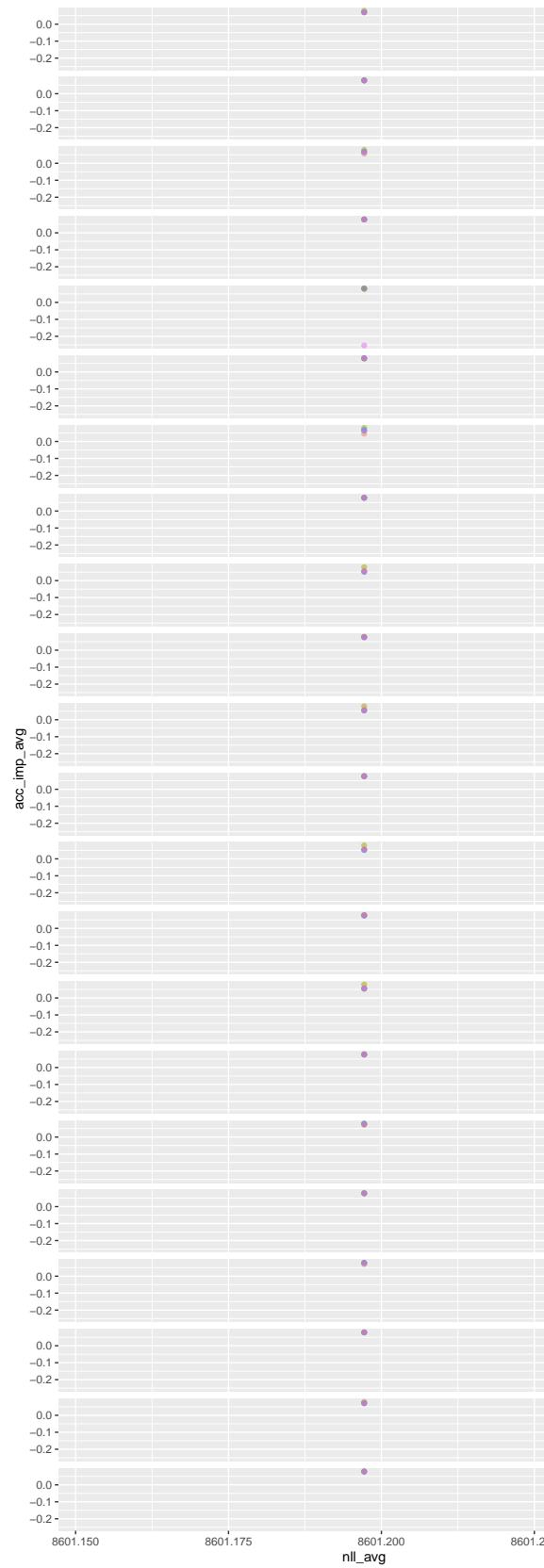
Accuracy improvement of ensemble over
the average of networks vs nll_max.
Ensemble size 6



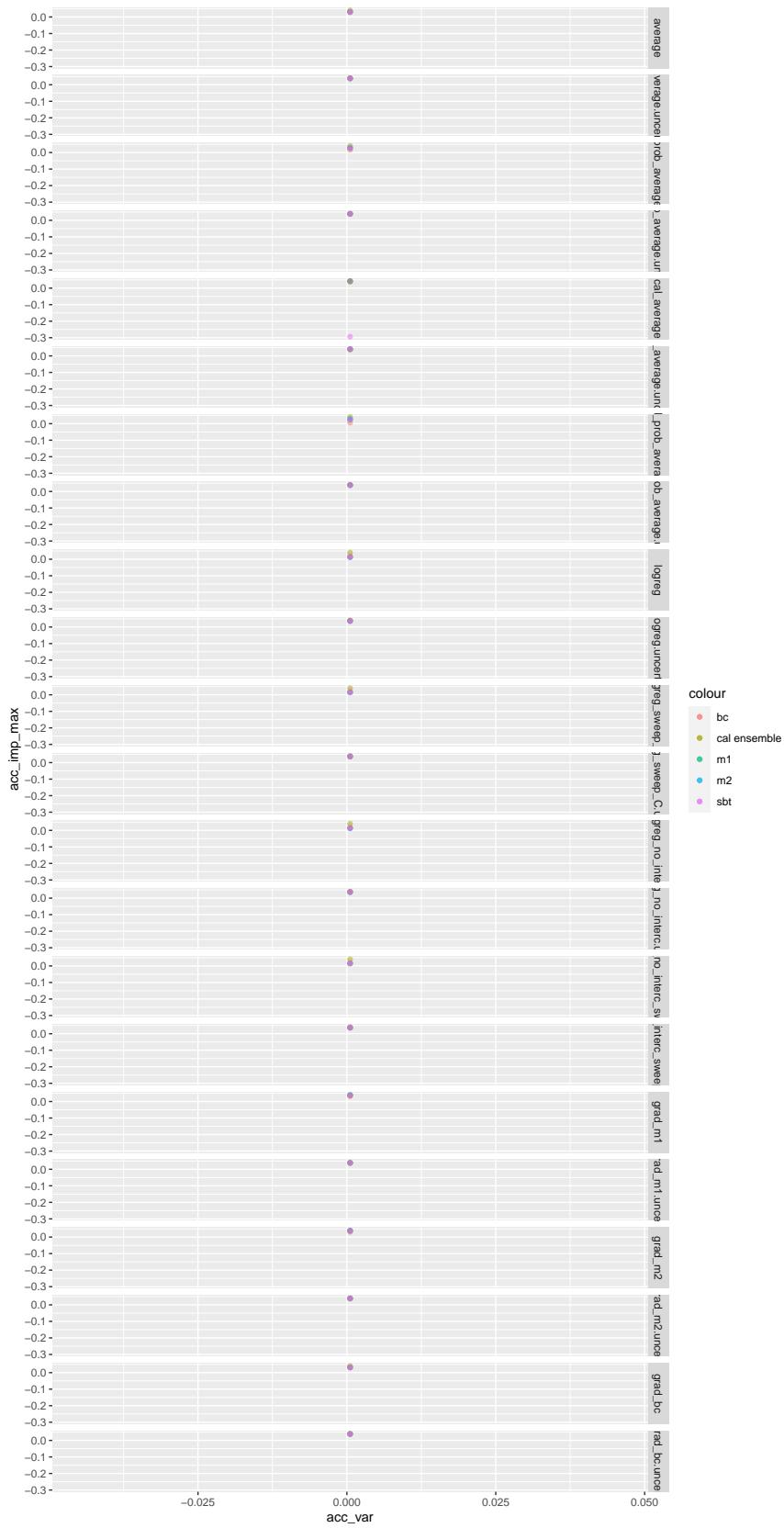
Accuracy improvement of ensemble over
the best of networks vs nll_avg.
Ensemble size 6



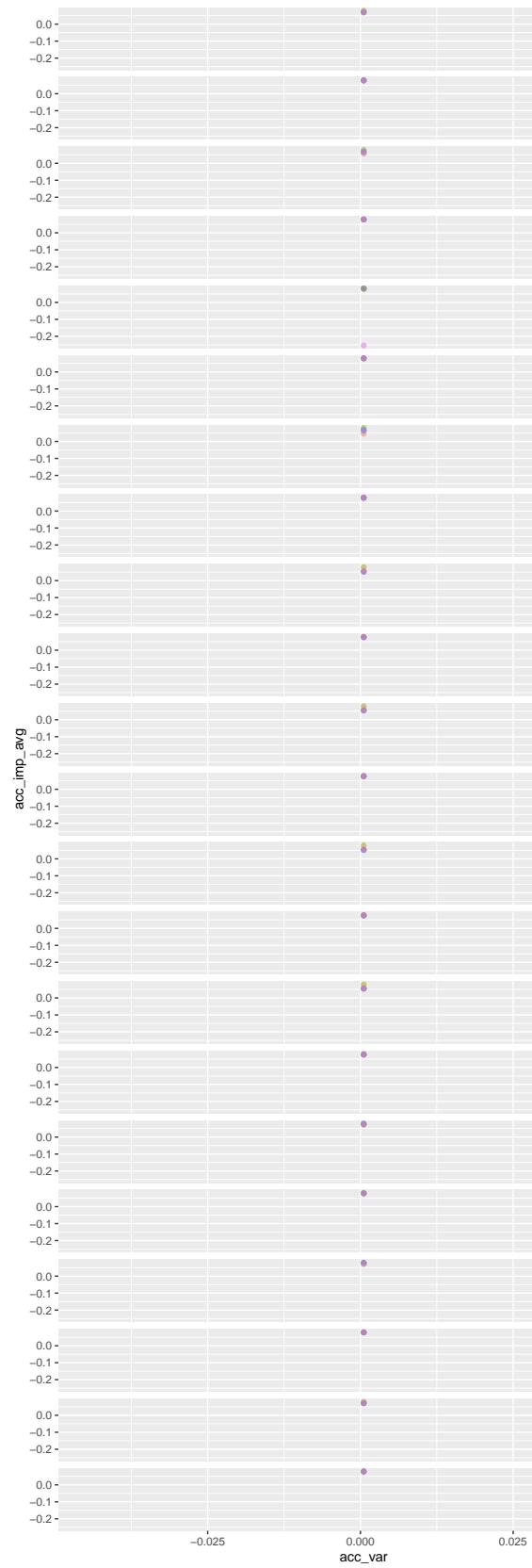
Accuracy improvement of ensemble over
the average of networks vs nll_avg.
Ensemble size 6



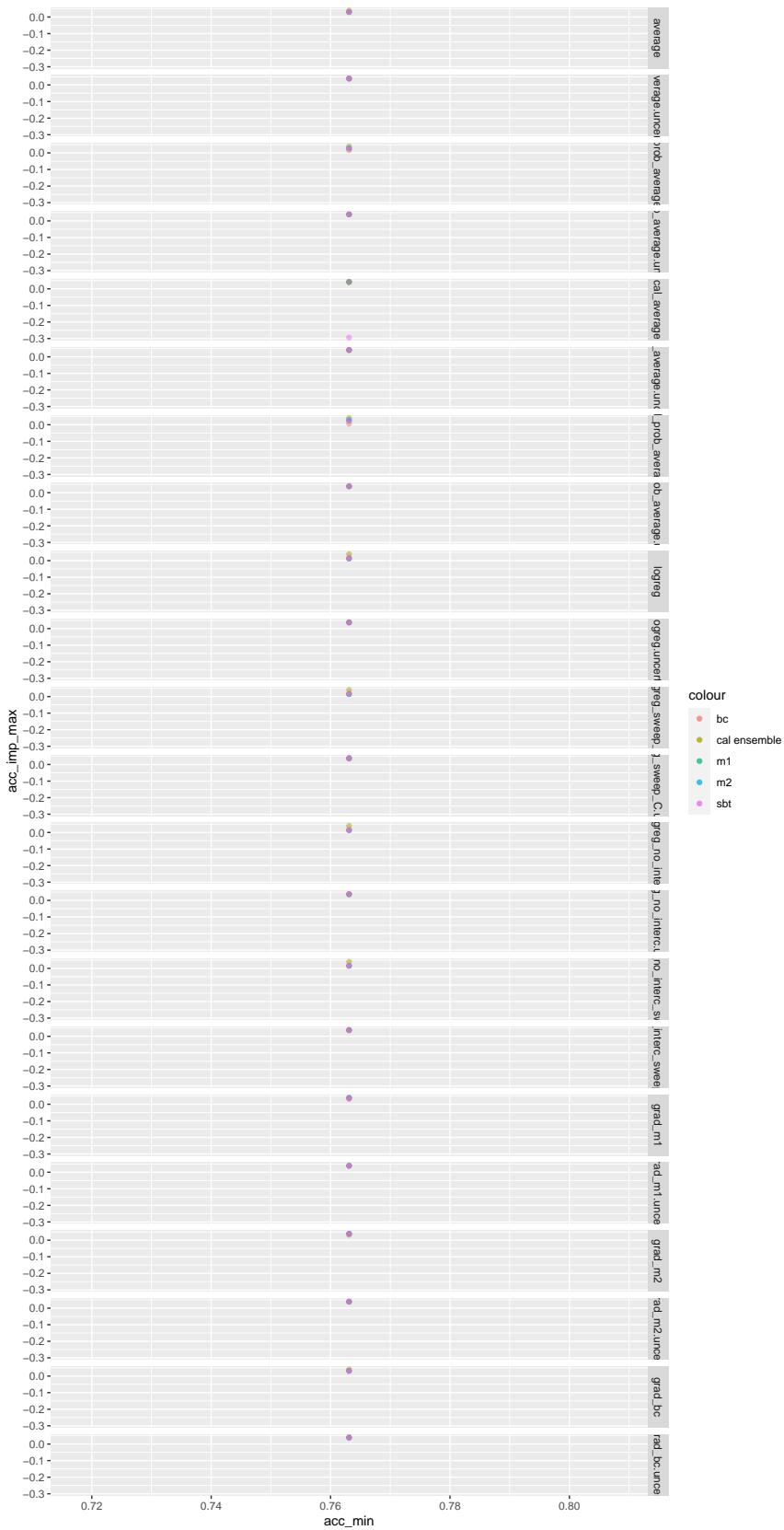
Accuracy improvement of ensemble over
the best of networks vs acc_var.
Ensemble size 6



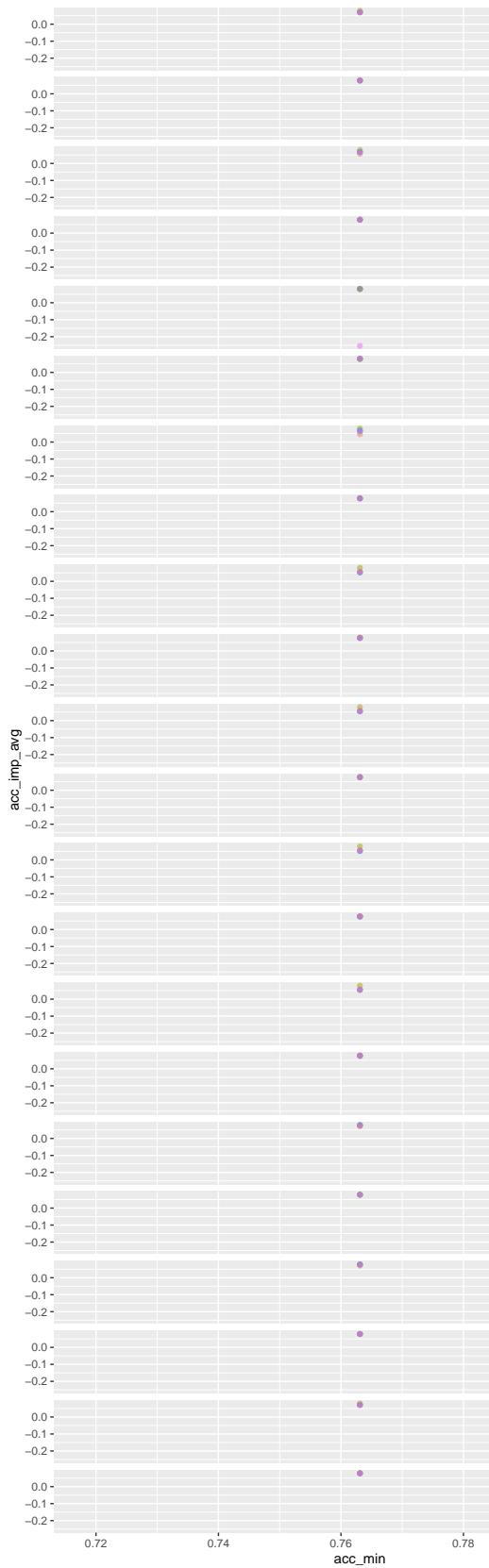
Accuracy improvement of ensemble over
the average of networks vs acc_var.
Ensemble size 6



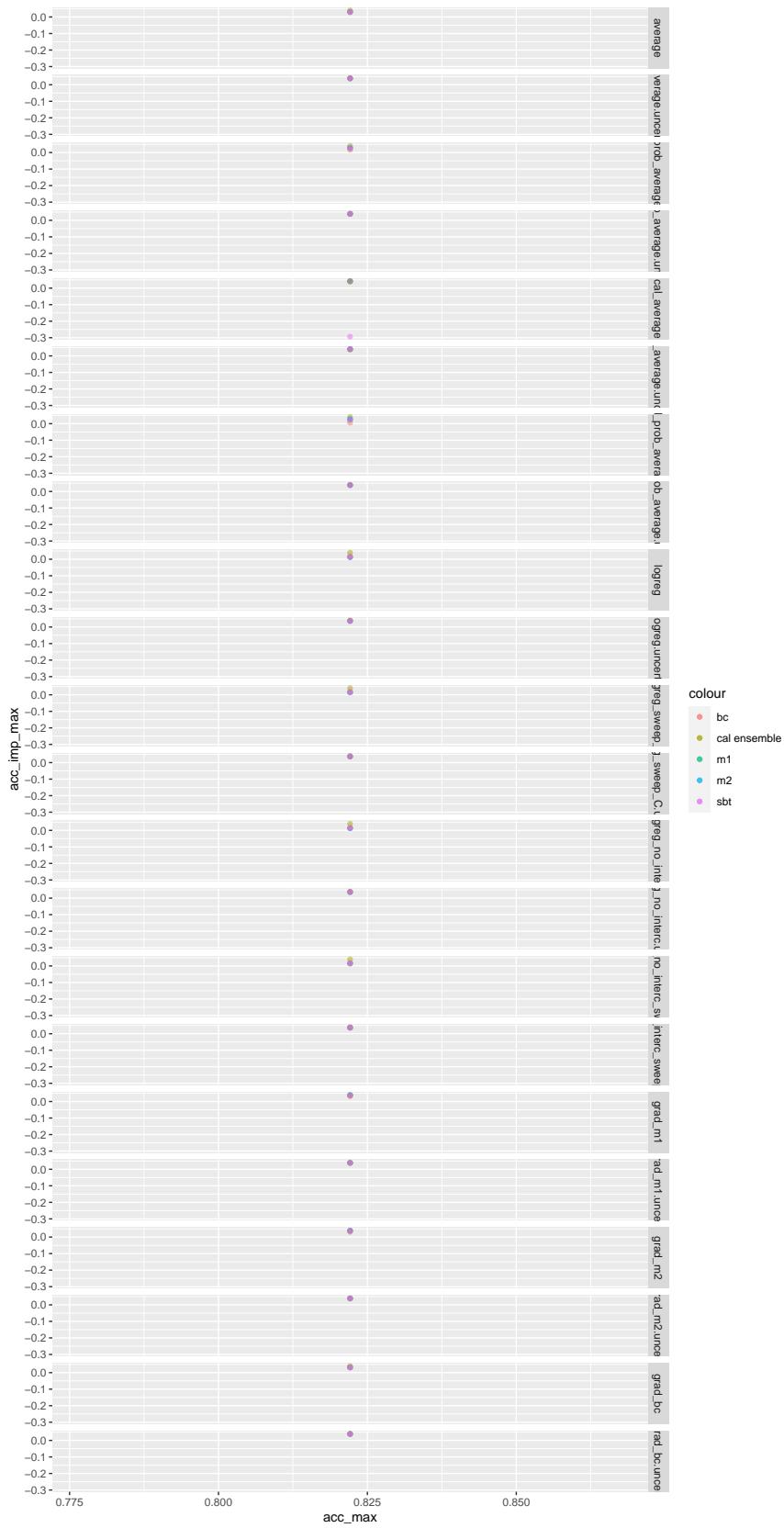
Accuracy improvement of ensemble over
the best of networks vs acc_min.
Ensemble size 6



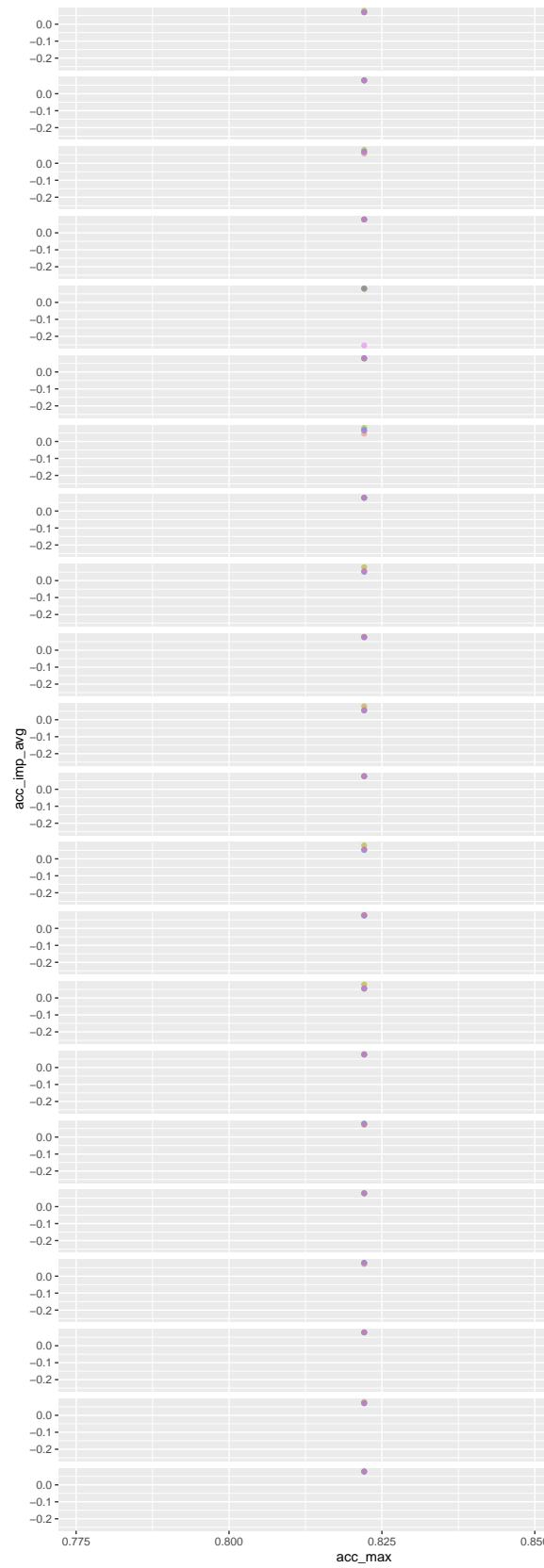
Accuracy improvement of ensemble over
the average of networks vs acc_min.
Ensemble size 6



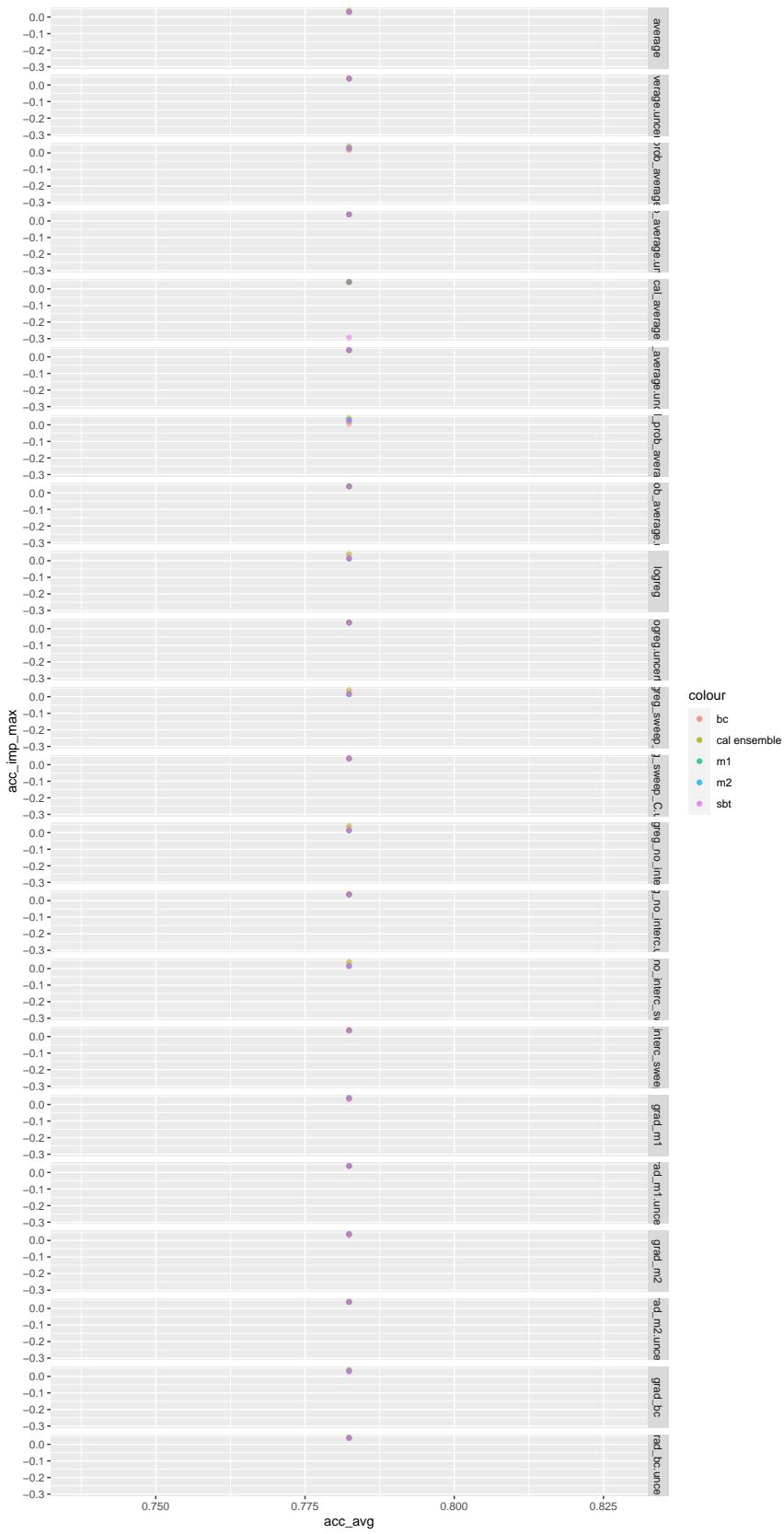
Accuracy improvement of ensemble over
the best of networks vs acc_max.
Ensemble size 6



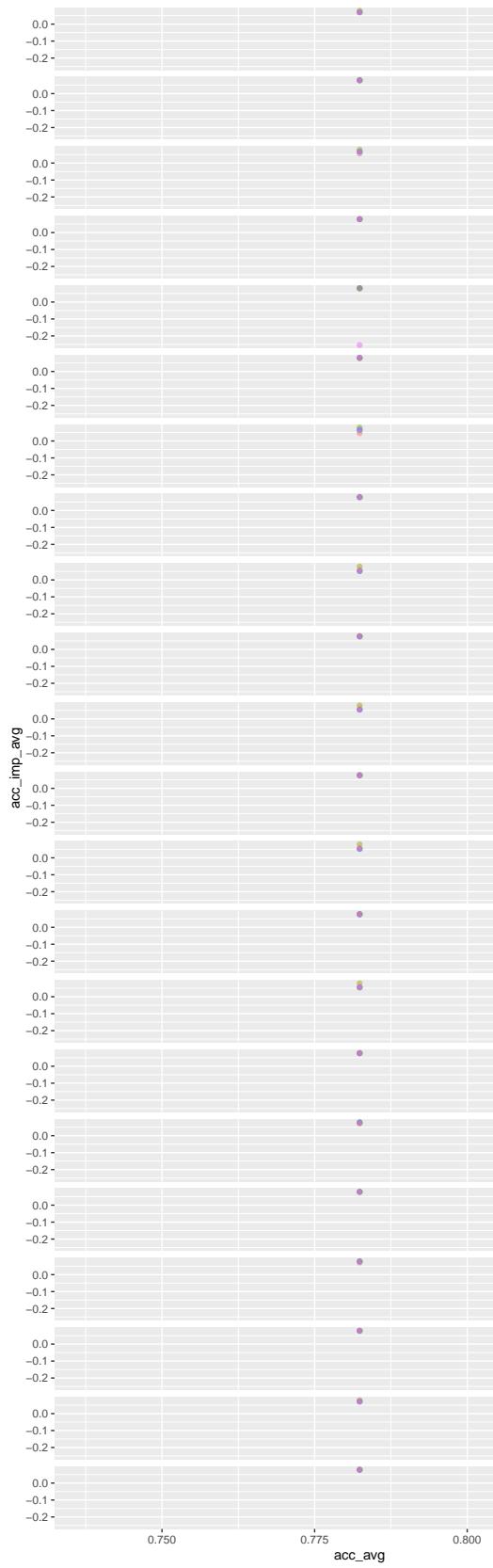
Accuracy improvement of ensemble over
the average of networks vs acc_max.
Ensemble size 6



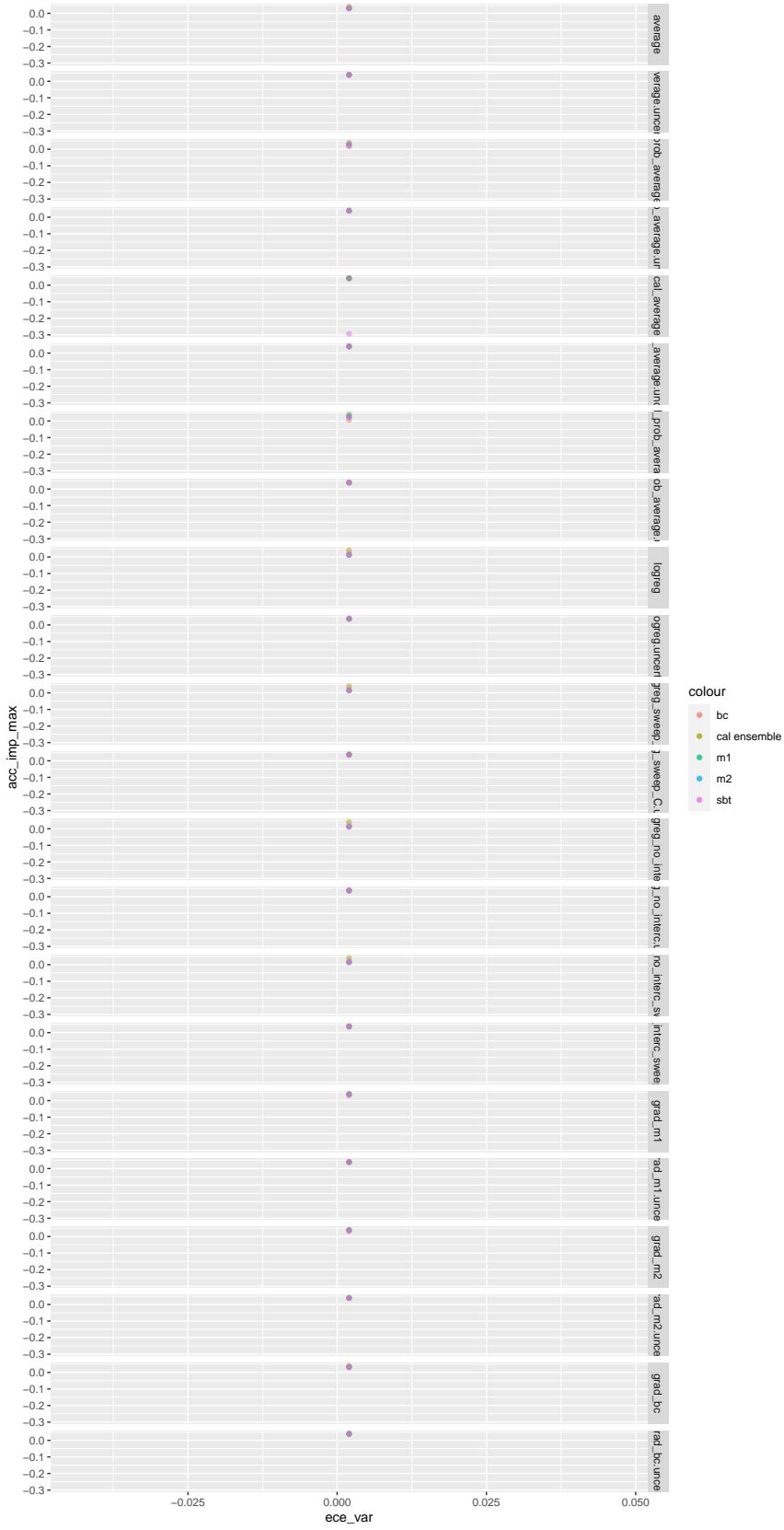
Accuracy improvement of ensemble over
the best of networks vs acc_avg.
Ensemble size 6



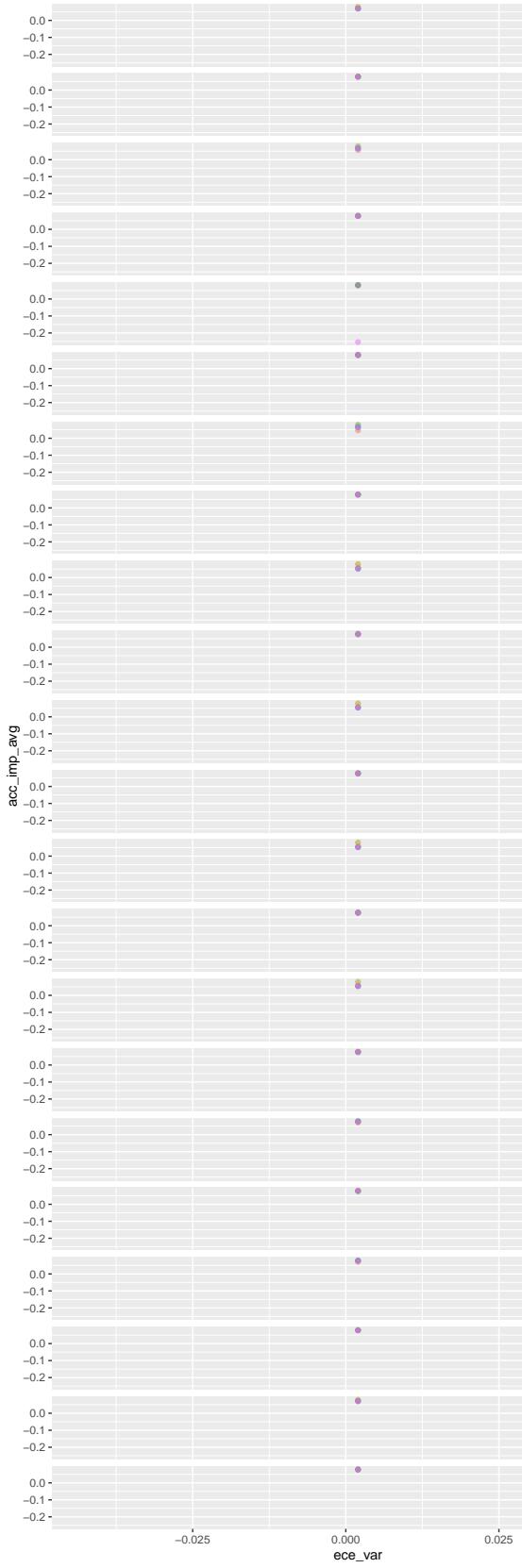
Accuracy improvement of ensemble over
the average of networks vs acc_avg.
Ensemble size 6



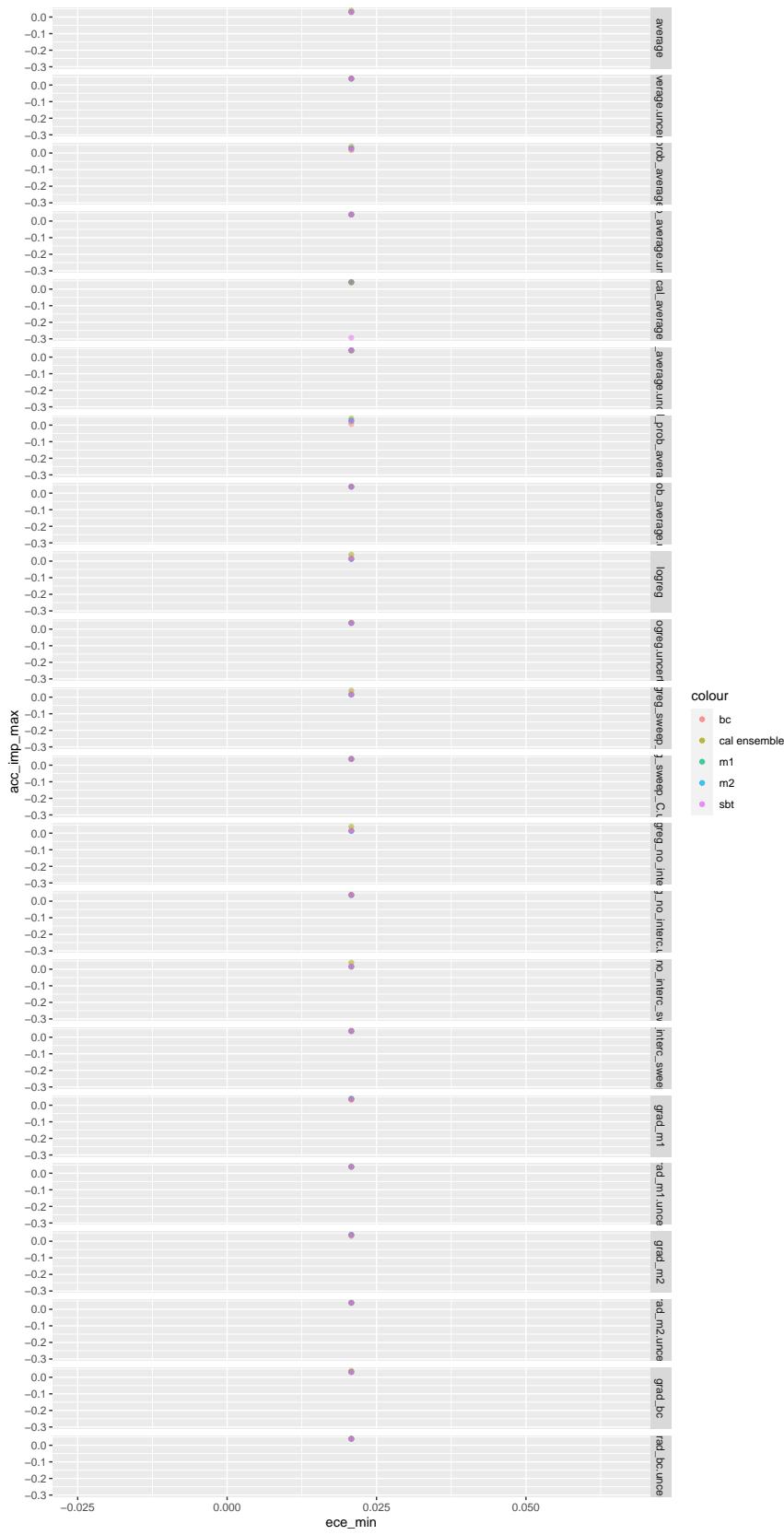
Accuracy improvement of ensemble over
the best of networks vs ece_var.
Ensemble size 6



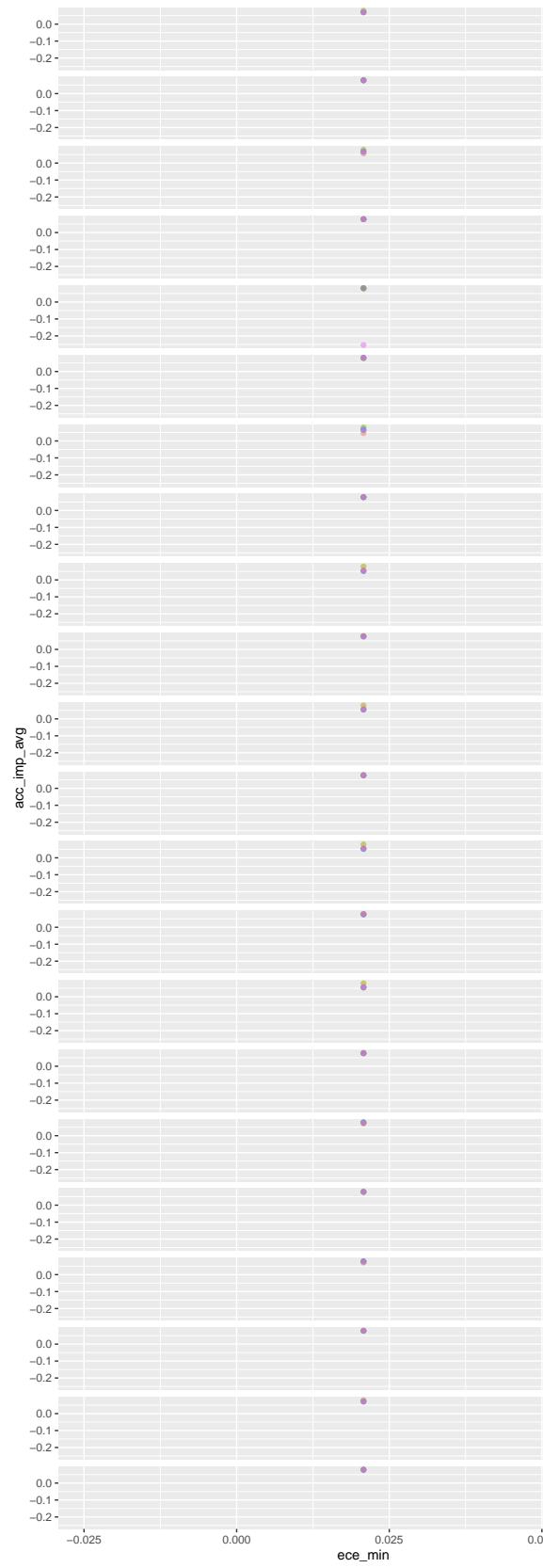
Accuracy improvement of ensemble over
the average of networks vs ece_var.
Ensemble size 6



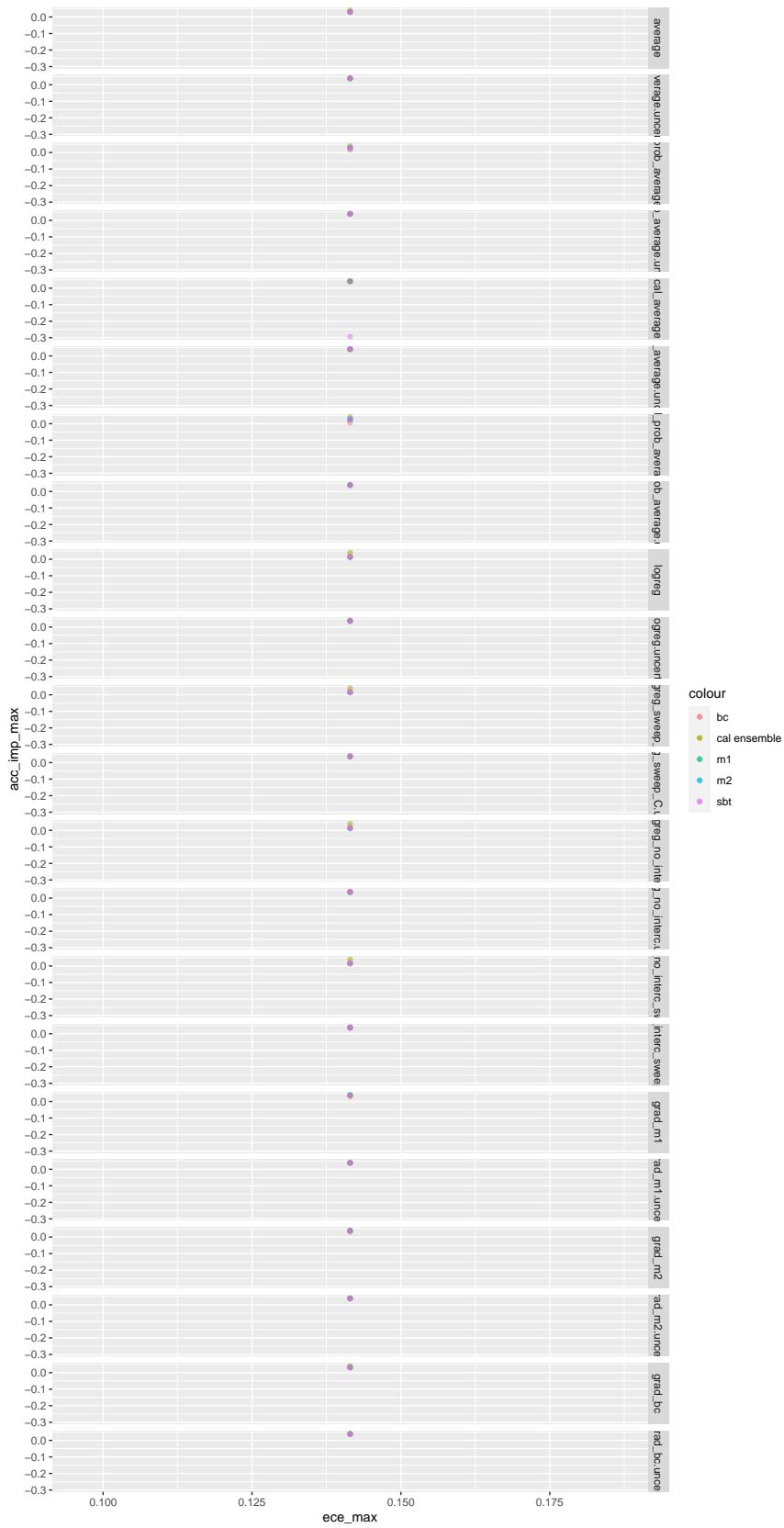
Accuracy improvement of ensemble over
the best of networks vs ece_min.
Ensemble size 6



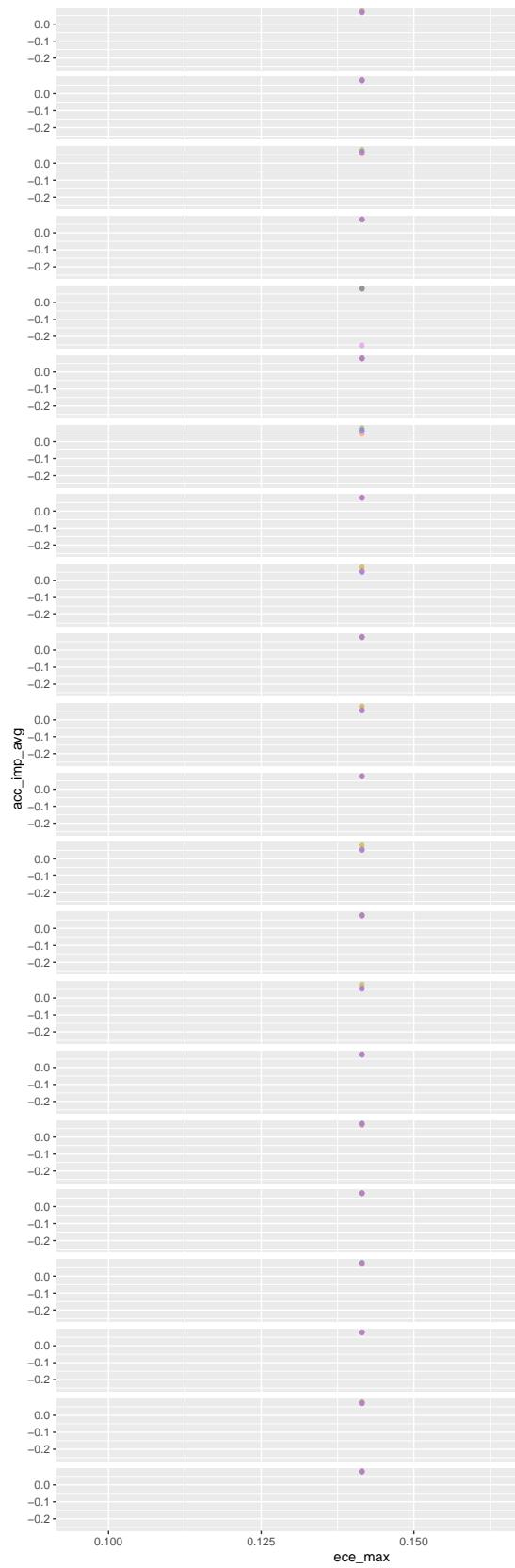
Accuracy improvement of ensemble over
the average of networks vs ece_min.
Ensemble size 6



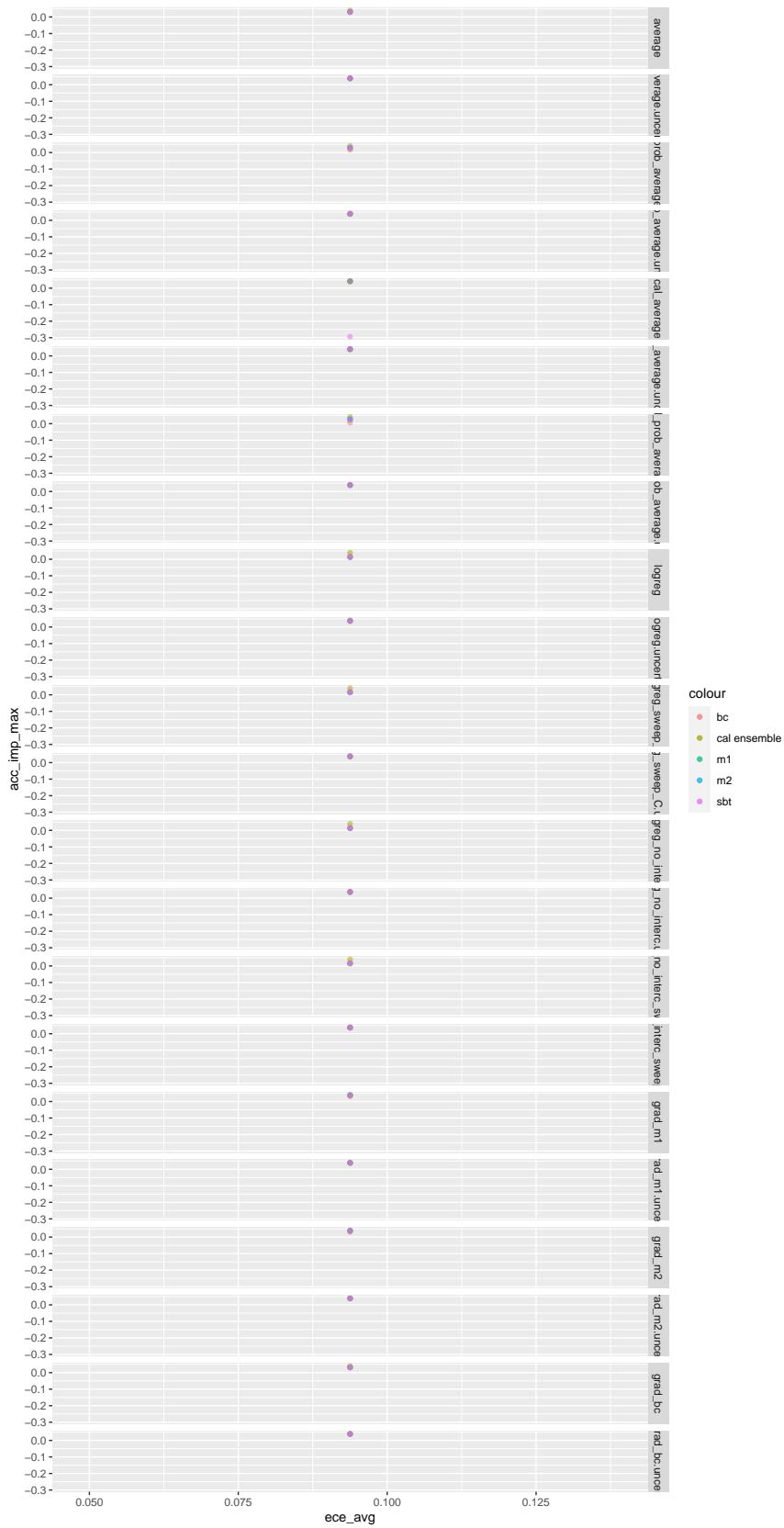
Accuracy improvement of ensemble over
the best of networks vs ece_max.
Ensemble size 6



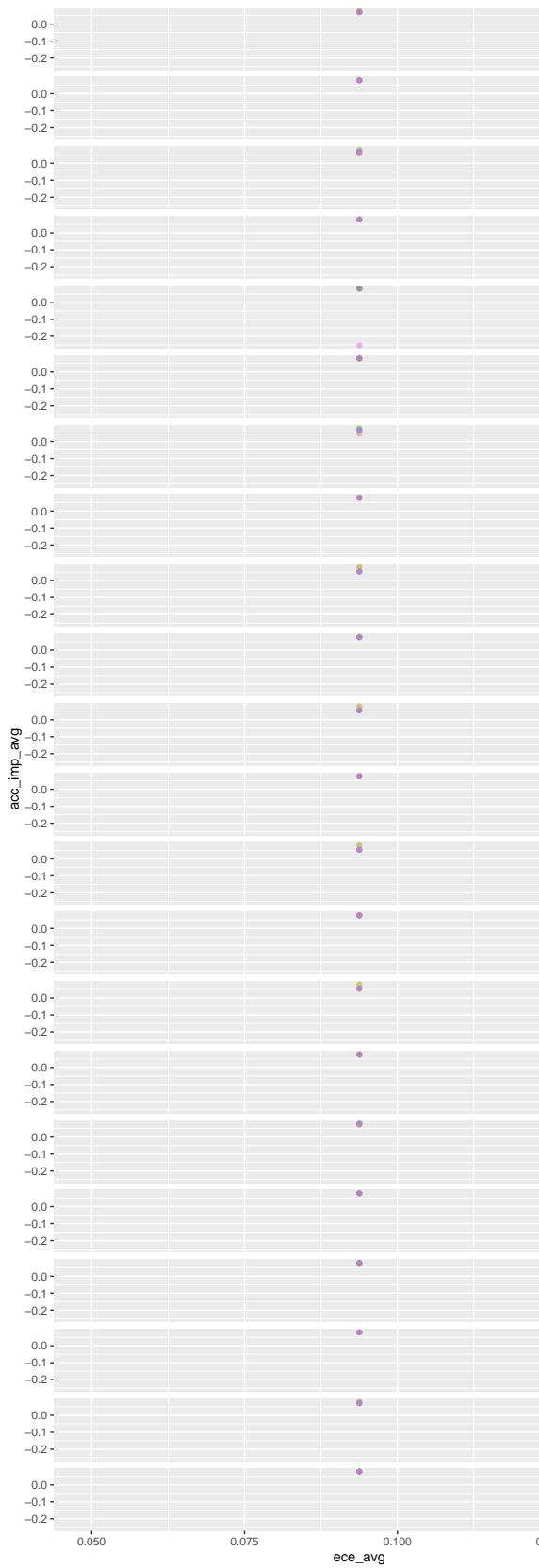
Accuracy improvement of ensemble over
the average of networks vs ece_max.
Ensemble size 6



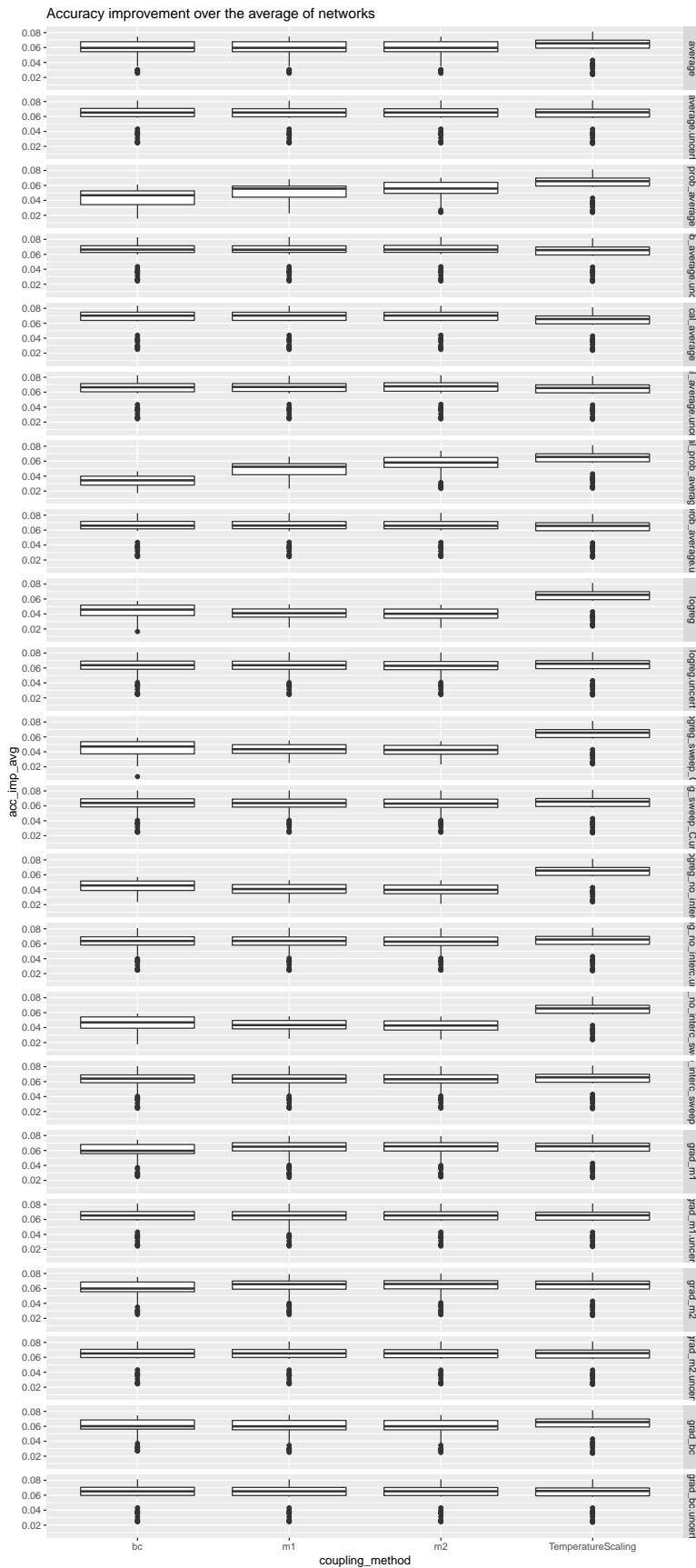
Accuracy improvement of ensemble over
the best of networks vs ece_avg.
Ensemble size 6



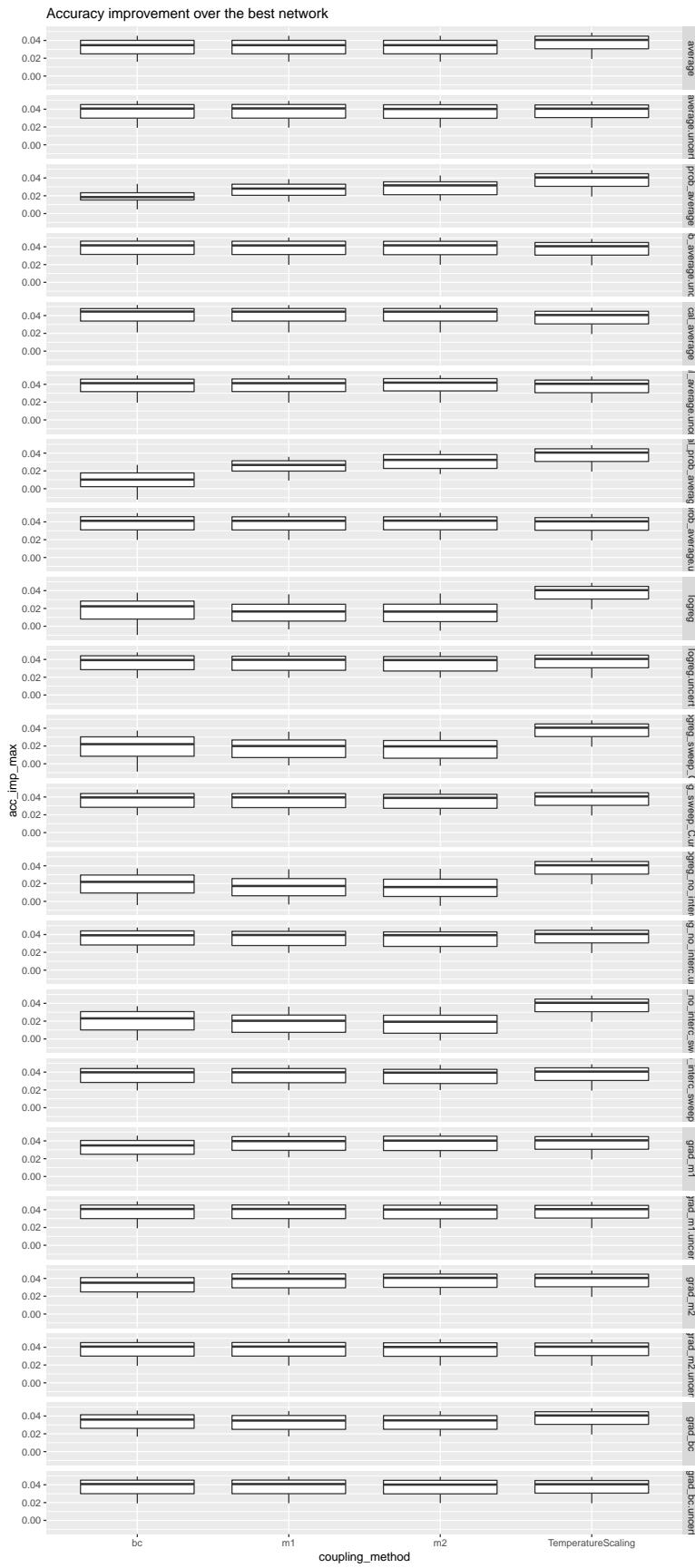
Accuracy improvement of ensemble over
the average of networks vs ece_avg.
Ensemble size 6



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



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



```

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

  print(imp_avg_plot)

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

  print(imp_max_plot)
}

```

