

Mapping patterns of occupancy and richness in a community characterized by strong philopatry

Appendix S2: Posterior Predictive Check

The posterior predictive check was completed by first drawing a set of simulated observed states for each site, year, and visit during each iteration of the MCMC sampler. Recall that the probability of recording each state in each site and year, named “eval” in the model code, was returned in the model output. It was calculated as the sum of the products of each true state probability and the detection probabilities that would have resulted in recording the observed state in question:

$$E_{state_{s,y,v}} = \sum_{i=1}^3 (\phi_{s,y} * p_{i,state})$$

Test statistics were then calculated as follows:

for the simulated (i.e. predicted) data:

$$T_{pred} = \sum_{state=1}^3 \sum_{s=1}^{181} \sum_{y \in Y} \sum_{v=1}^{V_{s,y}} \frac{(Simulated_{s,y,v} - E_{state_{s,y,v}})^2}{E_{state_{s,y,v}}}$$

and for the actual data:

$$T_{actual} = \sum_{state=1}^3 \sum_{s=1}^{181} \sum_{y \in Y} \sum_{v=1}^{V_{s,y}} \frac{(Observed_{s,y,v} - E_{state_{s,y,v}})^2}{E_{state_{s,y,v}}}$$

Bayesian p-values were then calculated as the proportion of iterations for which $T_{pred} > T_{actual}$.

R code for extracting the correct values from the model output and running the posterior predictive check, as well as all of the results, are below.

```
# First ensure that the model for the correct species is loaded,
# and that the parameters "eval" and "y.new" were included in the output.

# Each species was run one at a time,
# making it more useful to have generic code:
zm<-zm.GEPE

##### Extracting the predictive data #####
y.new.results <- array(dim = c(lc, S, Y, V))
for (i in 1:S) {
  for (j in 1:Y) {
    for (k in 1:V) {
      y.new.results[,i,j,k] <-
        c(as.numeric(zm[[1]][,which(colnames(zm[[1]]) ==
```



```

        ],
        ),
        as.numeric(zm[[2]][,which(colnames(zm[[2]]) ==
          (paste("eval[",
            as.character(i), ",",
            as.character(j), ",",
            as.character(k), "]",
            sep="")))
        )
      ]
    ),
    as.numeric(zm[[3]][,which(colnames(zm[[3]]) ==
      (paste("eval[",
        as.character(i), ",",
        as.character(j), ",",
        as.character(k), "]",
        sep="")))
      )
    ]
  )
}
}
}

```

Test statistic for predicted data:

```

t.pred <- rep(NA, times = lc)
system.time(
for (i in 1:lc) {
  for (j in 1:S) {
    for (k in 1:Y) {
      for (l in 1:V) {

        temp.y.vect <-
          c(as.numeric(y.new.results[i,j,k,l]==1),
            as.numeric(y.new.results[i,j,k,l]==2),
            as.numeric(y.new.results[i,j,k,l]==3))

        t.pred[i] <-
          sum(t.pred[i], sum((temp.y.vect - eval.results[i,j,k,]) ^2 /
            eval.results[i,j,k,], na.rm=T), na.rm=T)

      }
    }
  }
}
)
```

```

# Test statistic for actual data:

t.actual <- rep(NA, times = lc)
system.time(
  for (i in 1:lc) {
    for (j in 1:S) {
      for (k in 1:Y) {
        for (l in 1:V) {

          temp.y.vect <-
            c(as.numeric(y[j,k,l]==1),
              as.numeric(y[j,k,l]==2),
              as.numeric(y[j,k,l]==3))

          t.actual[i] <-
            sum(t.actual[i], sum((temp.y.vect - eval.results[i,j,k,]) ^2 /
              eval.results[i,j,k,], na.rm=T), na.rm=T)

        }
      }
    }
  }

# to work with the results later:
save(t.pred, t.actual, file="GEPE.post.pred.Rdata")



# Bayesian P-value:
sum((t.pred - t.actual) > 0) / length(t.actual)

```

Posterior Predictive Check Results

Bayesian P-value

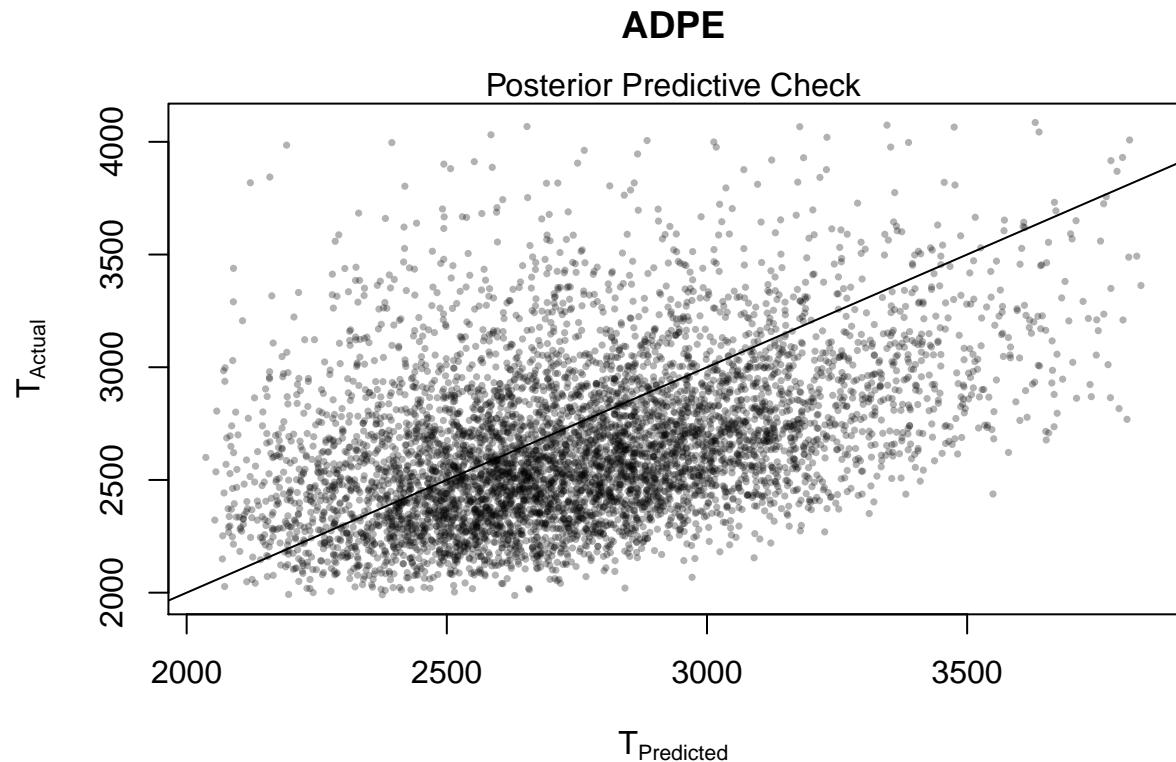
Values greater than 0.9 are highlighted in bold.

Species	P-value
Adelie Penguin	0.662
Antarctic Shag	0.970
Antarctic Tern	0.558
Black-bellied Storm-petrel	0.497
Brown Skua	0.749
Cape Petrel	0.431
Chinstrap Penguin	0.823
Gentoo Penguin	0.929
Kelp Gull	0.652
Macaroni Penguin	0.704
Snow Petrel	0.755
Snowy Sheathbill	0.727
South Polar Skua	0.689

Species	P-value
Southern Fulmar	0.398
Southern Giant Petrel	0.768
Wilson's Storm-petrel	0.484

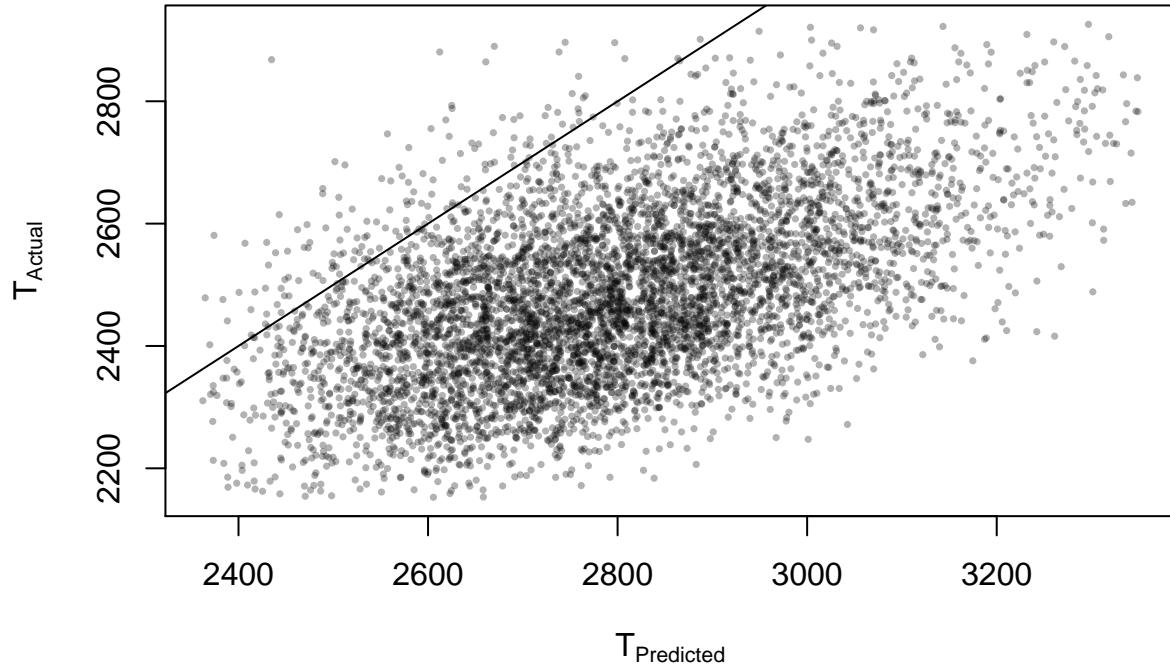
Graphical posterior predictive check

In the following figures, the test-statistic for the actual data is plotted against the test-statistic from the predicted data for each MCMC iteration. The 1:1 line is depicted. For ease of visualization, points more extreme than the 0.5% and 99.5% quantiles for either the predicted or the actual data test-statistics were removed.



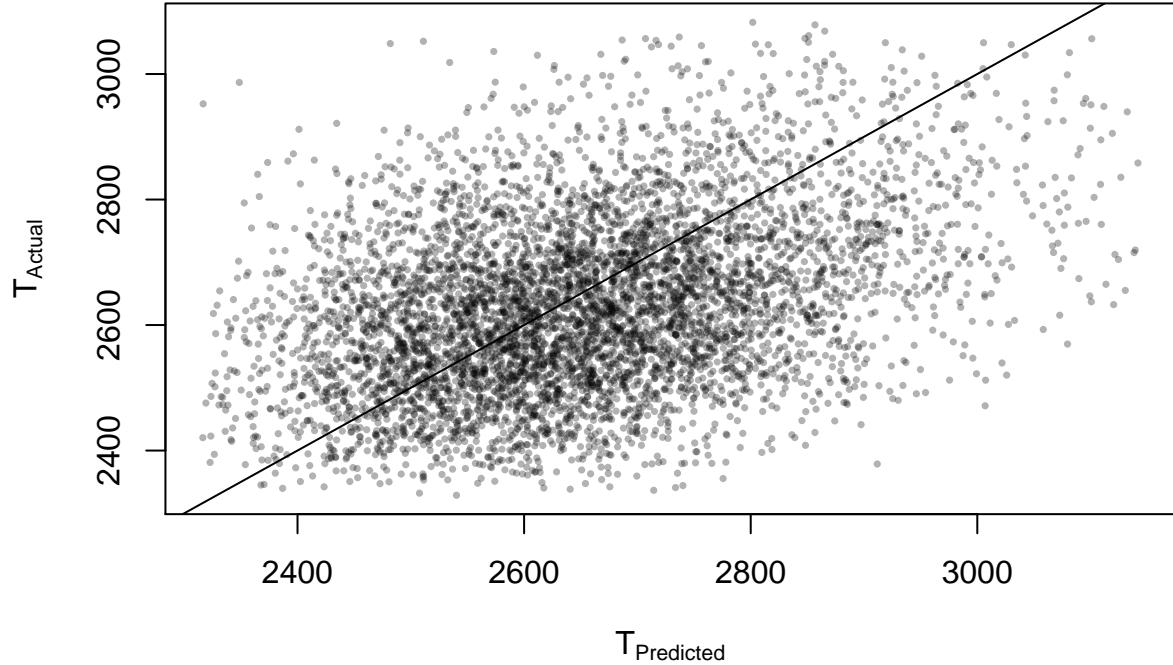
ANSH

Posterior Predictive Check



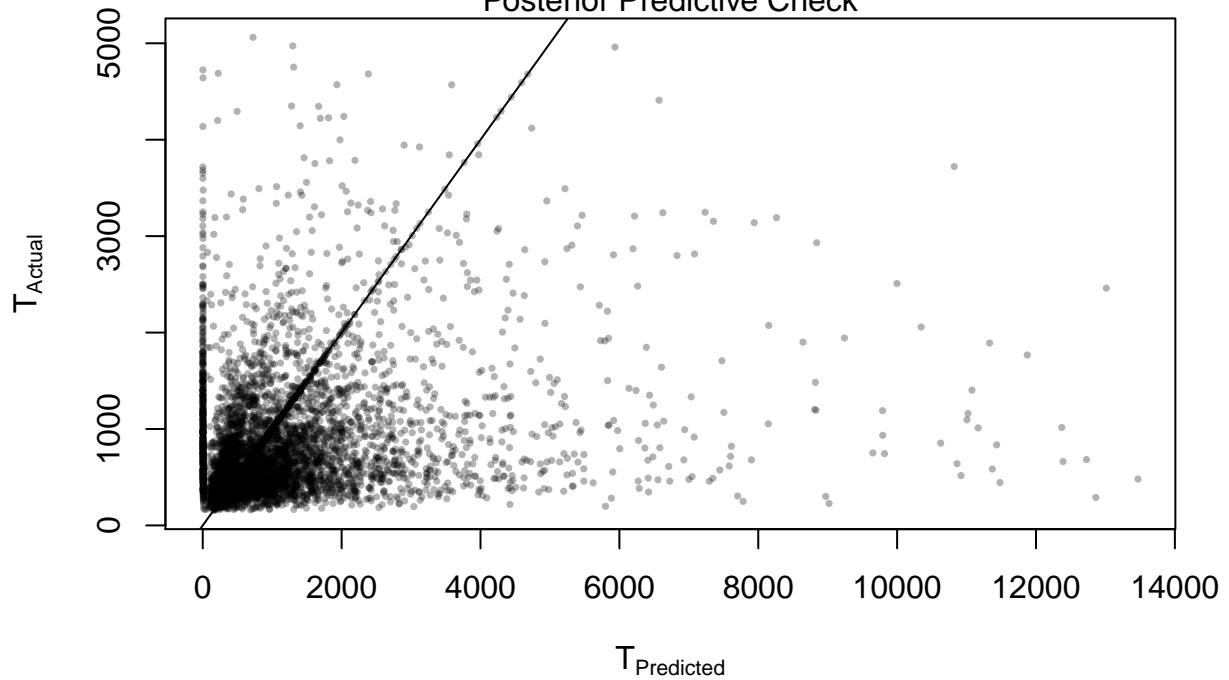
ANTE

Posterior Predictive Check



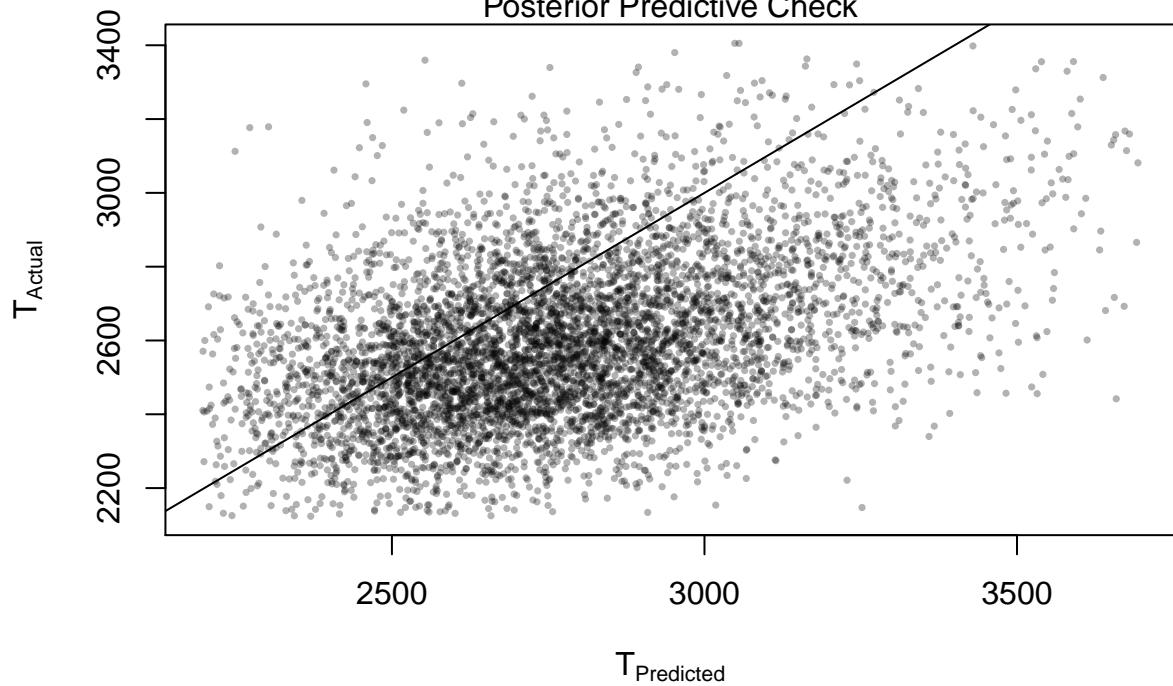
BBSP

Posterior Predictive Check



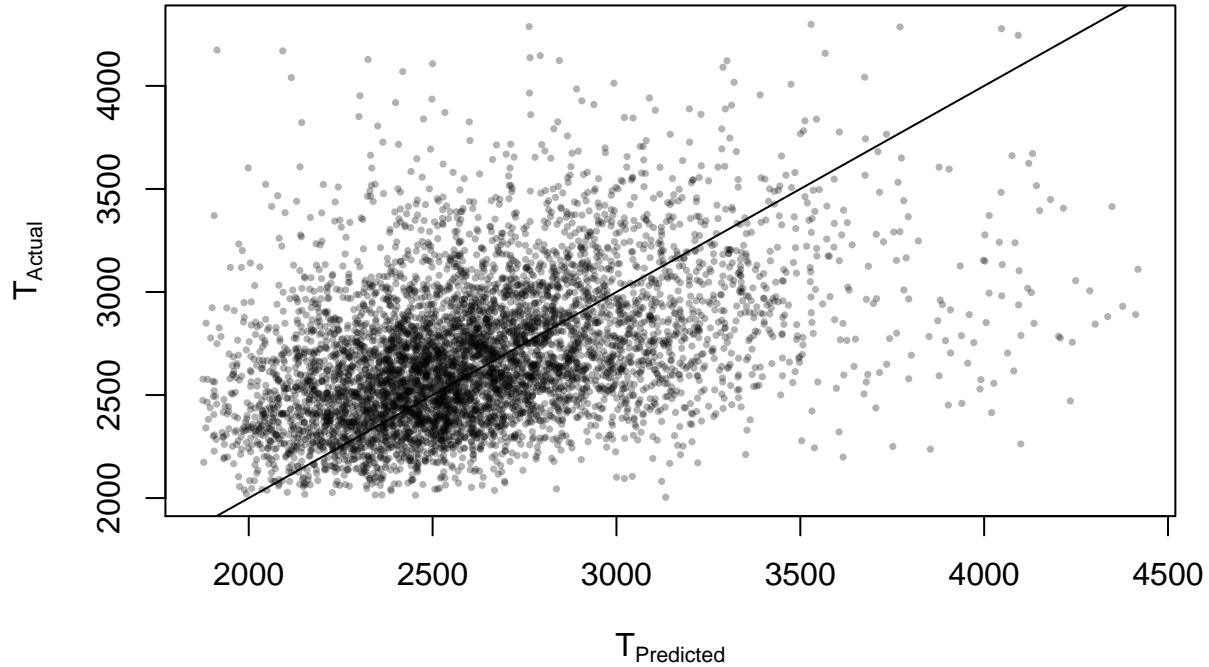
BRSK

Posterior Predictive Check



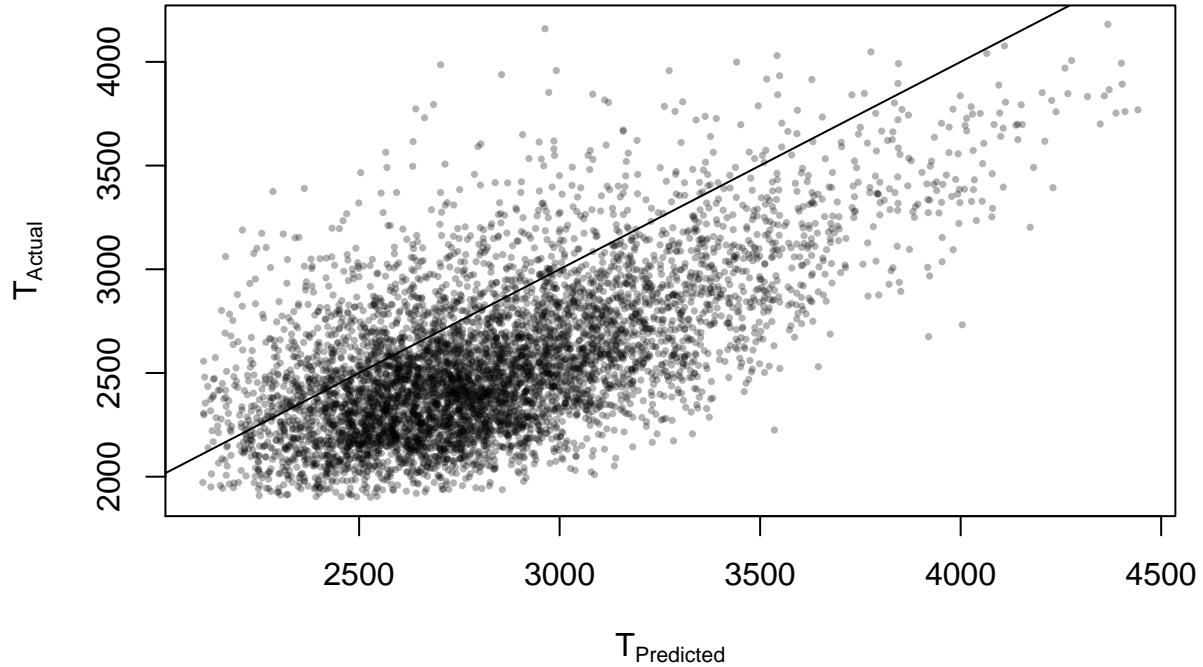
CAPE

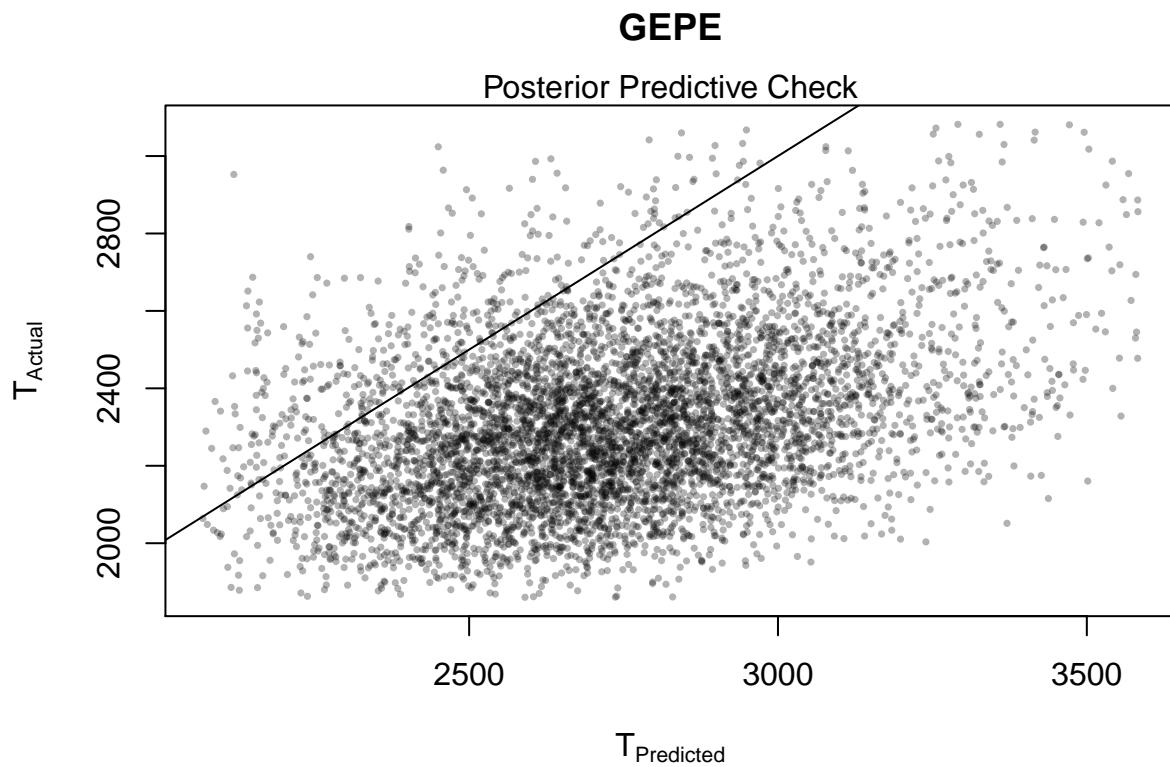
Posterior Predictive Check



CHPE

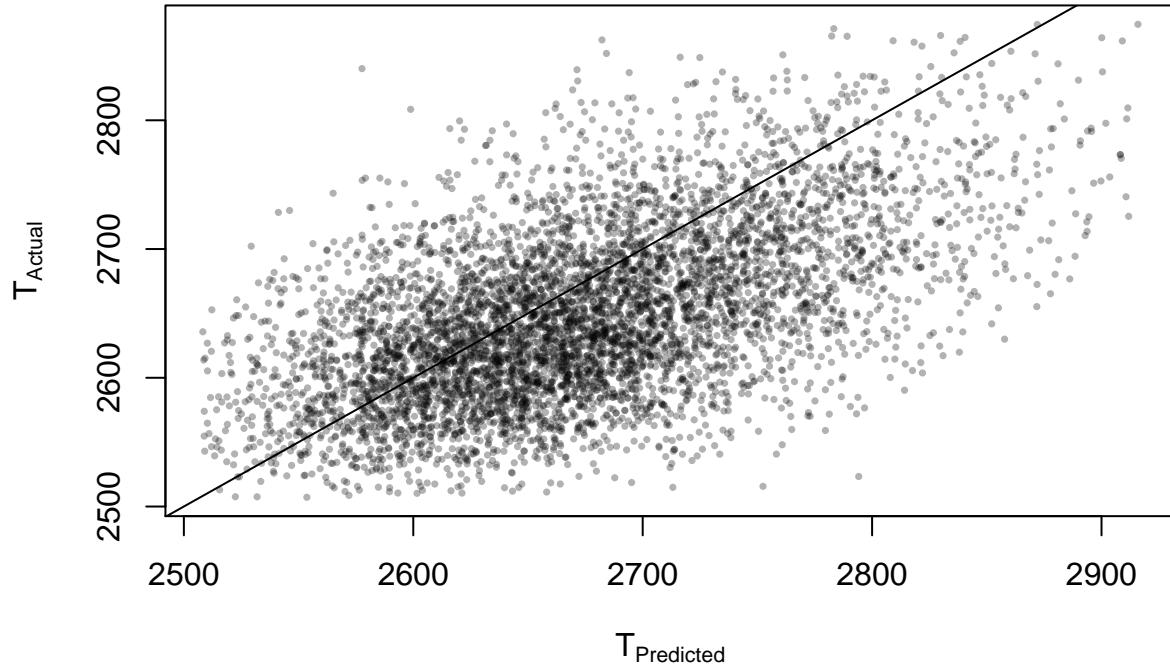
Posterior Predictive Check

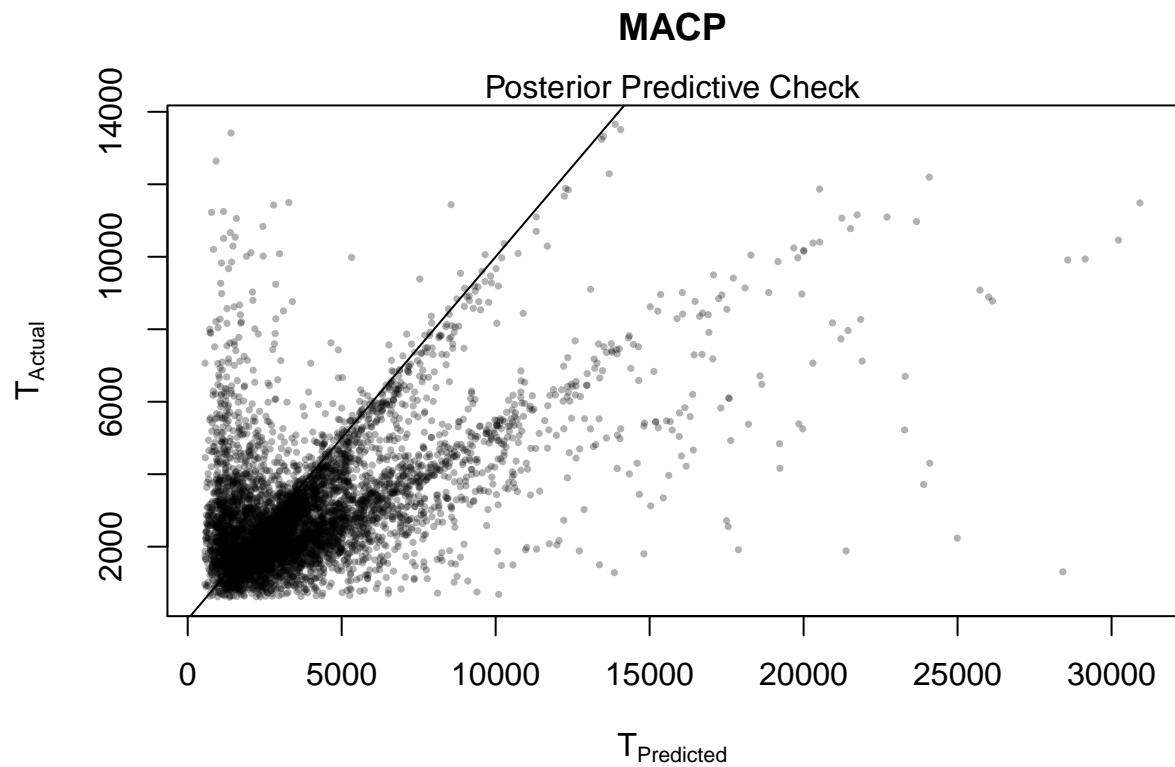


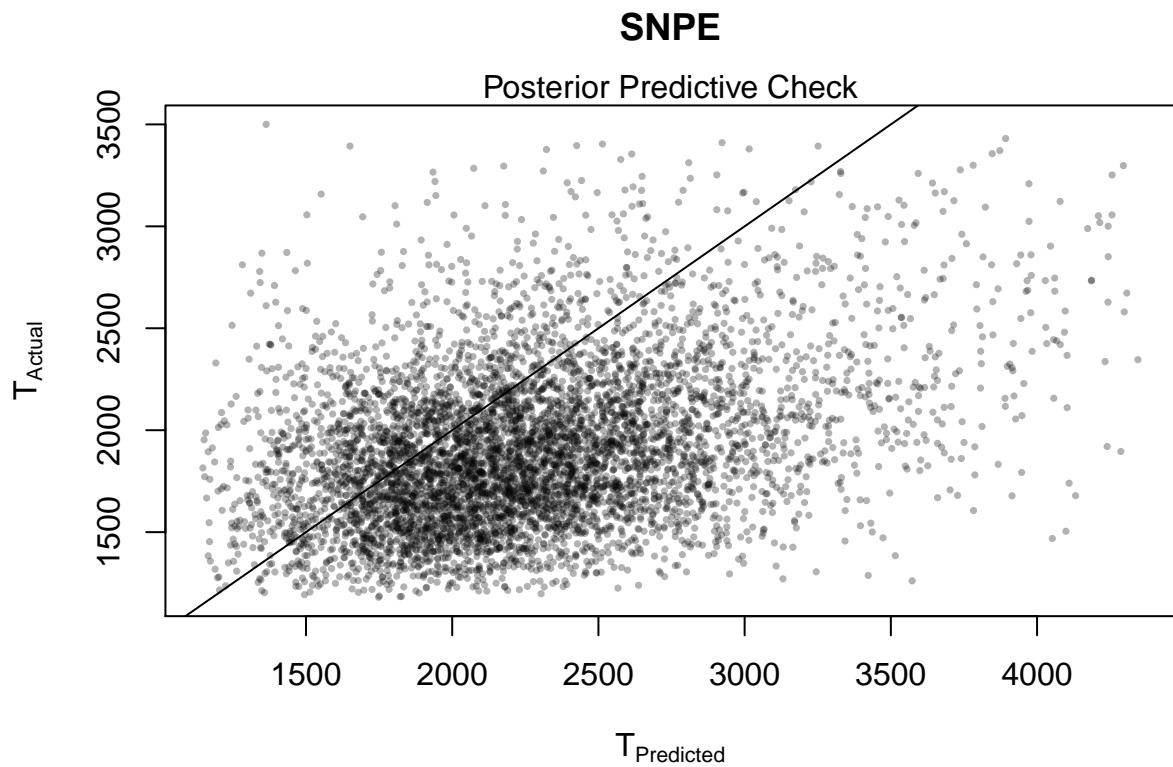


KEGU

Posterior Predictive Check

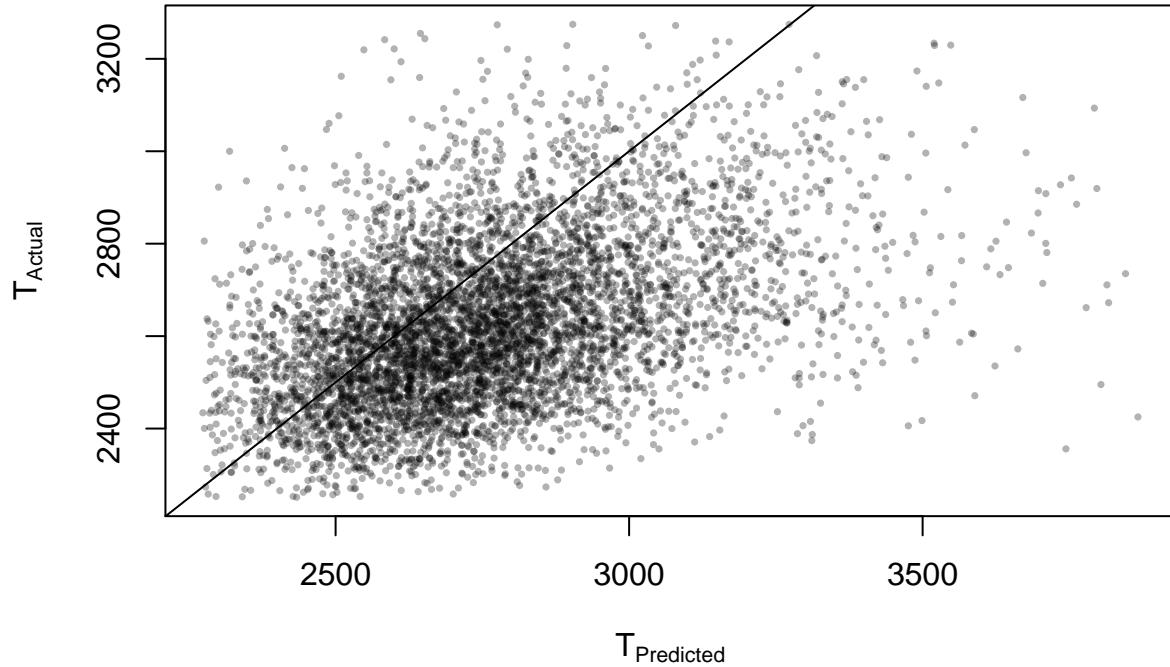


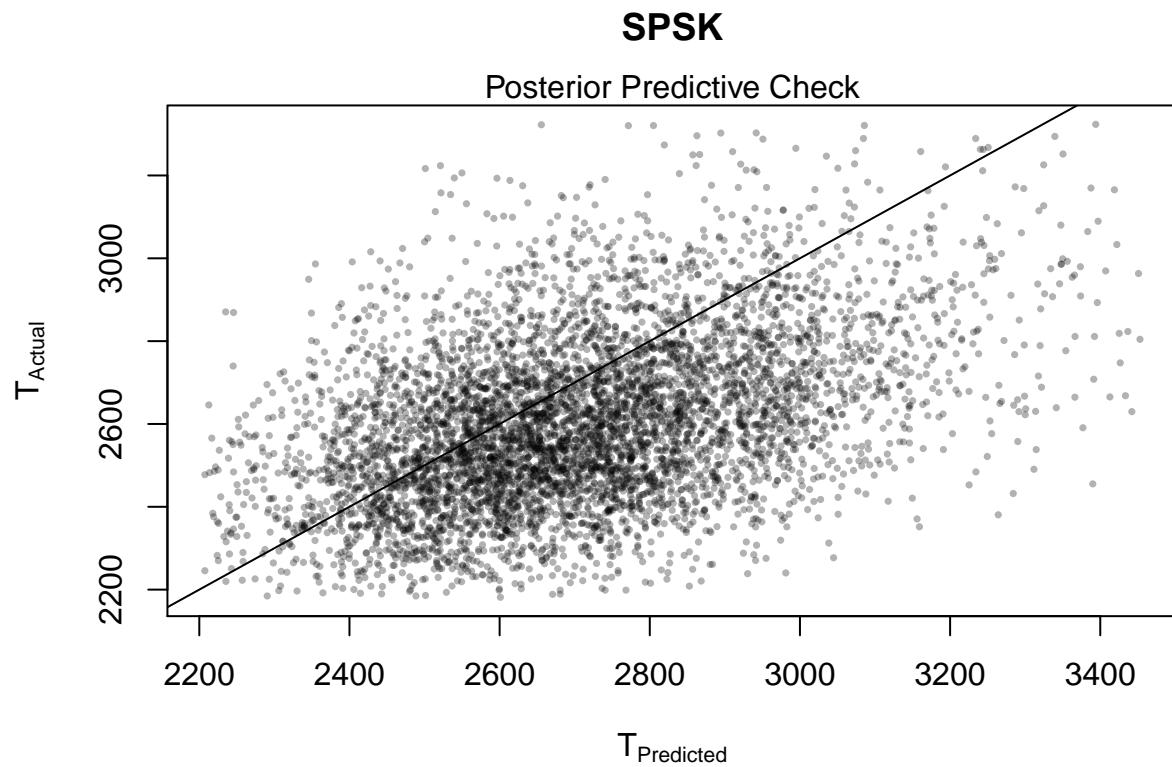




SNSH

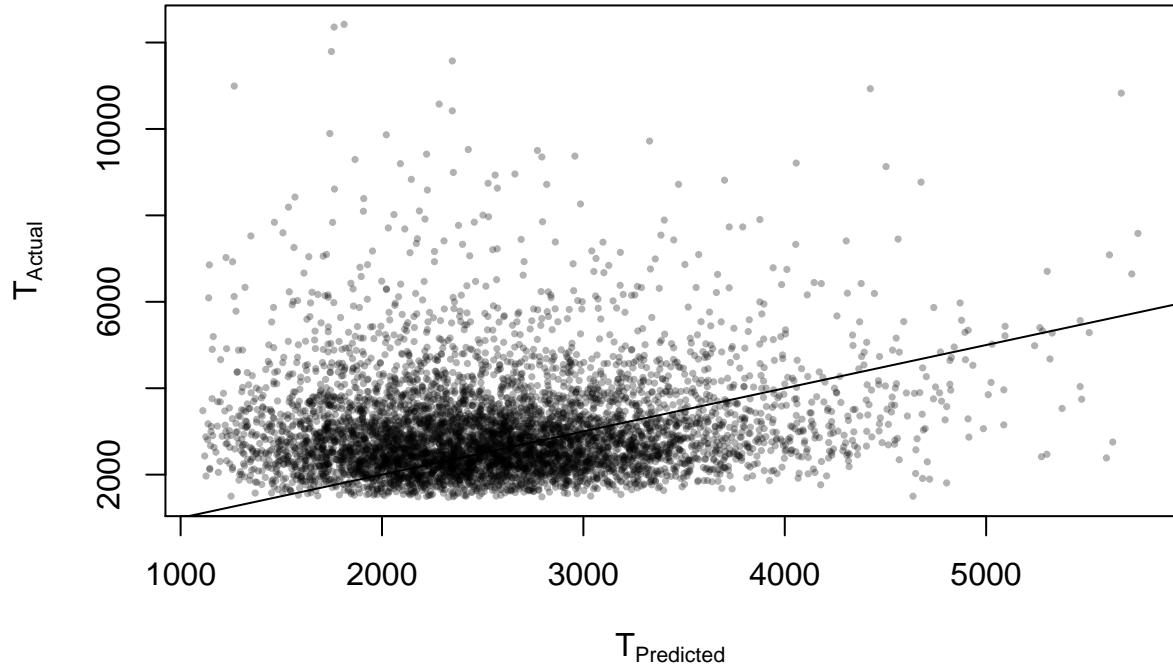
Posterior Predictive Check





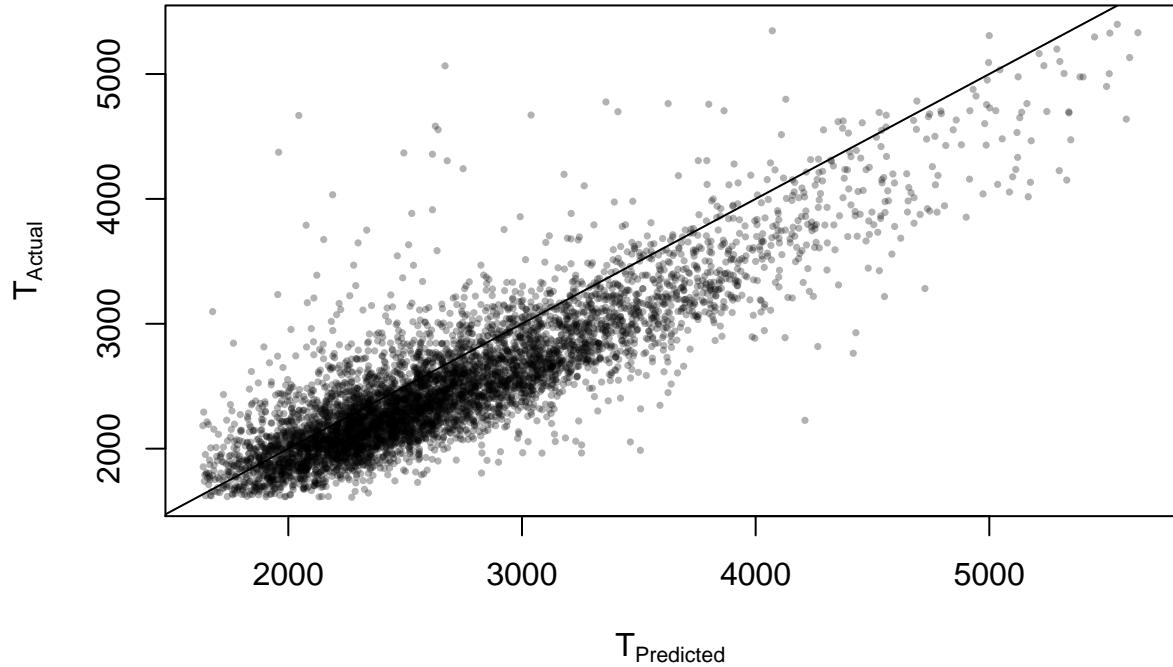
SOFU

Posterior Predictive Check



SGPE

Posterior Predictive Check



WISP

Posterior Predictive Check

