

Including nn[k]DistNotConsumed

Gan Yao

2022-12-15

Intro

This analysis is to reproduce the trick performed in [TR658](#) by Shaw et al. In their analysis, variable `flat` only makes sense to nodes `ldsi`, while variables `posi` and `row` only make sense to nodes `ld0i` and `r0i`. In our analysis, `nn[k]Dist` makes sense to all nodes, while `nn[k]DistNotConsumed` only makes sense to nodes after `isHarvested`.

In TR658, it is suggested that model matrix should be manually created in this case. But as studied in *HandCrafted_ModMat.pdf*, this is not necessary.

Another tricky thing about `nn[k]DistNotConsumed` is that it only has valid values for sampled individuals, and are NAs for the rest. We deal with these NAs by replacing them with zeros. One might think this behavior incorrectly assigns zero `nn[k]DistNotConsumed` values to those individuals which may harm our model. However, as pointed out in TR658, this is an illusion and we do not need to worry about it.

New modelling choices

Thanks to Dr. Jared Beck and Prof. Stuart Wagenius, who pointed out that `nn[k]Dist` are highly correlated (so are `nn[k]DistDistNotConsumed`), we decide to only include `nn5Dist` and `nn5DistNotConsumed` as covariates and fit new models. Also, it may not be appropriate to include `site` as random effect in model. Hence we step back to only consider fixed effects model.

Data

After some correction, the latest version of data is complete and tidy. Inconsistencies found in previous analyses are solved.

Aaster graph

root → *flCt* → *flCtNotConsumed* → *flCtUndamaged* → *capsuleCt* → *isHarvested* → *ovuleCt* → *embryoCt(fitness)*

Load library and data, fit basic models

```
library(aster)
library(tidyverse)
```

```
data <- read.csv("data/output/remLilium2021Data30Nov2022.csv")
names(data)
```

```
## [1] "id" "site" "year"
```

```

## [4] "Ax" "Ly" "flCt"
## [7] "capsuleCt" "nCpsulesHarvested" "ovuleCt"
## [10] "embryoCt" "nn1Dist" "nn2Dist"
## [13] "nn3Dist" "nn4Dist" "nn5Dist"
## [16] "nn6Dist" "nn7Dist" "nn8Dist"
## [19] "nn9Dist" "nn10Dist" "nn1DistNotConsumed"
## [22] "nn2DistNotConsumed" "nn3DistNotConsumed" "nn4DistNotConsumed"
## [25] "nn5DistNotConsumed" "nn6DistNotConsumed" "nn7DistNotConsumed"
## [28] "nn8DistNotConsumed" "nn9DistNotConsumed" "nn10DistNotConsumed"
## [31] "fecundity" "flCtNotConsumed" "flCtUndamaged"

data <- data[data$site != "lf",]
data <- data[data$site != "wrrx",]
data[is.na(data$nCapsulesHarvested), 'nCpsulesHarvested'] <- 0
data[is.na(data$ovuleCt), 'ovuleCt'] <- 0
data[is.na(data$embryoCt), 'embryoCt'] <- 0
names(data)[names(data) == 'nCpsulesHarvested'] <- 'isHarvested'

pred <- c(0,1,2,3,4,5,6)
fam <- c(2,1,1,1,1,2,1)
vars <- c("flCt", "flCtNotConsumed", "flCtUndamaged", "capsuleCt", "isHarvested", "ovuleCt", "embryoCt")

redata <- reshape(data, varying = list(vars), direction="long", timevar="varb", times = as.factor(vars))

redata <- data.frame(redata, root = 1)
model.null <- aster(resp ~ -1 + varb,
                    pred, fam, varb, id, root, data=redata)
summary(model.null)

##
## Call:
## aster.formula(formula = resp ~ -1 + varb, pred = pred, fam = fam,
##               varvar = varb, idvar = id, root = root, data = redata)
##
##               Estimate Std. Error z value Pr(>|z|)
## varbcapsuleCt -1.135e+00 1.703e-01 -6.666 2.64e-11 ***
## varbembryoCt -5.784e-01 1.189e-02 -48.645 < 2e-16 ***
## varbflCt -2.478e-01 6.063e-02 -4.087 4.36e-05 ***
## varbflCtNotConsumed -6.847e-02 1.075e-01 -0.637 0.524
## varbflCtUndamaged 1.366e+00 1.090e-01 12.531 < 2e-16 ***
## varbisHarvested -3.215e+02 1.855e+00 -173.370 < 2e-16 ***
## varbovuleCt 6.027e+00 7.127e-03 845.591 < 2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

eigen(model.null$fisher)$val

## [1] 1.125517e+07 6.263672e+03 1.179341e+03 1.694817e+02 5.191389e+01
## [6] 3.340106e+01 2.906923e-01

pred <- c(0,1,2,3,4,5,6)
fam <- c(2,1,1,1,1,2,1)
vars <- c("flCt", "flCtNotConsumed", "flCtUndamaged", "capsuleCt", "isHarvested", "ovuleCt", "embryoCt")
test <- data %>% mutate(ovuleCt = round(ovuleCt/1), embryoCt = round(embryoCt/1),
                      nn5Dist_s = scale(nn5Dist))

```

```

redata <- reshape(test, varying = list(vars), direction="long", timevar="varb", times = as.factor(vars))
#redata <- reshape(data, varying = list(vars), direction="long", timevar="varb", times = as.factor(vars))

redata <- data.frame(redata, root = 1)
redata$fit <- as.numeric(redata$varb == "embryoCt")
redata$Nid <- as.numeric(gsub("[^0-9.-]", "", redata$id))
#names(redata)
model1 <- aster(resp ~ -1 + varb + fit:nn5Dist_s,
               pred, fam,varb,id,root,data=redata)
#summary(model1, info.tol = 1e-15)
summary(model1)

##
## Call:
## aster.formula(formula = resp ~ -1 + varb + fit:nn5Dist_s, pred = pred,
##               fam = fam, varvar = varb, idvar = id, root = root, data = redata)
##
##               Estimate Std. Error z value Pr(>|z|)
## varbcapsuleCt      -1.135e+00  1.703e-01  -6.666 2.64e-11 ***
## varbembryoCt       -5.788e-01  1.189e-02 -48.659 < 2e-16 ***
## varbflCt          -2.478e-01  6.063e-02  -4.087 4.36e-05 ***
## varbflCtNotConsumed -6.847e-02  1.075e-01  -0.637  0.5241
## varbflCtUndamaged   1.366e+00  1.090e-01  12.531 < 2e-16 ***
## varbisHarvested    -3.215e+02  1.855e+00 -173.358 < 2e-16 ***
## varbovuleCt         6.027e+00  7.127e-03  845.591 < 2e-16 ***
## fit:nn5Dist_s       1.342e-03  5.703e-04   2.354  0.0186 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

eigen(model1$fisher)$val

## [1] 1.136300e+07 3.048860e+06 6.260684e+03 1.179341e+03 1.694817e+02
## [6] 5.191389e+01 3.340106e+01 2.906928e-01

```

Attempts to include nn[k]DistNotConsumed

Scale nn5Dist, divide nn5DistNotConsumed by 1000

```

pred <- c(0,1,2,3,4,5,6)
fam <- c(2,1,1,1,1,2,1)
vars <- c("flCt", "flCtNotConsumed", "flCtUndamaged", "capsuleCt", "isHarvested", "ovuleCt", "embryoCt")
test <- data %>% mutate(nn5Dist_s = scale(nn5Dist), nnDist_s2 =replace_na(nn5DistNotConsumed, 0)) %>% m
#test <- data %>% mutate(nn5Dist_s = nn5Dist/100, nn5DistNotConsumed =replace_na(nn5DistNotConsumed, 0))
redata <- reshape(test, varying = list(vars), direction="long", timevar="varb", times = as.factor(vars))

redata <- data.frame(redata, root = 1)
redata$fit <- as.numeric(redata$varb == "embryoCt")
redata$Nid <- as.numeric(gsub("[^0-9.-]", "", redata$id))
#names(redata)
model2 <- aster(resp ~ -1 + varb + fit:nn5Dist_s + fit:nn5DistNotConsumed_s,
               pred, fam,varb,id,root,data=redata)
#summary(model2, info.tol=1e-12)
summary(model2)

##

```

```
## Call:
## aster.formula(formula = resp ~ -1 + varb + fit:nn5Dist_s + fit:nn5DistNotConsumed_s,
##   pred = pred, fam = fam, varvar = varb, idvar = id, root = root,
##   data = redata)
##
##               Estimate Std. Error  z value Pr(>|z|)
## varbcapsuleCt      -1.135e+00  1.703e-01   -6.666 2.64e-11 ***
## varbembryoCt       -5.797e-01  1.190e-02  -48.703 < 2e-16 ***
## varbflCt          -2.478e-01  6.063e-02   -4.087 4.36e-05 ***
## varbflCtNotConsumed -6.847e-02  1.075e-01   -0.637  0.524
## varbflCtUndamaged   1.366e+00  1.090e-01   12.531 < 2e-16 ***
## varbisHarvested    -3.215e+02  1.855e+00  -173.321 < 2e-16 ***
## varbovuleCt         6.027e+00  7.127e-03   845.591 < 2e-16 ***
## fit:nn5Dist_s      -6.828e-04  7.063e-04   -0.967  0.334
## fit:nn5DistNotConsumed_s 2.796e-03  6.042e-04    4.627 3.71e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
eigen(model2$fisher)$val
```

```
## [1] 1.209651e+07 5.479858e+06 1.441083e+06 6.252286e+03 1.179341e+03
## [6] 1.694817e+02 5.191389e+01 3.340106e+01 2.906944e-01
```

Divide nn5Dist by 1000, divide nn5DistNotConsumed by 1000

```
pred <- c(0,1,2,3,4,5,6)
fam <- c(2,1,1,1,1,2,1)
vars <- c("flCt", "flCtNotConsumed", "flCtUndamaged", "capsuleCt", "isHarvested", "ovuleCt", "embryoCt")
test <- data %>% mutate(nn5Dist_s = nn5Dist/1000, nn5DistNotConsumed = replace_na(nn5DistNotConsumed, 0))
redata <- reshape(test, varying = list(vars), direction="long", timevar="varb", times = as.factor(vars))

redata <- data.frame(redata, root = 1)
redata$fit <- as.numeric(redata$varb == "embryoCt")
redata$Nid <- as.numeric(gsub("[^0-9.-]", "", redata$id))
#names(redata)
model3 <- aster(resp ~ -1 + varb + fit:nn5Dist_s + fit:nn5DistNotConsumed_s,
  pred, fam, varb, id, root, data=redata)
#summary(model3, info.tol=1e-12)
summary(model3)
```

```
##
## Call:
## aster.formula(formula = resp ~ -1 + varb + fit:nn5Dist_s + fit:nn5DistNotConsumed_s,
##   pred = pred, fam = fam, varvar = varb, idvar = id, root = root,
##   data = redata)
##
##               Estimate Std. Error  z value Pr(>|z|)
## varbcapsuleCt      -1.135e+00  1.703e-01   -6.666 2.64e-11 ***
## varbembryoCt       -5.807e-01  1.192e-02  -48.709 < 2e-16 ***
## varbflCt          -2.478e-01  6.063e-02   -4.087 4.36e-05 ***
## varbflCtNotConsumed -6.847e-02  1.075e-01   -0.637  0.524
## varbflCtUndamaged   1.366e+00  1.090e-01   12.531 < 2e-16 ***
## varbisHarvested    -3.215e+02  1.855e+00  -173.321 < 2e-16 ***
## varbovuleCt         6.027e+00  7.127e-03   845.591 < 2e-16 ***
## fit:nn5Dist_s      -4.136e-02  4.278e-02   -0.967  0.334
## fit:nn5DistNotConsumed_s 1.250e-01  2.702e-02    4.627 3.71e-06 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

eigen(model3$fisher)$val

## [1] 1.125621e+07 6.297170e+03 2.477553e+03 1.179341e+03 4.635116e+02
## [6] 1.694817e+02 5.191389e+01 3.340105e+01 2.906943e-01
```