\*Modeling the influence of blooming forb coverage on total bee abundance;

\*Year 1

\*Load in the data;

proc import out = plantcoverage\_year1

datafile = 'C:/Users/Morgan/Documents/ISU/Project/SAS/Data Files/PlantCoverage\_Year1.xlsx'

dbms = xlsx

replace;

getnames = yes;

datarow = 2;

run;

\*Print the data to make sure it loaded okay;

proc print data = plantcoverage\_year1;

run;

\*Poisson mixed effects model with average plant coverage as a covariate;

proc glimmix data = plantcoverage\_year1 plots = studentpanel;

class Sampling\_Day Site;

model Total\_Bees = Average\_Coverage|Sampling\_Day / solution link = log dist = poisson ddfm = satterthwaite;

random Site;

lsmeans Sampling\_Day / cl;

run;

\*The model fits with no problems in SAS! However, the residual plot doesn’t look great, and

the value of Gener. Chi-Square / DF suggests that there is overdispersion.

\*Try fitting a negative binomial model:

\*Negative binomial mixed effects model with average plant coverage as a covariate -

\*(this model is fit as a way of dealing with overdispersion);

proc glimmix data = plantcoverage\_year1 plots = studentpanel;

class Sampling\_Day Site;

model Total\_Bees = Average\_Coverage|Sampling\_Day / solution link = log dist = negbinomial ddfm = satterthwaite;

random Site;

lsmeans Sampling\_Day / cl diff;

run;

\*No more overdispersion and the residual plot looks great!;

\*Let’s see what happens if we include the random effect for each observation in the model…;

proc glimmix data = plantcoverage\_year1 plots = studentpanel;

class Sampling\_Day Site A;

model Total\_Bees = Average\_Coverage|Sampling\_Day / solution link = log dist = poisson ddfm = satterthwaite;

random Site A;

lsmeans Sampling\_Day / cl;

run;

\*The residual plot looks scary, so best to use the negative binomial model;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*Year 2

\*Load in the data;

proc import out = plantcoverage\_year2

datafile = 'C:/Users/Morgan/Documents/ISU/Project/SAS/Data Files/PlantCoverage\_Year2.xlsx'

dbms = xlsx

replace;

getnames = yes;

datarow = 2;

run;

\*Print the data to make sure it loaded okay;

proc print data = plantcoverage\_year2;

run;

\*Negative binomial mixed effects model with average plant coverage as a covariate;

proc glimmix data = plantcoverage\_year2 plots = studentpanel;

class Sampling\_Day Site;

model Total\_Bees = Average\_Coverage|Sampling\_Day / solution link = log dist = negbinomial ddfm = satterthwaite;

random Site;

lsmeans Sampling\_Day / cl diff;

run;

\*Doesn't converge, bummer;

\*Let's try the Poisson distribution again;

proc glimmix data = plantcoverage\_year2 plots = studentpanel;

class Sampling\_Day Site;

model Total\_Bees = Average\_Coverage|Sampling\_Day / solution link = log dist = poisson ddfm = satterthwaite;

random Site;

lsmeans Sampling\_Day / cl;

run;

\*Still crazy overdispersed;

\*Try adding each observation as a random effect;

proc glimmix data = plantcoverage\_year2 plots = studentpanel;

class Sampling\_Day Site A;

model Total\_Bees = Average\_Coverage|Sampling\_Day / solution link = log dist = poisson ddfm = satterthwaite;

random Site A;

lsmeans Sampling\_Day / cl;

run;

\*Residual plot is garbage;

\*Let's try some log stuff

\*log transform the “Total\_Bees” variable, no need to add +1 because there are no zeroes;

data plantcoverage\_year2;

set plantcoverage\_year2;

log\_Total\_Bees = log(Total\_Bees);

run;

\*Sort it by log\_Total\_Bees value;

proc sort data = plantcoverage\_year2;

by log\_Total\_Bees;

run;

\*Print the data to make sure SAS included the new variable in correctly;

proc print data = plantcoverage\_year2;

run;

\*Mixed model with the log transformed bee numbers;

Proc Mixed Data = plantcoverage\_year2 plots = studentpanel;

class Site Year Sampling\_Day;

model log\_Total\_Bees = Average\_Coverage|Sampling\_Day|Year / s ddfm = sat;

random Site Site\*Sampling\_Day;

run;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*Year 3

\*Load in the data;

proc import out = plantcoverage\_year3

datafile = 'C:/Users/Morgan/Documents/ISU/Project/SAS/Data Files/PlantCoverage\_Year3.xlsx'

dbms = xlsx

replace;

getnames = yes;

datarow = 2;

run;

\*Print the data to make sure it loaded okay;

proc print data = plantcoverage\_year3;

run;

\*Negative binomial mixed effects model with average plant coverage as a covariate;

proc glimmix data = plantcoverage\_year3 plots = studentpanel;

class Sampling\_Day Site;

model Total\_Bees = Average\_Coverage|Sampling\_Day / solution link = log dist = negbinomial ddfm = satterthwaite;

random Site;

lsmeans Sampling\_Day / cl diff;

run;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*Years 1 and 2

\*Load in the data;

proc import out = plantcoverage\_years12

datafile = 'C:/Users/Morgan/Documents/ISU/Project/SAS/Data Files/PlantCoverage\_Years12.xlsx'

dbms = xlsx

replace;

getnames = yes;

datarow = 2;

run;

\*Print the data to make sure it loaded okay;

proc print data = plantcoverage\_years12;

run;

\*log transform the “Total\_Bees” variable, no need to add +1 because there are no zeroes;

data plantcoverage\_years12;

set plantcoverage\_years12;

log\_Total\_Bees = log(Total\_Bees);

run;

\*Sort it by log\_Total\_Bees value;

proc sort data = plantcoverage\_years12;

by log\_Total\_Bees;

run;

\*Print the data to make sure SAS included the new variable in correctly;

proc print data = plantcoverage\_years12;

run;

\*Mixed model with the log transformed bee numbers;

Proc Mixed Data = plantcoverage\_years12 plots = studentpanel;

class Site Year Sampling\_Day;

model log\_Total\_Bees = Average\_Coverage|Sampling\_Day|Year / s ddfm = sat;

random Site Site\*Sampling\_Day;

run;

\*Converges, no overdispersion, and residual plot looks okay!;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*Years 1, 2, and 3

\*Load in the data;

proc import out = plantcoverage\_years123

datafile = 'C:/Users/Morgan/Documents/ISU/Project/SAS/Data Files/PlantCoverage\_Years123.xlsx'

dbms = xlsx

replace;

getnames = yes;

datarow = 2;

run;

\*Print the data to make sure it loaded okay;

proc print data = plantcoverage\_years123;

run;

\*log+1 transform the “Total\_Bees” variable (need the +1 in this case because there are zeroes in the data);

data plantcoverage\_years123;

set plantcoverage\_years123;

log\_Total\_Bees = log(Total\_Bees + 1);

run;

\*Sort it by log\_Total\_Bees value;

proc sort data = plantcoverage\_years123;

by log\_Total\_Bees;

run;

\*Print the data to make sure SAS included the new variable in correctly;

proc print data = plantcoverage\_years123;

run;

\*Mixed model with the log transformed bee numbers;

Proc Mixed Data = plantcoverage\_years123 plots = studentpanel;

class Site Year Sampling\_Day;

model log\_Total\_Bees = Average\_Coverage|Sampling\_Day|Year / s ddfm = sat;

random Site Site\*Sampling\_Day;

run;

\*Converges, no overdispersion, and residual plot looks okay!