Osmia Apple Orchard Nesting Structure Survival Rate

Batoule Hyjazie

18 April, 2020

1. Introduction

1.1 Background

Like many species around the world, pollinators are heavily affected by habitat loss, often driven by agricultural intensification, among other factors like pesticide use, disease and climate change (Potts et al. 2010). Natural habitats provide floral and nesting resources to pollinators which are both essential to survival and growth (Roulston and Goodell 2011). Ironically, while, most pollinator population declines are often attributed to agriculture, pollinators have been shown to not only enhance crop production but also, to be uniquely essential to the production of certain crops like macadamia, cantaloupe and watermelon (Klein et al. 2006). Several studies have used artificial nesting structures, also known as trap nests or bee hotels (Geslin et al. 2020), to not only study wild bees but also, to promote their conservation in urban and agricultural settings (MacIvor and Packer 2015; N. K. Boyle and Pitts-Singer 2017). Understanding how bees respond to their environment, and in particular, to changes in resource availability like with floral and nesting resources, can allow us to better predict their abundance and by extension their survival. Most importantly, this can aid in developping better conservation strategies in both natural and agricultural environments.

1.2 Study System

While nesting structures are widely used in the study of bees, no study has yet examined whether the provision of nesting structures actually increases local bee populations in comparison to control populations, establishing that bees are nest site limited (Westerfelt, Weslien, and Widenfalk 2018; Roulston and Goodell 2011). In order to conduct such a study, I must first test that nesting structures installed in anthropogenic habitat are not a population sink, attracting nesting female bees to a potentially unfit environment for larval survival, which could theoretically harm, rather than protect, local wild bee populations.

In Hungarian apple orchards, it has been found that when comparing nesting structures installed in paired apple orchard and natural habitat sites, higher colonization rates and counts of bee and wasp live offspring were observed in semi-natural habitat (Bihaly et al. 2020). This is possibly due to a higher prevalence of floral resources in semi-natural habitat when compared to orchard floral resources (Bihaly et al. 2020).

In fruit orchards, a wide variety of pollinators such as butterflies, wasps and of course, bees, both domestic and wild, can be found. *Osmia* spp. is a common wild bee genus in fruit orchards that distinguishes itself from other bees, like honey bees, by its ability to fly at cooler temperatures (Isaacs et al. 2017; Natalie K. Boyle and Pitts-Singer 2019) and deposit higher pollen loads on a wide-range of flowers (Földesi et al. 2015). To test whether similar observations can be made in orchards within the greater Ottawa area, nesting structure survival rates of *Osmia* spp. ("mason bees") offspring in orchard and natural habitats will be compared with one another.

Figure 1: Osmia Entering Nesting Structure

1.3 Hypothesis

The installation of nesting structures in apple orchards, are potential population sinks rather than populations sources, conceivably due to lesser floral resource availability over the entire growing season. As a result, nesting structures in natural habitat will have a greater ratio of survival of *Osmia spp.* offspring when compared to apple orchard habitat.

1.4 Prediction

If nesting structures placed in natural environments have a significantly greater *Osmia spp.* survival ratio than orchard environments, then it is likely that nesting structures placed in orchard environments are an *Osmia spp.* population sink.

2. Methods

2.1 Data Collection

This study was conducted from May to August 2017 in 7 apple orchards and 8 natural habitats (6 on National Capital Commission (NCC) grounds and 2 on City of Ottawa grounds) around the Ottawa-area for a total of 15 sites. Overall, 75 nesting blocks were installed in both natural and orchard habitats with 5 blocks per site. Each block had 10 nests or holes.

Throughout these 4 months, various insects, mostly wasps and bees, inhabited the holes, and some nested within them, laying their eggs. The eggs developed first as larvae, then as pupae. In September, the blocks were collected and stored in a laboratory environment where the insects over-wintered (hibernated). Before their expected scheduled emergence, the nests were inspected and *Osmia* spp. cell survival count was estimated using visual markers such as nest material and cell size.

2.2 Data Set

The "Osmia Survival.xlsx" file contains the data that will be used for subsequent statistical analyses. It includes information such as the original number of cells observed in the straw (*No.cells*), the initial perceived number of surviving cells (*No.surv.cells*) as well as a validated revised number of surviving *Osmia* spp. cells (*No.surv.osmia*). The number of *Osmia* spp. cell deaths was calculated as the difference between the number of cells and the revised number of surviving *Osmia* spp. cells (*No.death*).

It is important to note that the number of surviving cells does not account for bees that died over winter due to unfavorable storage conditions. Therefore, we are only accounting for deaths of bees that have clearly died as larvae either through parasitization or other un-diagnosed causes. Deaths at the larval stage can only be attributed to causes occurring during the growing season or associated with either orchard or natural conditions. By the developmental stage of adult eclosion, nests were already transferred to a laboratory environment. Any deaths occurring at the adult stage are ignored.

2.2 Statistical Methods

To explore frequentist and bayesian statistics, as an exercise, I will be running the same model using these two different approaches. Due to differences in code structure, the response for the frequentist approach will be written as ratio (*No.surv.osmia/No.cells*) while the bayesian approach will be written as a combined column (*cbind(No.surv.osmia, No.death)*). The random effects of each approach will be the same (block ID nested within site ID).

Because the response that is measured is the proportion of surviving Osmia offspring, the data is not normally distributed but binomially distributed. This was verified using both a histogram (**Figure 2a**) of the distribution of the frequency of *Osmia* spp. survival ratios and a q-q plot (**Figure 2b**). Thus, all models will be run using a binomial distribution.
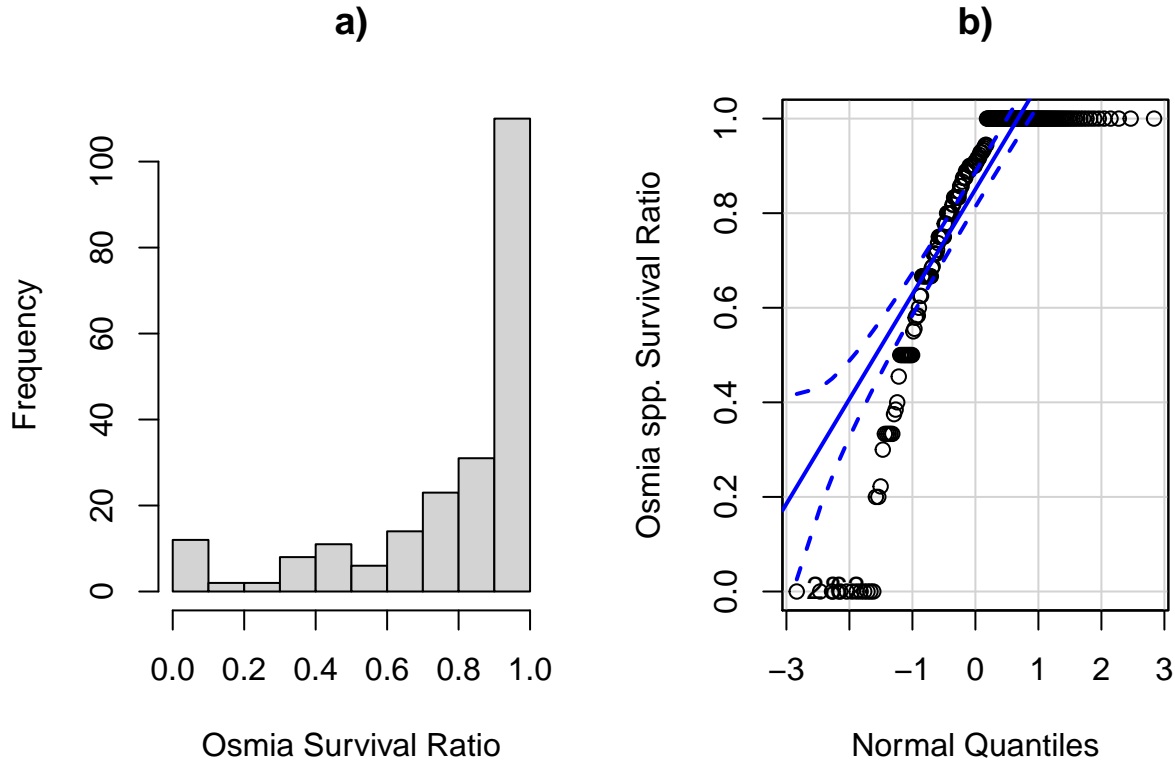
a)   **b)**

Figure 2: Normality Tests of *Osmia* spp. survival ratio a) Histogram and b) Q-Q Plot

2.2.1 Frequentist Approach

Knowing that the data was hierarchical with random effects of block nested within site, I wanted to test the significance of these terms. To do this, I ran a simple generalized linear model (GLM).

```
##
## Call:
## glm(formula = No.surv.osmia/No.cells ~ s.type, family = binomial(link = "logit"),
##     data = Osmia, weights = No.cells)
##
## Deviance Residuals:
##     Min       1Q    Median       3Q       Max
## -5.6454  -0.6845    0.6313   1.4834    2.6942
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.5839     0.0918   17.25   <2e-16 ***
## s.typeo      -0.0851     0.1290   -0.66    0.509
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

4

```
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 576.71  on 218  degrees of freedom
## Residual deviance: 576.28  on 217  degrees of freedom
## AIC: 820.78
##
## Number of Fisher Scoring iterations: 4
```

Out of curiosity, I also wanted to test the results that I would get if I ignored the hierarchical structure of the data and simply fit site as a random effect in a generalized linear mixed model (GLMM).

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##    Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: No.surv.osmia/No.cells ~ s.type + (1 | s.ID)
##     Data: Osmia
##
##      AIC      BIC   logLik deviance df.resid
##    198.0    208.2    -96.0    192.0      216
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.4651 -0.4188  0.1935  0.4577  0.4577
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  s.ID   (Intercept) 0        0
## Number of obs: 219, groups:  s.ID, 15
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.5629     0.2345   6.666 2.64e-11 ***
## s.typeo       0.2416     0.3802   0.635    0.525
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##         (Intr)
## s.typeo -0.617
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

Afterwards, I accounted for the hierarchical structure of the data and ran a more adequate GLMM using block nested within site as a random effect.

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##    Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: No.surv.osmia/No.cells ~ s.type + (1 | s.ID/Block_ID)
##     Data: Osmia
## Weights: No.cells
##
##      AIC      BIC   logLik deviance df.resid
```

```
##    765.5    779.0    -378.7    757.5       215
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.6866 -0.6447  0.4800  0.9038  1.8469
##
## Random effects:
##  Groups          Name         Variance Std.Dev.
##  Block_ID:s.ID (Intercept) 0.70518  0.8397
##  s.ID          (Intercept) 0.06338  0.2518
## Number of obs: 219, groups:  Block_ID:s.ID, 57; s.ID, 15
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.77702    0.22137   8.027 9.97e-16 ***
## s.typeo      0.04725    0.31990   0.148    0.883
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##         (Intr)
## s.typeo -0.666
```

2.2.1.1 Frequentist Key Assumptions   The main assumption of a GLMM, is that the chosen probability distribution and associated link function is appropriate for the data. As previously stated, due to the use of proportional data, the distribution used was binomial and the link function was the logarithm of the odds (logit). Another assumption is that the random effects are normally distributed. Finally, it is assumed that there is no over-dispersion or under-dispersion. This refers to the constancy of the variance and was verified by comparing the ratio of the residual deviance to the degrees of freedom using the *overdisp.glmm* function. A final implicit assumptions of GLMMs, as with all model types, is to match the model with the data and its structure and to properly identify the fixed versus the random effects.

To evaluate model fit, a likelihood ratio test using the *drop1* function was conducted. The GLMM and GLM versions of the model output (one with the random effect term of block nested within site and one without) were compared via a likelihood ratio test and also by simply comparing their Aikake Information Criteria (AIC) from their model outputs. The normality of the random effects was assessed visually using a q-q plot (***Figure 5***). While this is not an explicit assumption of GLMM, the normality of the residuals were evaluated visually using a residuals versus fitted values plot (***Figure 4***).

2.2.2 Bayesian Approach

Two Monte Carlo Markov Chains (MCMC) were run following the same model structure described previously (random effects of block nested within site and the binomial probability distribution). The first chain consisted of a test run and was sequenced using the default settings of the *MCMCglmm* package (sample size of 1,000, number of iterations of 13,000, burnin of 3,000, thin of 10). The second chain (with R model output below) was run with a larger sample size and number of iterations as well as adjusted priors (sample size of 4,000, number of iterations of 2,020,000, burnin of 20,000, thin of 500, V=1, nu=1).

```
##
##  Iterations = 20001:2019501
##  Thinning interval  = 500
##  Sample size  = 4000
##
```

```
##  DIC: 1301.287
##
##  G-structure:  ~s.ID
##
##      post.mean  l-95% CI u-95% CI eff.samp
## s.ID    0.2662 5.985e-10   0.8958     4000
##
##                 ~Block_ID:s.ID
##
##              post.mean  l-95% CI u-95% CI eff.samp
## Block_ID:s.ID   0.6155 1.011e-07   1.429     4000
##
##  R-structure:  ~units
##
##      post.mean l-95% CI u-95% CI eff.samp
## units   2.063    1.19    3.099     3465
##
##  Location effects: cbind(No.surv.osmia, No.death) ~ s.type
##
##             post.mean l-95% CI u-95% CI eff.samp  pMCMC
## (Intercept)    2.0469  1.3922   2.6528     3805 <3e-04 ***
## s.typeo        0.1248 -0.7762   1.0090     4000  0.773
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2.2.2.1 Bayesian Key Assumptions   Posterior trace and density plots were used to visually diagnose the variance and autocorrelation of each iterative value in the chains. The *autocorr.diag* functions were used to assess the autocorrelation of the chain in further detail. The *raftery.diag* function was used to diagnostically determine the optimal sample size for the final chain (sample size of 4,000).

3. Results

3.1 General Results

At the end of August 2017, 71 out of the initial 75 installed nests were occupied, with 37 occupied blocks in natural habitats and 34 occupied blocks in orchard habitats by various species of bees and wasps.

3.2 Statistical Results

3.2.1 Frequentist Results

Following the GLMM, in *Figure 2*, we see that orchard sites do not have a significant effect on *Osmia* spp. survival ($p = 0.883$) and that as a result, there is no detectable difference among survival ratios in orchard sites when compared to natural sites. This applies not only to the GLMM ouput in *Figure 2a* but also the means of the raw data in *Figure 2b*.

The GLMM output shows that when controlling for differences in site and block characteristics, there is no evidence that there is a difference in *Osmia* spp. larvae survival among orchard and natural habitat types.

Most importantly, the likelihood ratio testing for the predictive value of site type shows a marginal difference in AIC of 1.97 and a p-value of 0.882. Because of this, it is most likely that site type has little predictive influence on *Osmia* spp. survival.
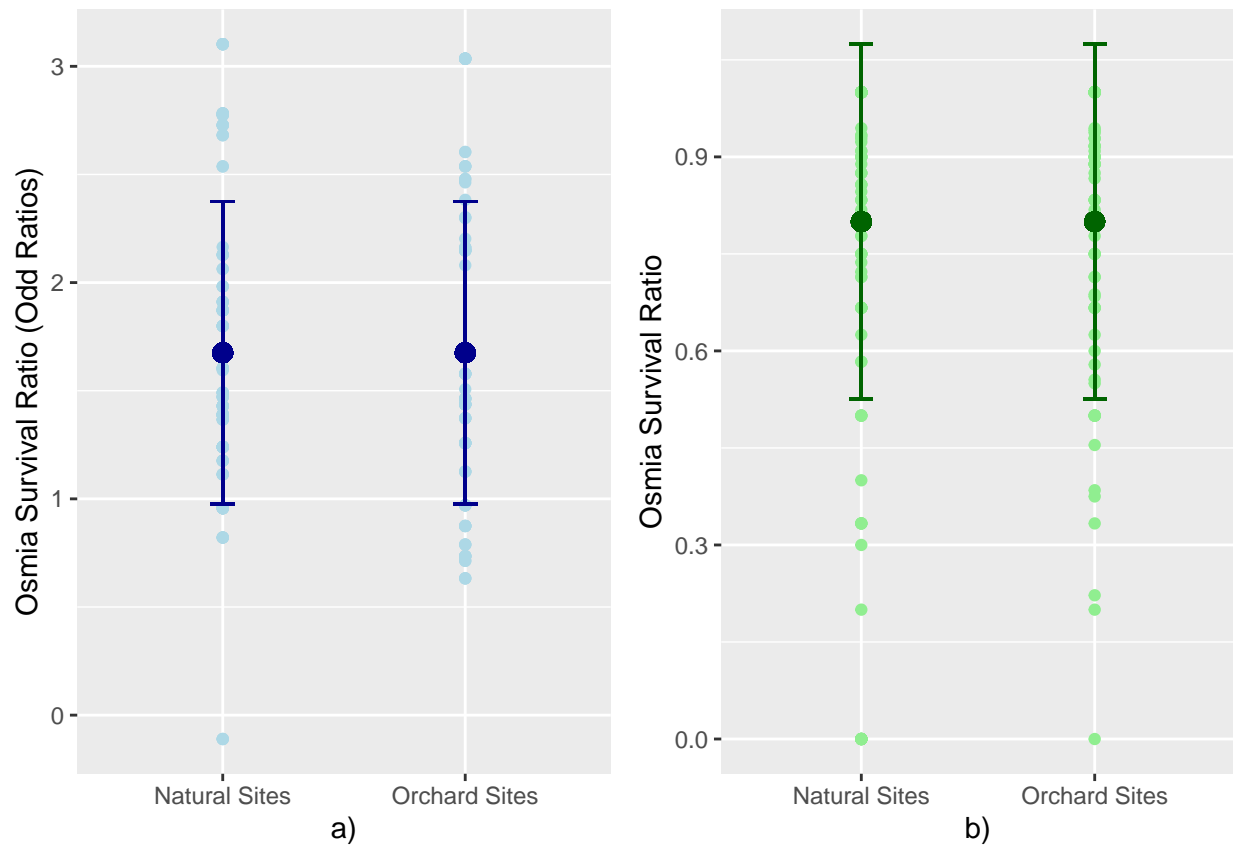
Figure 3: *Osmia* spp. survival versus habitat type with a) GLMM estimates using odds ratio and b) untransformed means ± standard deviation (raw data)

Observations	219
Dependent variable	No.surv.osmia/No.cells
Type	Mixed effects generalized linear model
Family	binomial
Link	logit

AIC	765.46
BIC	779.02
Pseudo-R² (fixed effects)	0.00
Pseudo-R² (total)	0.19

Fixed Effects					
	Est.	2.5%	97.5%	z val.	p
(Intercept)	1.78	1.34	2.21	8.03	0.00
s.typeo	0.05	-0.58	0.67	0.15	0.88

Random Effects		
Group	Parameter	Std. Dev.
Block_ID:s.ID	(Intercept)	0.84
s.ID	(Intercept)	0.25

Grouping Variables		
Group	# groups	ICC
Block_ID:s.ID	57	0.17
s.ID	15	0.02

```
## Single term deletions
##
## Model:
## No.surv.osmia/No.cells ~ s.type + (1 | s.ID/Block_ID)
##          npar    AIC      LRT Pr(Chi)
## <none>        765.46
## s.type   1 763.49 0.022051   0.882
```

3.2.1.1 Key Assumptions   In terms of model fit, when comparing the mixed effect model (GLMM) with the model without the random effect of site ID (GLM) in a Wald's Chi-Square test, model fit is improved (GLMM AIC of 765.46 versus GLM AIC of 820.78). This confirms that a hierarchical model taking into account the inherent structure of the data is better suited.

```
## Data: Osmia
## Models:
## osmiaglm: No.surv.osmia/No.cells ~ s.type
## osmiaglmm: No.surv.osmia/No.cells ~ s.type + (1 | s.ID/Block_ID)
```

```
##          npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## osmiaglm     2 820.78 827.56 -408.39   816.78
## osmiaglmm    4 765.46 779.02 -378.73   757.46 59.319  2  1.315e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

To test the effect of nesting the random effects of site and block ID on model fit, the Wald's Chi-Square test demonstrates that the nested random value causes AIC values to increase from 198.03 to 765.46. All AIC comparisons are summarized in the table below.

	Final GLMM	Un-Nested GLMM
Intercept	1.78 ***	1.56 ***
	(CI [1.34, 2.21]	(CI [1.10, 2.02]
Orchard Site Type	0.05	0.24
	(CI [-0.58, 0.67]	(CI [-0.50, 0.99]
N	219	219
N (Block_ID:s.ID)	57	
N (s.ID)	15	15
AIC	765.46	198.03
BIC	779.02	208.20
R2 (fixed)	0.00	0.00
R2 (total)	0.19	0.00

*** p < 0.001; ** p < 0.01; * p < 0.05.

```
## Data: Osmia
## Models:
## osmiaglmmtest: No.surv.osmia/No.cells ~ s.type + (1 | s.ID)
## osmiaglmm: No.surv.osmia/No.cells ~ s.type + (1 | s.ID/Block_ID)
##               npar    AIC    BIC  logLik deviance Chisq Df Pr(>Chisq)
## osmiaglmmtest    3 198.03 208.20  -96.02   192.03
## osmiaglmm        4 765.46 779.02 -378.73   757.46     0  1          1
```

However, despite this supposed improved model fit, this so-called simpler model has issues with convergence. A test checking for the singularity of the model fit (r model output below) shows that some of the constrained parameters of the random effects are equal to 0 for the un-nested model. This implies a higher probability of false positives and that the model has mis-converged due to optimization issues.

```
tt <- getME(osmiaglmmtest,"theta")
ll <- getME(osmiaglmmtest,"lower")
min(tt[ll==0])
```

```
## [1] 0
```

When looking at the issue of over-dispersion, we find that the GLM had a deviance of 572.27 over 217 degrees of freedom. The simpler GLMM with one random effect had issues of under-dispersion with a residual deviance of 108.10 over 216 degrees of freedom. The nested GLMM had a deviance of 413.40 over 215 degrees of freedom. The final GLMM model with block ID nested within site ID as a random effect had a ratio of deviance to degrees of freedom of 1.9, close to the accepted value of 1.

```
##
## Call:
## glm(formula = No.surv.osmia/No.cells ~ s.type, family = binomial(link = "logit"),
##     data = Osmia, weights = No.cells)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -5.6454  -0.6845   0.6313   1.4834   2.6942
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.5839     0.0918   17.25   <2e-16 ***
## s.typeo      -0.0851     0.1290   -0.66    0.509
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 576.71  on 218  degrees of freedom
## Residual deviance: 576.28  on 217  degrees of freedom
## AIC: 820.78
##
## Number of Fisher Scoring iterations: 4


## Residual deviance: 108.105 on 216 degrees of freedom (ratio: 0.5)


## Residual deviance: 413.402 on 215 degrees of freedom (ratio: 1.923)
```

When taking into account the fact that the nested random effect model better matches the data-set structure and the fact that it did not have any convergence warnings or issues with under-dispersion, I will choose the nested random effect model as my final model and by focusing the remaining GLMM assumptions on this final model. Therefore, returning to the initial assumptions of a GLMM that the residuals of the random effects should be normal and homoscedastic, in **Figure 4**, the residuals are constant throughout all levels of the random effect. Moreover, the q-q plots in **Figure 5** show that the nested random effects of site and block ID are appropriately normal.

Having considered model fit, data structure, dispersion of the variance, issues with convergence, normality of the residuals and normality of the random effects, the GLMM with nested random-effects is the best model for my data and adequately meets the major assumptions of a GLMM.

3.2.2 Bayesian Results

Similar to the frequentist results, we find that there is not enough evidence from the data collected to show that site type has an effect on *Osmia* spp. survival since the credible interval [-0.79, 1.02] overlaps 0. This can also be surmised from **Figure 6**.
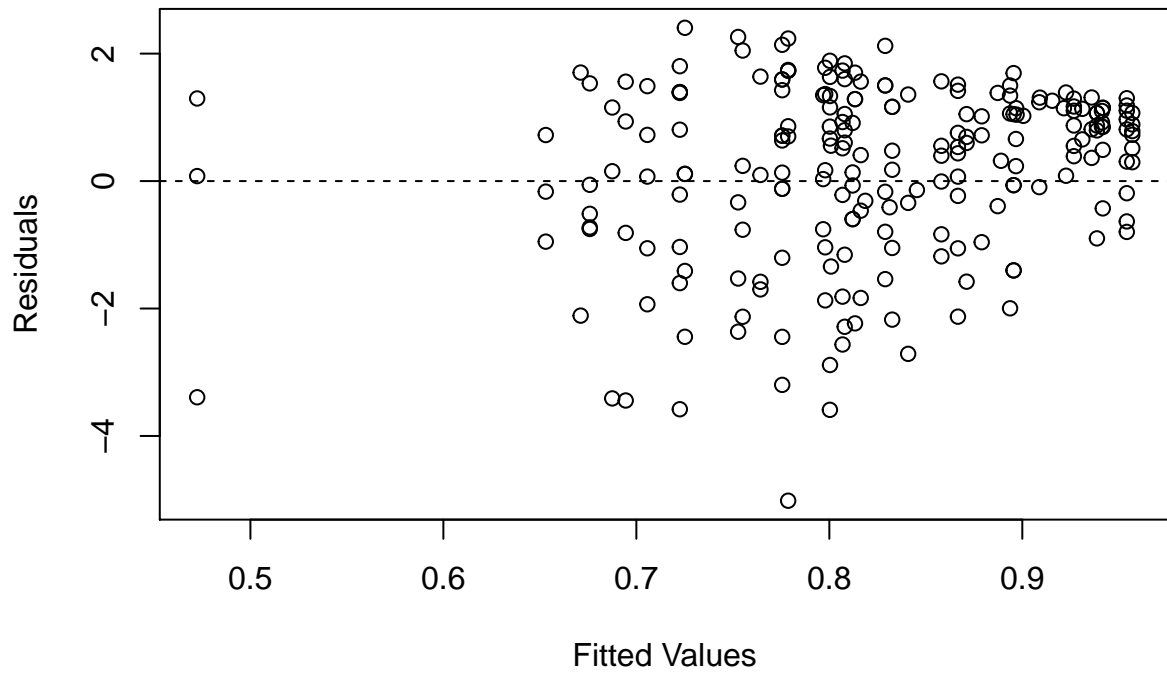
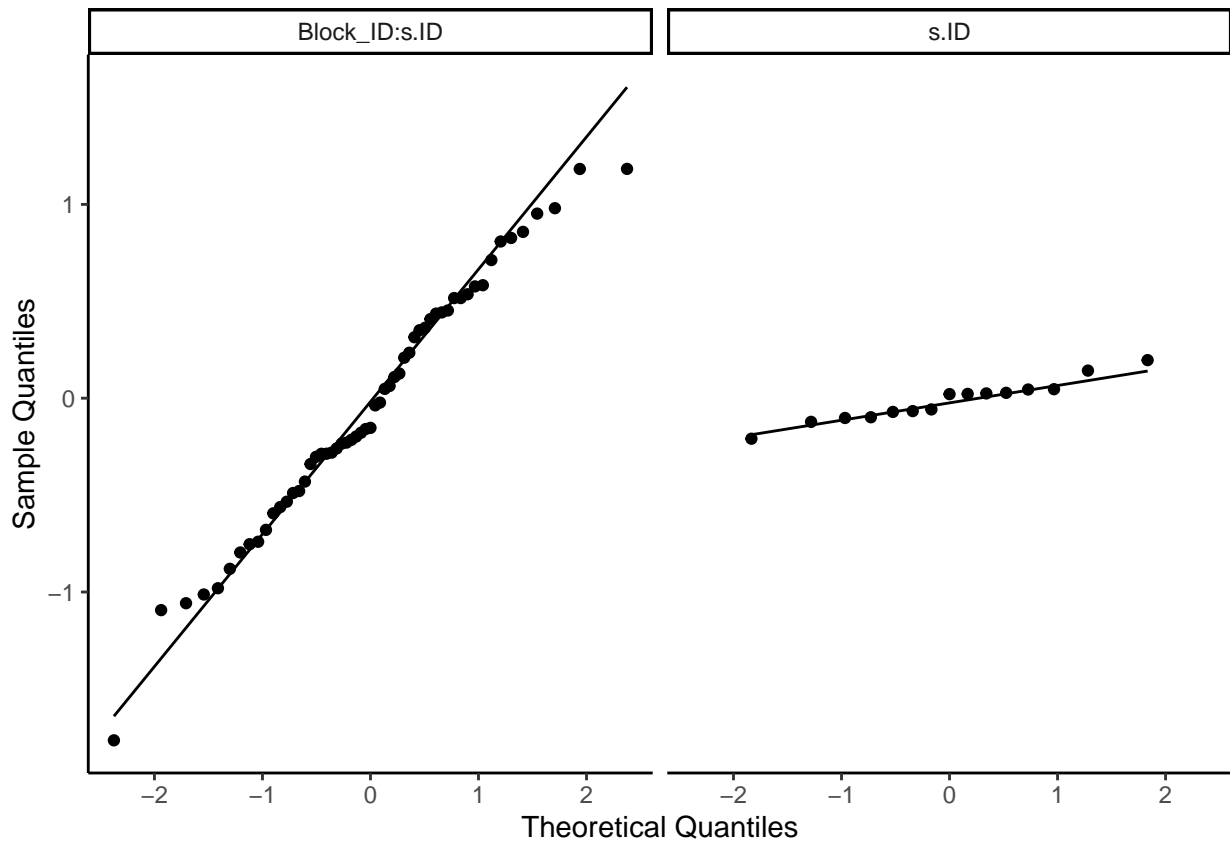Figure 4: Residuals versus Fitted Values of Final Osmia Survival Model

Figure 5: Q-Q Plots of Nested Random Effects of Block and Site ID (BLUPs)

12

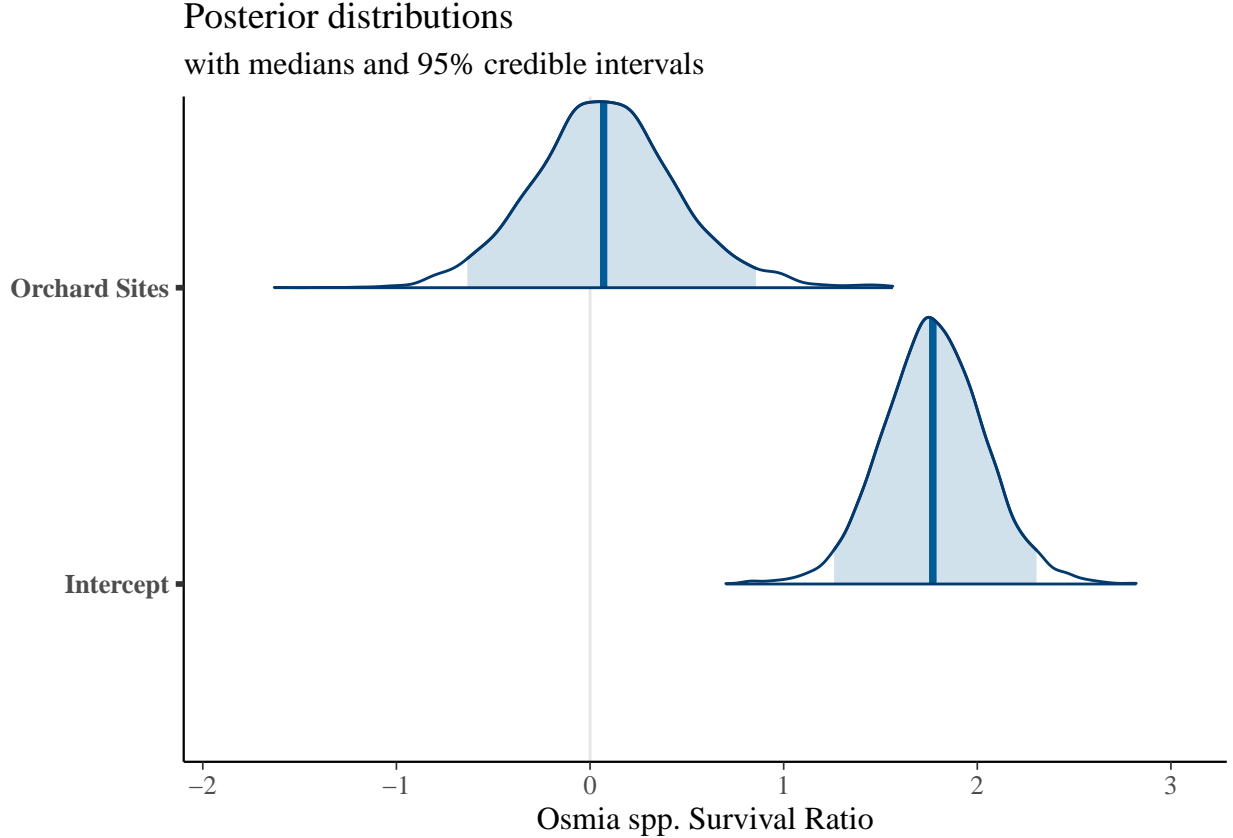| Parameter	Posterior Mode	Lower Credible Interval	Upper Credible Interval
Intercept	2.02	1.42	2.67
Orchard Sites	0.08	-0.79	1.02

Figure 6: Q-Q Plots of Nested Random Effects of Block and Site ID (BLUPs)

3.2.2.1 Bayesian Key Assumptions From a visual assessment of the posterior trace and density plots of the first chain (***Figures 7 and 8***), there was high evidence of autocorrelation and variability, especially for the random effects. There was also some degree of autocorrelation since most lag iterations had auto-correlation values exceeding 0.1. Moreover, the *raftery.diag* function recommended a sample size of at least 3,746. To remedy this, for the second and final chain, sample size was increased from 1,000 to 4,000, number of iterations was increased from 13,000 to 2,020,000 and priors were adjusted (V=1, nu=1).

```
autocorr.diag(osmiabayes$Sol)
```

```
##          (Intercept)       s.typeo
## Lag 0     1.000000000  1.0000000000
## Lag 10    0.211519599  0.0575871950
## Lag 50   -0.002496244 -0.0235296389
## Lag 100  -0.051142350 -0.0147581135
## Lag 500   0.016400651 -0.0007690553
```
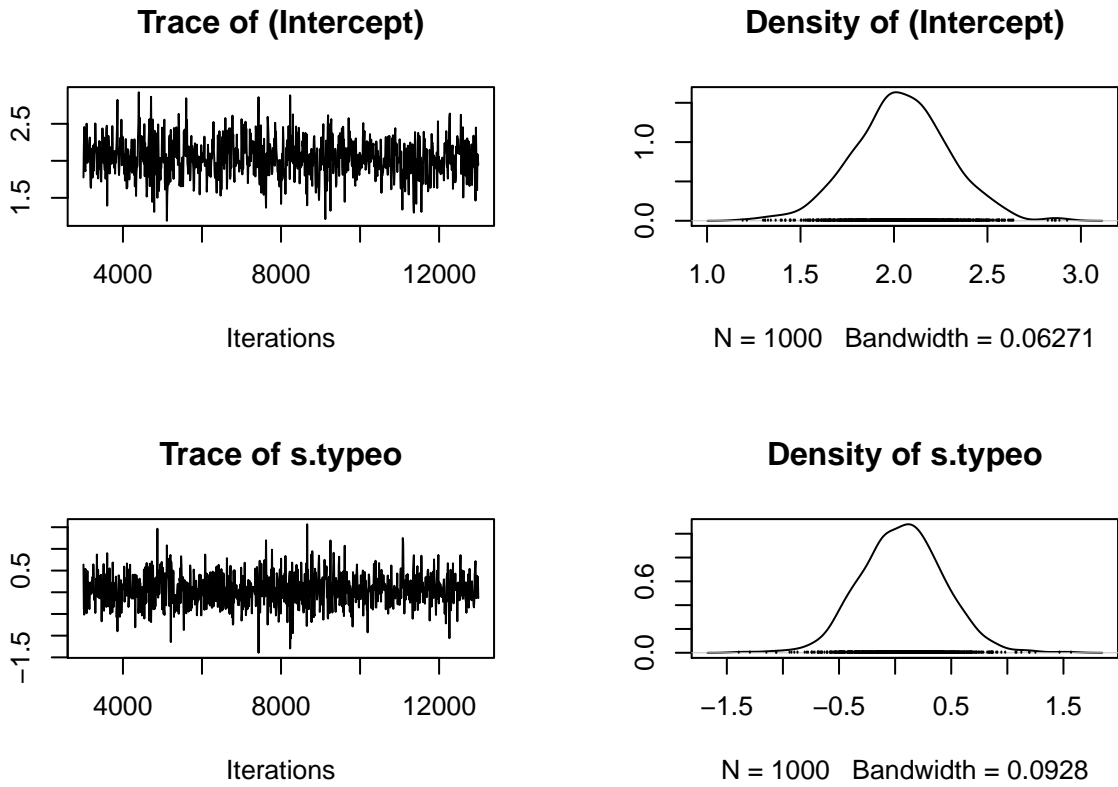
Trace of (Intercept)

Density of (Intercept)

N = 1000   Bandwidth = 0.06271

Trace of s.typeo

Density of s.typeo

N = 1000   Bandwidth = 0.0928

Figure 7: Initial Posterior MCMC Trace and Distribution of the Fixed Effects

Trace of s.ID

Density of s.ID

N = 1000   Bandwidth = 0.008085

Trace of Block_ID:s.ID

Density of Block_ID:s.ID

N = 1000   Bandwidth = 0.1185

Trace of units

Density of units

N = 1000   Bandwidth = 0.1374
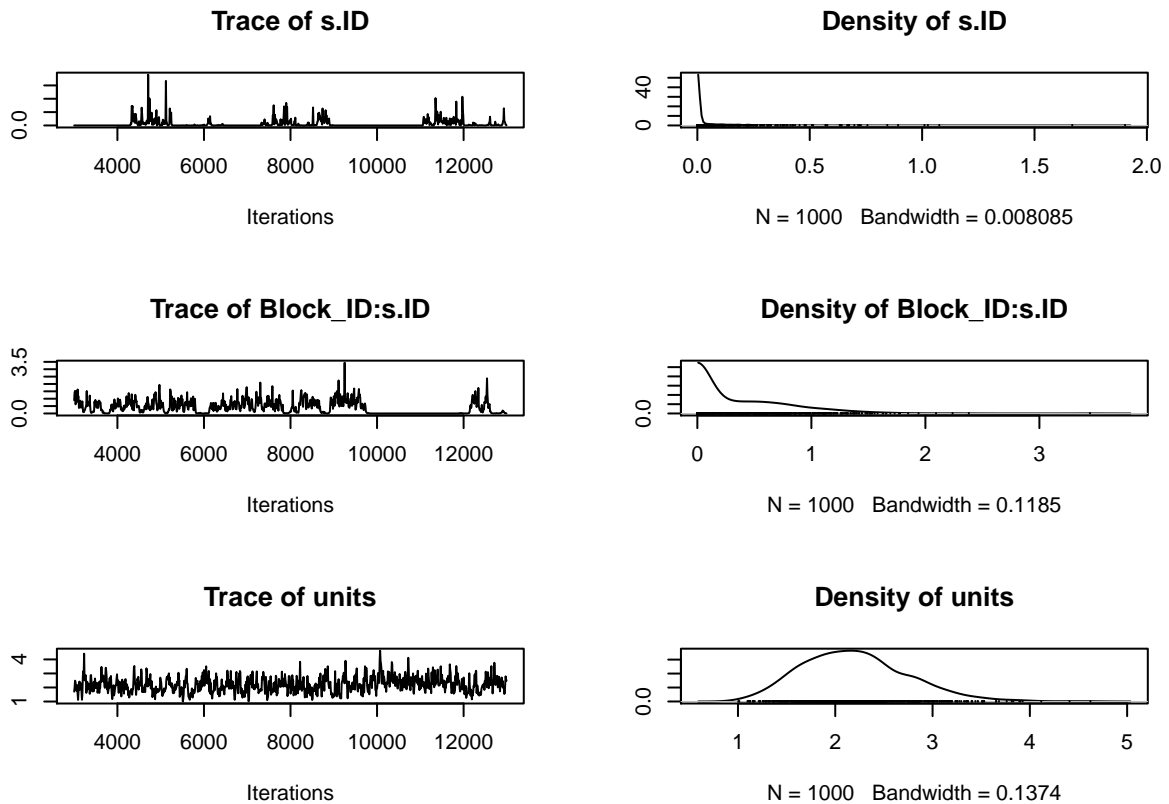
Figure 8: Initial Posterior MCMC Trace and Distribution of the Random Effects

```
autocorr.diag(osmiabayes$VCV)
```

```
##              s.ID Block_ID:s.ID      units
## Lag 0   1.00000000    1.0000000 1.000000000
## Lag 10  0.41361927    0.7388257 0.529968729
## Lag 50  0.21600491    0.4916362 0.084506649
## Lag 100 0.17975037    0.3649395 0.002904702
## Lag 500 0.08774687    0.1136287 0.038242094
```

```
raftery.diag(osmiabayes)
```

```
##
## Quantile (q) = 0.025
## Accuracy (r) = +/- 0.005
## Probability (s) = 0.95
##
## You need a sample size of at least 3746 with these values of q, r and s
```

A visual assessment of the modified posterior trace and density plots (**Figures 9 and 10**) show that there is no identifiable pattern in the chain as what was evident in the first chain and much less proof of autocorrelation.
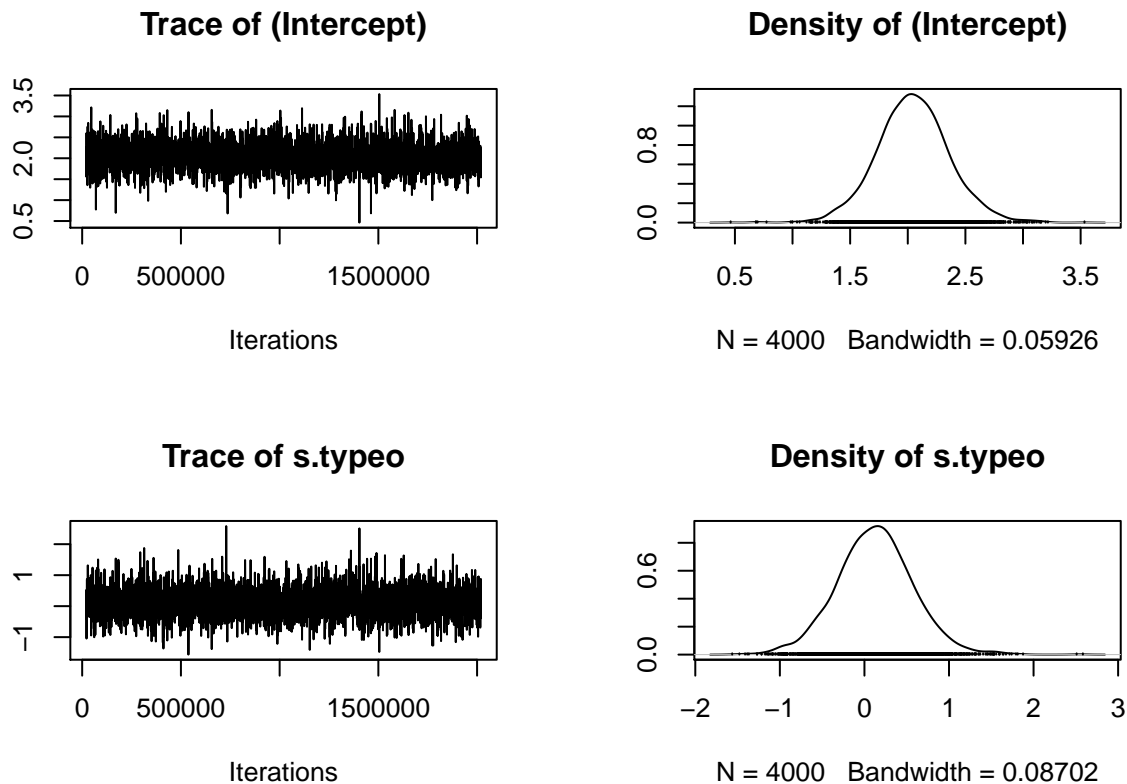

Figure 9: Final Posterior MCMC Trace and Distribution of the Fixed Effects

This is further confirmed by the diagnostic autocorrelation test in which all autocorrelation values are well below the 0.1 threshold.
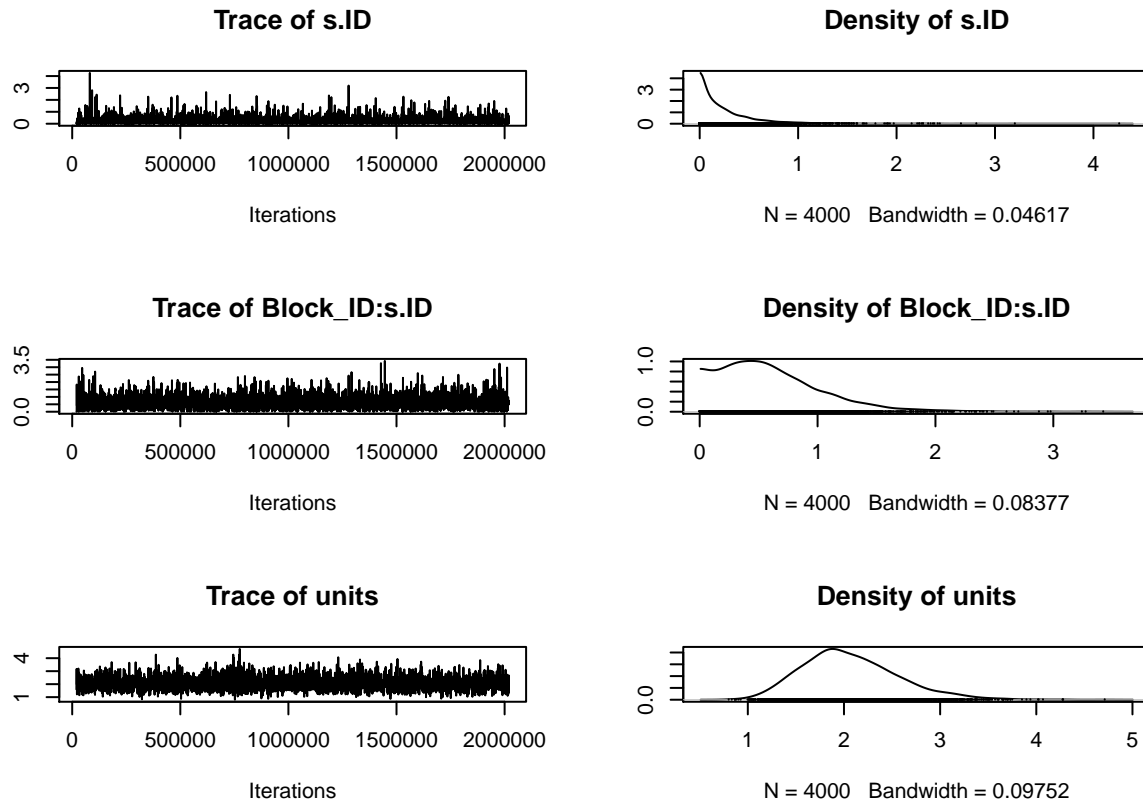
Figure 10: Final Posterior MCMC Trace and Distribution of the Random Effects

```
autocorr.diag(osmiabayesfinal$Sol)
```

```
##              (Intercept)        s.typeo
## Lag 0       1.000000000   1.000000000
## Lag 500     0.024811375   0.009443638
## Lag 2500   -0.026121203  -0.016447871
## Lag 5000    0.010804614   0.006192597
## Lag 25000   0.005513838   0.015610974
```

```
autocorr.diag(osmiabayesfinal$VCV)
```

```
##                    s.ID Block_ID:s.ID        units
## Lag 0       1.000000000    1.00000000  1.000000000
## Lag 500     0.018129524   -0.00268481  0.018571428
## Lag 2500   -0.026162952   -0.02349306  0.002303032
## Lag 5000    0.016883618   -0.01046079  0.014991384
## Lag 25000   0.008939716    0.01968110 -0.040399899
```

4. Discussion

4.1 Summary

Based on the data collected and subsequent statistical analyses, I can conclude that there is no evidence of a difference between the survival ratio of *Osmia* spp. in orchard and natural sites. This is evident when simply looking at the raw *Osmia* spp. survival ratios in natural versus orchard sites (***Figure 3b***). When correcting for differences in sites (s.ID) and blocks (Block_ID) in order to account for any noise or variation in the data, this lack of a difference among site types is still prevalent (***Figure 3a***).

However, these conclusions do not take into account the wider implications of land-use change on pollinators. For instance, species richness has been shown to be heavily influenced by percent agriculture, orchard size and most notably, sampling effort (Russo et al. 2015). Even though *Osmia* spp. survival rates in orchards have not been shown to differ from natural sites in my analysis, habitat loss due to agricultural intensification is still of major concern in the declines of native bee populations (Kline and Joshi 2020).

A possible explanation for this discrepancy in results is that the apple orchard sites included in this study are not as agriculturally intense as other fruit orchards (i.e. almond orchards in California (Koh et al. 2017)) and thus, are not as isolated from the floral and nesting resources required for wild bee population survival and growth. Thus, the difference among these orchard and natural sites are not as pronounced as in other studies.

4.2 Implications

While further investigation is needed, this could possibly indicate that the installation of nesting structures pose little risk to the development and survival of *Osmia* spp. offspring and most likely, will not act a population sink. This means that in my thesis experiment, testing the efficacy of nesting structures in orchard sites, the nesting structures for my experiment will most likely not pose an ecological threat to wild bee populations either as an ecological trap or population sink. Furthermore, the issue of promoting a resource that is potentially a population sink will most likely not have an impact on any of the conclusions that I make relating to the efficacy of the nesting structures for boosting pollinator populations.

4.3 Limitations of Analysis

Because the hierarchical nature of the data was acknowledged, I was able to avoid issues of autocorrelation within the data structure and issues of non-independent residuals which will lead to an increased rate of false positives (McNeish and Stapleton 2014). However, it must be acknowledged that the number of sites visited in the orchard and natural habitat types is very small (7 and 8, respectively), which overall, leads to bias in the results. It is typically recommended to have a minimum cluster sample size of 30 which is well below the 15 sites or clusters from this data set. Some authors even argue that mixed models become potentially untrustworthy with small sample sizes and should not be run if cluster sizes are below 10 (McNeish and Stapleton 2014). A power analysis using the *simr* package shows that the power to reject the null hypothesis of no effect of site type on *Osmia* survival given this particular set-up is 5.6% which is very low especially when considering that most studies aim for a power of 80% (Green and MacLeod 2016). This low power is most likely due to a small sample size.

```
osmia_sim <- powerSim(osmiaglmm, nsim=1000)
```

Power for Site Type
5.60%
Based on 1000 simulations, alpha = 0.05, n =219

When it comes to a critique of the data, it would have been informative to have more data relating to the cause of death since having this value as a fixed effect would have accounted for some of the over-dispersion detected

in the final GLMM. Controlling for cause of death would have also helped with recognizing overarching factors in larval mortality that are not necessarily related to habitat type but rather to normal threats to larval development common across all landscapes. This would have helped with pinpointing the true differences among habitat types which can be overshadowed by these potentially common observable factors. From a different perspective, it would have also helped to determine whether death due to parasitization is more common in orchard habitat types or natural habitat types since this could give some by-proxy insight into parasite population densities.

4.4 Future Directions

In the future, it would be useful to rerun this experiment with a larger sample size to address the issue with power. By extension, it would also be interesting to collect some information on *Osmia* spp. survival in urban and suburban habitats in addition to both natural and agricultural habitats since data on *Osmia* spp. populations in these human-dominant landscapes is scarce. Cross-referencing the likelihood of *Osmia* spp. survival in these different habitats can help with re-evaluating the usefulness of installing these nesting structures in these different habitat types from a conservation perspective. For instance, if mortality rates are high in suburban landscapes when compared to natural landscapes, nesting structures may be better suited to orchard landscapes where larval mortality has not been shown to differ from natural areas.

Moreover, it would be interesting to expand observations to include other bee and wasp genera and to document their survival. In addition, while logistically complicated, since it would involve extensive and detailed monitoring of nesting structures throughout the growing season, documenting and comparing the successful attainment of key developmental stages such as hatching, pupation and eclosion across landscape types could be an interesting extension to this question on successful *Osmia* spp. reproduction in anthropogenic environments.

Data Availability Statement

All data is available in the **Final Report GitHub Repository**.

References

Bihaly, Áron Domonkos, Anikó Kovács-Hostyánszki, Márk Szalai, and Miklós Sárospataki. 2020. "Nesting Activity of Cavity-Nesting Bees and Wasps Is Lower in Small-Scale Apple Orchards Compared to Nearby Semi-Natural Habitats." *Agricultural and Forest Entomology* 23 (1): 49–58. https://doi.org/10.1111/afe.12403.

Boyle, N. K., and T. L. Pitts-Singer. 2017. "The Effect of Nest Box Distribution on Sustainable Propagation of Osmia Lignaria (Hymenoptera: Megachilidae) in Commercial Tart Cherry Orchards." *Journal of Insect Science* 17 (2). https://doi.org/10.1093/jisesa/iex008.

Boyle, Natalie K., and Theresa L. Pitts-Singer. 2019. "Assessing Blue Orchard Bee (Osmia Lignaria) Propagation and Pollination Services in the Presence of Honey Bees (Apis Mellifera) in Utah Tart Cherries." *PeerJ* 7 (September): e7639. https://doi.org/10.7717/peerj.7639.

Földesi, Rita, Anikó Kovács-Hostyánszki, Ádám Kőrösi, László Somay, Zoltán Elek, Viktor Markó, Miklós Sárospataki, et al. 2015. "Relationships Between Wild Bees, Hoverflies and Pollination Success in Apple Orchards with Different Landscape Contexts." *Agricultural and Forest Entomology* 18 (1): 68–75. https://doi.org/10.1111/afe.12135.

Geslin, Benoît, Sophie Gachet, Magali Deschamps-Cottin, Floriane Flacher, Benjamin Ignace, Corentin Knoploch, Éric Meineri, et al. 2020. "Bee Hotels Host a High Abundance of Exotic Bees in an Urban Context." *Acta Oecologica* 105 (May): 103556. https://doi.org/10.1016/j.actao.2020.103556.

Green, Peter, and Catriona J. MacLeod. 2016. "SIMR : An R Package for Power Analysis of Generalized Linear Mixed Models by Simulation." Edited by Shinichi Nakagawa. *Methods in Ecology and Evolution* 7 (4): 493–98. https://doi.org/10.1111/2041-210x.12504.

Isaacs, Rufus, Neal Williams, James Ellis, Theresa L. Pitts-Singer, Riccardo Bommarco, and Mace Vaughan. 2017. "Integrated Crop Pollination: Combining Strategies to Ensure Stable and Sustainable Yields of Pollination-Dependent Crops." *Basic and Applied Ecology* 22 (August): 44–60. https://doi.org/10.1016/j.baae.2017.07.003.

Klein, Alexandra-Maria, Bernard E Vaissière, James H Cane, Ingolf Steffan-Dewenter, Saul A Cunningham, Claire Kremen, and Teja Tscharntke. 2006. "Importance of Pollinators in Changing Landscapes for World Crops." *Proceedings of the Royal Society B: Biological Sciences* 274 (1608): 303–13. https://doi.org/10.1098/rspb.2006.3721.

Kline, Olivia, and Neelendra K. Joshi. 2020. "Mitigating the Effects of Habitat Loss on Solitary Bees in Agricultural Ecosystems." *Agriculture* 10 (4): 115. https://doi.org/10.3390/agriculture10040115.

Koh, Insu, Eric V Lonsdorf, Derek R Artz, Theresa L Pitts-Singer, and Taylor H Ricketts. 2017. "Ecology and Economics of Using Native Managed Bees for Almond Pollination." *Journal of Economic Entomology* 111 (1): 16–25. https://doi.org/10.1093/jee/tox318.

MacIvor, J. Scott, and Laurence Packer. 2015. "'Bee Hotels' as Tools for Native Pollinator Conservation: A Premature Verdict?" Edited by Fabio S. Nascimento. *PLOS ONE* 10 (3): e0122126. https://doi.org/10.1371/journal.pone.0122126.

McNeish, Daniel M., and Laura M. Stapleton. 2014. "The Effect of Small Sample Size on Two-Level Model Estimates: A Review and Illustration." *Educational Psychology Review* 28 (2): 295–314. https://doi.org/10.1007/s10648-014-9287-x.

Potts, Simon G., Jacobus C. Biesmeijer, Claire Kremen, Peter Neumann, Oliver Schweiger, and William E. Kunin. 2010. "Global Pollinator Declines: Trends, Impacts and Drivers." *Trends in Ecology & Evolution* 25 (6): 345–53. https://doi.org/10.1016/j.tree.2010.01.007.

Roulston, T'ai H., and Karen Goodell. 2011. "The Role of Resources and Risks in Regulating Wild Bee Populations." *Annual Review of Entomology* 56 (1): 293–312. https://doi.org/10.1146/annurev-ento-120709-144802.

Russo, Laura, Mia Park, Jason Gibbs, and Bryan Danforth. 2015. "The Challenge of Accurately Documenting Bee Species Richness in Agroecosystems: Bee Diversity in Eastern Apple Orchards." *Ecology and Evolution* 5 (17): 3531–40. https://doi.org/10.1002/ece3.1582.

Westerfelt, Per, Jan Weslien, and Olof Widenfalk. 2018. "Population Patterns in Relation to Food and Nesting Resource for Two Cavity-Nesting Bee Species in Young Boreal Forest Stands." *Forest Ecology and Management* 430 (December): 629–38. https://doi.org/10.1016/j.foreco.2018.08.053.