Advanced Stats - Bank Swallow Movement

2023-04-24

## BIO8940 Final Project

### Bank swallow movements to roosting habitat

### By: Manon Veselovsky - 8259387

## Introduction:

### Various animals across taxa are known to form aggregate groups during their lifecycle. This phenomenon can become problematic when faced with rapid environmental change such as habitat modification. and the alteration of suitable habitats where this clustering behaviour is centered. For example, during the winter, the eastern migratory monarch butterfly population forms a single roosting colony of millions of individuals in a very small region of oyamel forest in Mexico. While many factors are contributing to declining monarch butterfly populations, one of the most concerning is the human alteration of the oyamel forests which provide this roosting habitat (\_\_\_\_\_\_\_\_\_\_\_\_\_). Today, only \_\_\_\_\_\_\_\_\_\_\_ remains of the original \_\_\_\_\_\_\_\_\_\_ which could house these overwintering insects. Unfortunately, habitat alteration and loss is only projected to increase as human infrastructure expands, which is likely to continue to threaten certain habitats and species that utilize these habitats. For conservation of habitats to be effective, it is important to understand the degree to which these habitats are being used by endangered species, especially when endangered species aggregate in these areas.

### One highly threatened species whose aggregation ecology requires greater study is the bank swallow (*Riparia riparia*). Much of the North American bank swallow’s behavioural ecology remains unknown, a shortcoming that must be addressed if this species is to be conserved. Bank swallows form breeding colonies on the banks of lakes and in gravel pits across North America. Currently, the threatened status of these birds in Ontario means that 50 m surrounding their breeding colonies must be protected from alteration (Bank Swallow General Habitat Description | Ontario.Ca, n.d.). However, it has been discovered that bank swallows not only form aggregate groups in breeding colonies, but also congregate at night to roost on wetland vegetation. Unfortunately, the use of wetlands at night by these birds has not been taken into consideration in their habitat protection plan due to the recent discovery of this phenomenon. Indeed, much remains unknown about the movement ecology of these birds, and studies must be conducted to evaluate the degree of wetland use by these birds to inform recovery strategies. Currently, we have limited data on the proportiono of nights bank swallows use wetland roosts during the breeding period. Preliminary studies indicate that bank swallow roosting behaviour is negatively associated with night time light availability, and that males utilize wetland roosts more than females during the early breeding period (@Falconer2016; @Saldanha2019). Unfortunately, these studies were only able to evaluate the movement behaviour of a small number of individuals, limiting the conclusions that could be drawn. For recovery plans to be effective it is important to identify the degree to which bank swallows are roosting in wetlands and determine what conditions may modify this roosting behaviour.

### In this paper, I will use a database of 52 bank swallows observed on 49 nights to explore bank swallow roosting behaviour. The results of this paper will determine how often bank swallows are using major wetlands to roost at night. If bank swallows are using wetland roosts for a majority of nights, and are only using wetland roosts when they are close to their colony, two conclusions can be drawn: (1) Wetlands are likely an important habitat for these birds, and (2) wetland habitats close to these colonies should be taken into protection consideration for the overall recovery of the swallows.

### While a hypothesis cannot be tested with this data and our current state of knowledge, predictions can be made. Firstly, based on the tentative results of previous studies, I predict that nighttime light availibility will negatively affect roosting probability. Secondly, I predict that males will use wetland roosts proportionately more in the early breeding period than females.

## Methods:

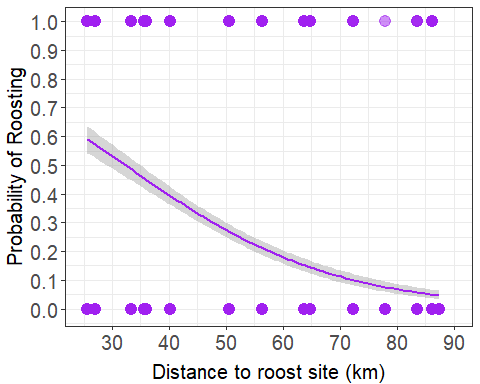
### In summer 2015, 52 bank swallows were captured, sexed, and tagged at breeding colonies along the shore of Lake Erie and in sand and gravel pits near Long Point, Ontario. Automated radio telemtry systems were also installed at each colony and throughout the Long Point wetland complex. Between June 2nd and July 21st, 2015, these birds were tracked for movement between their colony and the main wetland complex. These data were then visually analyzed and a “roost” value assigned to each bird’s evening movement (0 = did not roost in wetland, 1 = roosted in wetland). Birds that were detected in the wetland complex in the late evening and early the following morning were assigned a roost value of 1. Birds that were detected at their colony in the late evening and early the following morning were assigned a roosting value of 0. Any nights where birds were not detected in the same location both in the morning and the evening were removed from the dataset.

### Environmental conditions were collected using the package RNCEP. The time and location of birds in the evening were used with this package to determine the evening’s wind component vectors (N/S and E/W), the illumated proportion of the moon (1 being full moon, 0 being new moon), and the percentage cloud cover. While the effect of tailwind has not been evaluated for bank swallow roost movements, a study in New Brunswick indicated that light availability (a combination of moon illumination and cloud cover) affected the roosting movement behaviour of bank swallows (@Saldanha2019). Consequently, I multiplied the proportion of moon illumination with the proportion of open sky (1 - proportion of cloud cover) to get the light availability at the time of bird detection. Finally, RNCEP was used to calculate the tailwind flow assistance on the evening a bird was observed. The degree of flow assistance was determined using the RNCEP function NCEP.Tailwind(). Using the north-south and east-west wind components on the evening of observation, and the direction from the bird’s colony to the mean wetland location, the airflow assistance for that direction of travel was determined. The resulting tailwind variable takes into account degree of air movement parallel to the movement of the bird to give a unitless value, with negative values being airflow against the animal’s direction of travel and positive being airflow in the animal’s direction of travel.

### Additional variables that are included in my analysis are distance,

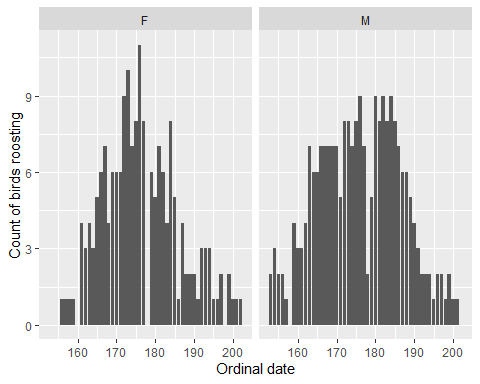
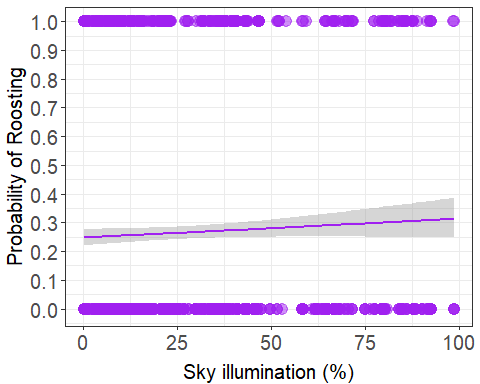
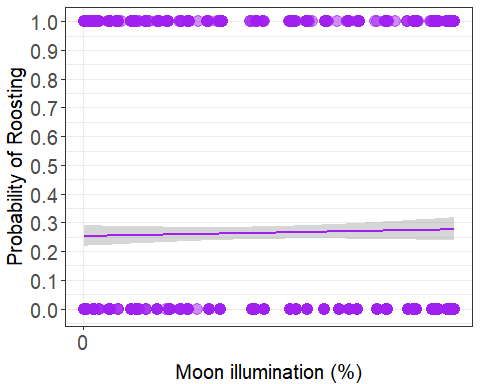
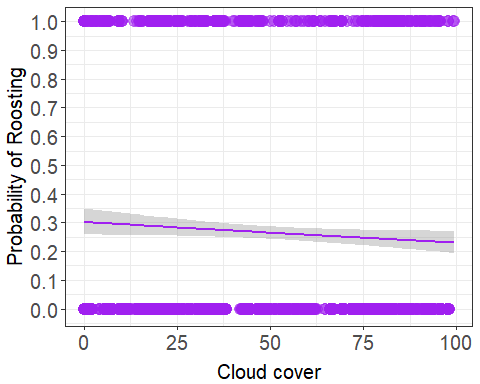
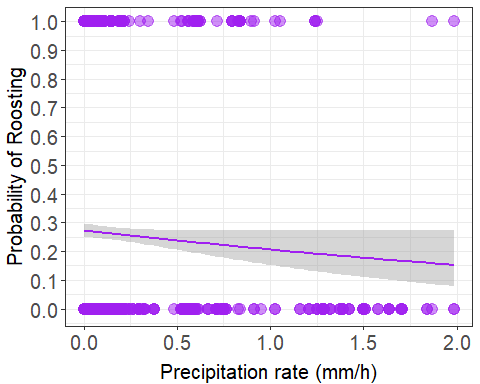
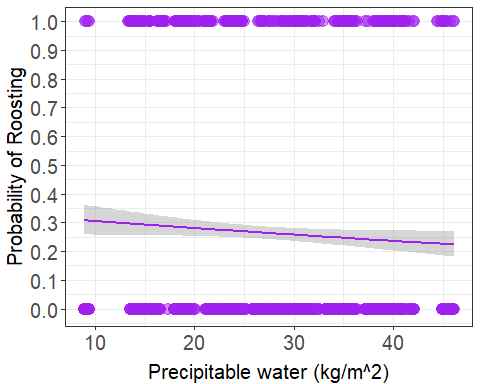
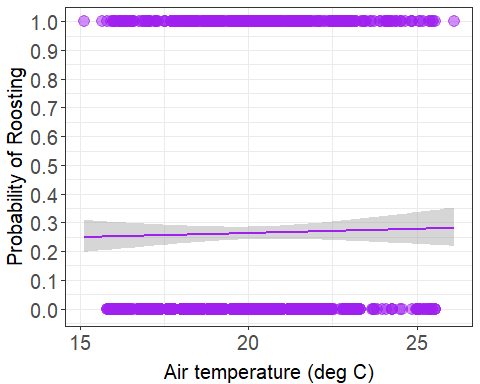
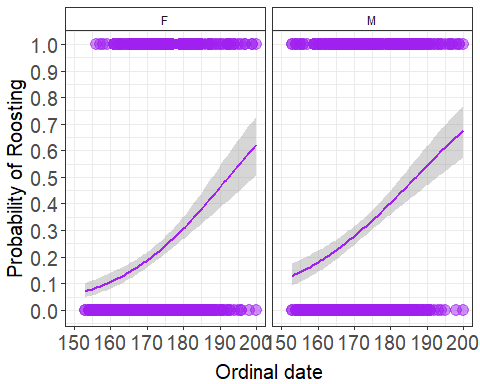
### Statistical Methods:

### I fit a generalized linear mixed effects model using a binomial distribution to which factors affect whether a bank swallow roosts overnight at the wetland (1) or remains at the colony (0). I included distance from the wetland (breeding colony location to mean wetland location), nighttime illumination, tailwind, ordinal date, sex of the individual, and the interaction of sex and ordinal date as fixed predictors. I also included the random variable of inividual ID nested within colony, as individuals were only detected at one breeding colony for the duration of monitoring, and individuals were followed on multiple nights (meaning night departure observations are not independent).



## Warning: Removed 5 rows containing non-finite values (`stat\_smooth()`).

## Warning: Removed 5 rows containing missing values (`geom\_point()`).

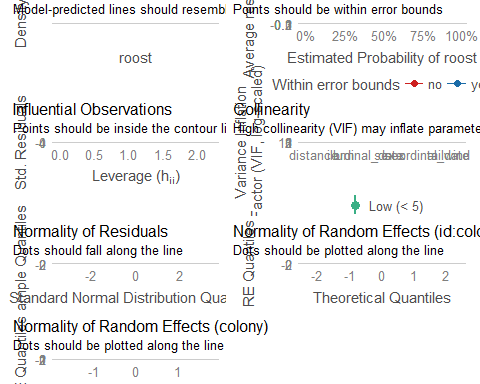
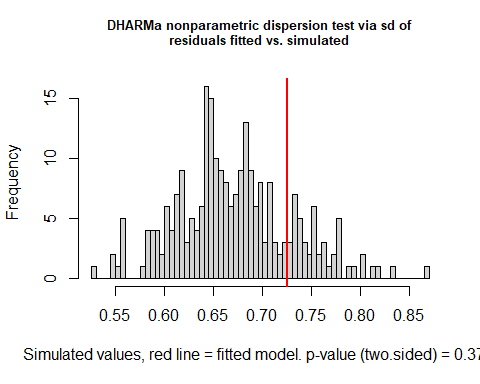


################### SECTION 1.4 MODEL BUILDING AND DIAGNOSTIC PLOTS ####################  
# model has problems using unscaled data, use df.scale (df where distance and ordinal date were scaled and centered)  
m = glmer(roost ~ distance + sex + ordinal\_date + sex:ordinal\_date + illum + tailwind + (1|colony/id), data=df.scale, family="binomial")

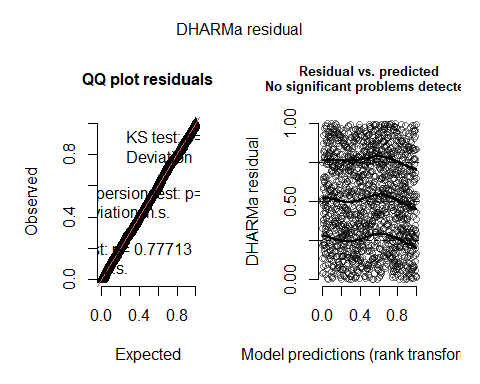
## Results:

### Overall, 52 birds were tracked between June 2nd and July 21st, 2015

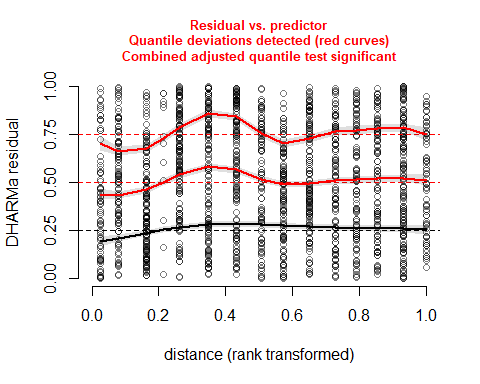
### First, I will examine the normality of the residuals from the random variables and the collinearity of fixed predictors (VIF values) using the performance package check\_model() function.

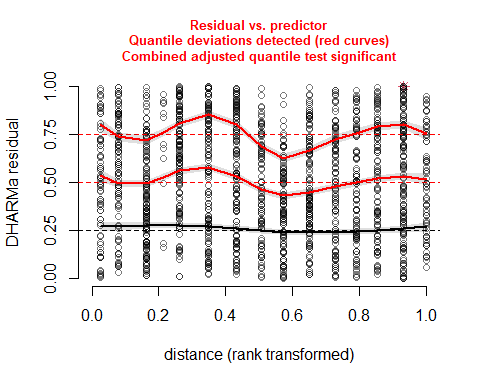
 ### VIF values are all below 3 indicating my fixed effects do not suffer from collinearity. Normality of the random effects also appears fine with no significant deviations from the plotted line (I will double check this with another plot). ### Next, I will examine the dispersion of the residuals for the overall model. This will be done using the DHARMa package. 

##   
## DHARMa nonparametric dispersion test via sd of residuals fitted vs.  
## simulated  
##   
## data: simulationOutput  
## dispersion = 1.0819, p-value = 0.376  
## alternative hypothesis: two.sided

 ### There does not appear to be any problems in the overall model. However, the overall residuals do not provide a full assessment for binomial models. I will look at the residuals for each individual variable to assess model fit and check for under or overdispersion.

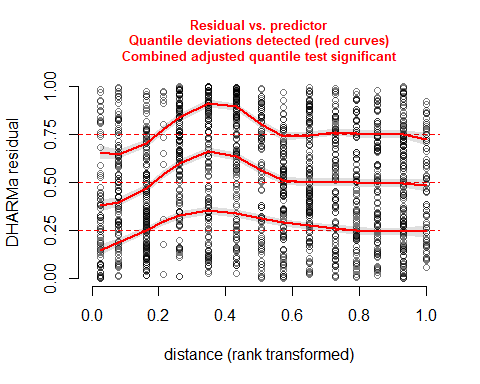
plotResiduals(dharm\_output, form=df.scale$distance) #some problem with model fit due to pattern in the residuals but not terrible, indicates possible non-linear relationship with distance

 ### There appears to be some issues in the residuals here, but it does not appear too problematic. A pattern would indicate a non-linear relationship with distance (and require a transformation), but I am not sure a particular pattern is evident. For good measure, I did try some transformations but none increased the suitability of the model fit, and many failed to converge. Given this, and that the Levene’s test may be overly sensitive with my sample size (1579 observations on 52 birds), I believe the fit is not concerning. ### I will continue with the evaluations of the residuals for the other predictors below.

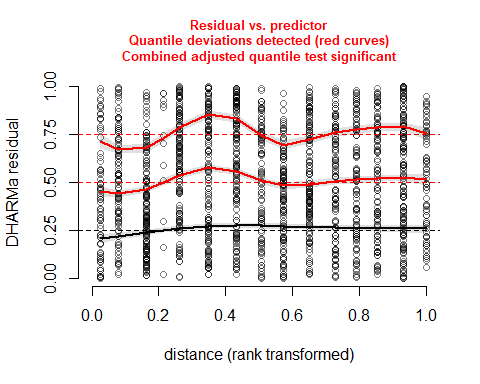


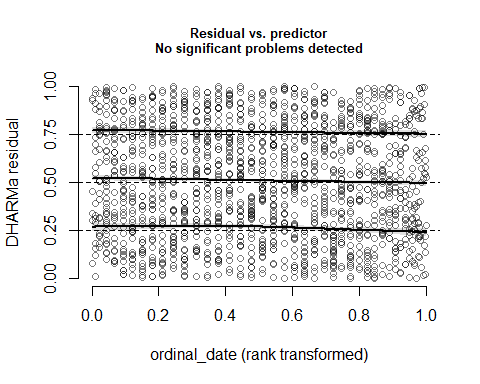
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :  
## Model failed to converge with max|grad| = 0.0733594 (tol = 0.002, component 1)

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Model is nearly unidentifiable: very large eigenvalue  
## - Rescale variables?

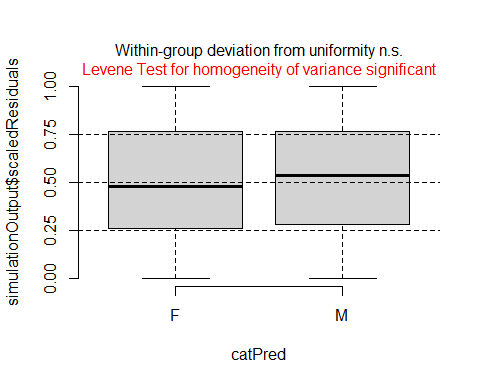


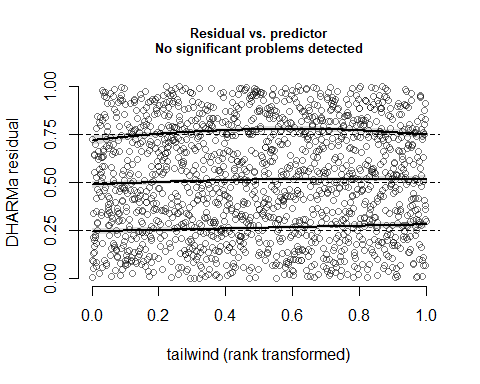
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :  
## Model failed to converge with max|grad| = 0.0103368 (tol = 0.002, component 1)

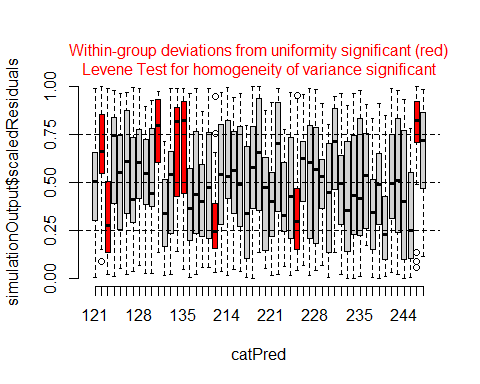


 ### Ordinal date appears to be fine.

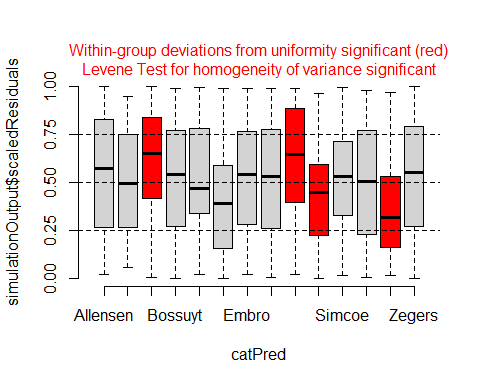
## Warning in ensurePredictor(simulationOutput, form): DHARMa:::ensurePredictor:  
## character string was provided as predictor. DHARMa has converted to factor  
## automatically. To remove this warning, please convert to factor before  
## attempting to plot with DHARMa.

 ### Sex also appears to be fine.

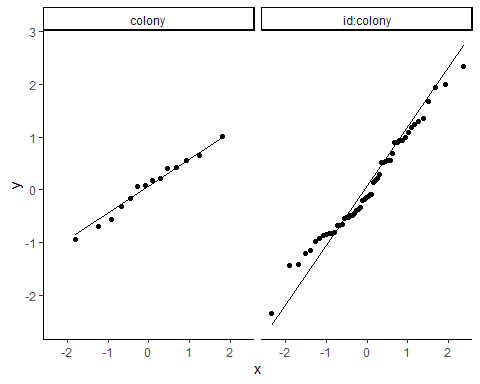
 ### Tailwind is fine.



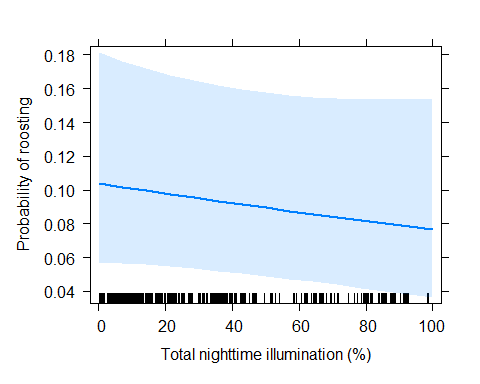
## Warning in ensurePredictor(simulationOutput, form): DHARMa:::ensurePredictor:  
## character string was provided as predictor. DHARMa has converted to factor  
## automatically. To remove this warning, please convert to factor before  
## attempting to plot with DHARMa.

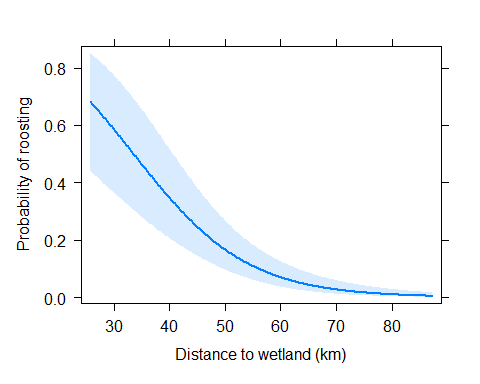
 ### Some issues in the random effects here but not to bad. Documentation on how to evaluate these issues is scarce. I will check the QQ plot for normality of the random effects.

r5 = as.data.frame(ranef(m))  
ggplot(data = r5, aes(sample = condval)) +  
 geom\_qq() + geom\_qq\_line() +  
 facet\_wrap(~ grpvar) +  
 theme\_classic()

 ### Normality of the random effects here looks good, so I will continue as the assumptions appear to be met. ###Now, I will begin evaluating the effects of the model.

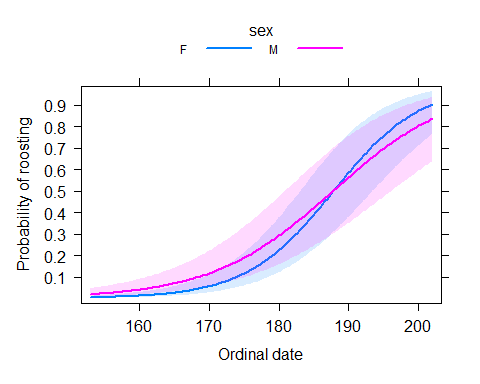
## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: roost ~ distance + sex + ordinal\_date + sex:ordinal\_date + illum +   
## tailwind + (1 | colony/id)  
## Data: df.scale  
##   
## AIC BIC logLik deviance df.resid   
## 1062.4 1110.6 -522.2 1044.4 1570   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -12.0398 -0.3228 -0.1198 0.1719 16.8109   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## id:colony (Intercept) 1.670 1.2921   
## colony (Intercept) 0.615 0.7842   
## Number of obs: 1579, groups: id:colony, 54; colony, 14  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -2.498347 0.385736 -6.477 9.37e-11 \*\*\*  
## distance -2.044346 0.316968 -6.450 1.12e-10 \*\*\*  
## sexM 0.725066 0.431201 1.682 0.09266 .   
## ordinal\_date 1.646245 0.161120 10.217 < 2e-16 \*\*\*  
## illum -0.003349 0.003224 -1.039 0.29896   
## tailwind 0.243765 0.085270 2.859 0.00425 \*\*   
## sexM:ordinal\_date -0.456073 0.207821 -2.195 0.02820 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) distnc sexM ordnl\_ illum talwnd  
## distance 0.216   
## sexM -0.518 -0.068   
## ordinal\_dat -0.165 -0.174 0.119   
## illum -0.205 -0.022 -0.014 -0.130   
## tailwind -0.029 -0.047 0.007 0.071 -0.023   
## sxM:rdnl\_dt 0.106 0.058 -0.127 -0.724 0.005 -0.010

 ### There is a significant relationship with each variable except for the nighttime illumination % (illum). When observing the effect plot for illum, there is only a slight negative relationship between illumination and the probability of roosting at a wetland. Indeed, 0% illumination estimated to have a mean estimate of ~10.5% of birds roosting, and 100% illumination having a mean estimate of 8% of birds roosting. The confidence interval does contain potential effects which could be biologically significant - anywhere for a decreasing relationship from 18% to 4% roosting rate with illumination, or 6% to 16% increasing rate with illumination. Because of this I cannot definitively say there no relationship between roosting rate/probability and nighttime illumination (as this differences would be biologically significant). However, it is likely that this is not an overall important relationship given the sample size here. ### Now I call my plot showing the relationship between distance to the wetland roost and the probability of roosting.



### Here, we can see that the distance to the wetland site has a significant effect on whether the wetland roost will be used. If the wetland habitat is within 30 km (25.62 km being the closest in my dataset), bank swallows are estimated to roost anywhere from ~45%-85% (mean estimate ~60%) of the time. There is a significant drop in roosting with distance to the wetland. At 50 km, only 15% of bank swallows will travel to the wetland to roost (with 95% confidence interval maximum being ~25%). Beyond this distance, the wetland starts to be used <20% of the time.

### I will continue with my exploration of effects with the interaction of sex and the ordinal date.

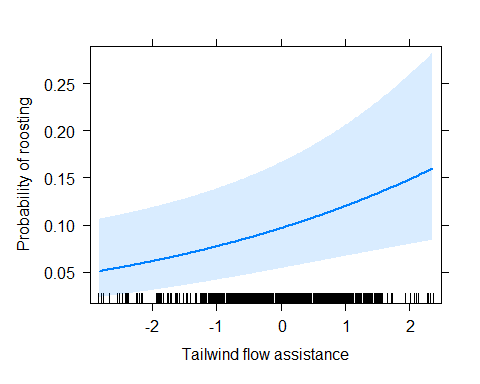


### Here we can see the effect of the interaction sex and ordinal date. During the early breeding season period, males are estimated to use wetlands to roost proportionately more than females. For example, at ordinal date 170, males appear to be using wetlands about twice as much as females (~10% of males roosting versus approximately 5% of females).

### Later it the season it appears that the relationship to ordinal date changes for males and females. Around ordinal date 188, we begin to see females utilizing wetland roosts more than males. By ordinal date 200, females are ~8% more likely to roost than males, although the rate for males is already quite high at ~83%. While the overall interaction between sex and ordinal date is significant, it should be noted that there are no clear points of separation between the 95% CIs of the two sexes.

### Next, I will look at the effect of tailwind.

#get predicted y values for tailwind on the scaled tailwind values  
predict\_tw = predictorEffect("tailwind",m, focal.levels=100,xlevels=1)  
  
#convert tailwind values back to the un-scaled, un-centered form  
predict\_dist$x$tailwind = predict\_tw$x$tailwind\*scaleList\_tw$scale + scaleList\_tw$center  
  
#plot the data (rescale.axis = F prevents it from changing the spacing between the y-tick marks to make the relationship appear linear)  
plot(predict\_tw,rescale.axis=F,main="", ylab="Probability of roosting",xlab="Tailwind flow assistance")

 ### Here it can be seen that the degree of tailwind has a positive effect on the probability of roosting for the bank swallow. At the highest degree of headwind (~-3 flow assistance),  
### Finally, I will look at variation in the random effects.

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: roost ~ distance + sex + ordinal\_date + sex:ordinal\_date + illum +   
## tailwind + (1 | colony/id)  
## Data: df.scale  
##   
## AIC BIC logLik deviance df.resid   
## 1062.4 1110.6 -522.2 1044.4 1570   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -12.0398 -0.3228 -0.1198 0.1719 16.8109   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## id:colony (Intercept) 1.670 1.2921   
## colony (Intercept) 0.615 0.7842   
## Number of obs: 1579, groups: id:colony, 54; colony, 14  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -2.498347 0.385736 -6.477 9.37e-11 \*\*\*  
## distance -2.044346 0.316968 -6.450 1.12e-10 \*\*\*  
## sexM 0.725066 0.431201 1.682 0.09266 .   
## ordinal\_date 1.646245 0.161120 10.217 < 2e-16 \*\*\*  
## illum -0.003349 0.003224 -1.039 0.29896   
## tailwind 0.243765 0.085270 2.859 0.00425 \*\*   
## sexM:ordinal\_date -0.456073 0.207821 -2.195 0.02820 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) distnc sexM ordnl\_ illum talwnd  
## distance 0.216   
## sexM -0.518 -0.068   
## ordinal\_dat -0.165 -0.174 0.119   
## illum -0.205 -0.022 -0.014 -0.130   
## tailwind -0.029 -0.047 0.007 0.071 -0.023   
## sxM:rdnl\_dt 0.106 0.058 -0.127 -0.724 0.005 -0.010

## grpvar term grp condval   
## Length:68 (Intercept):68 123:Allensen : 1 Min. :-2.34709   
## Class :character 223:Simcoe : 1 1st Qu.:-0.59104   
## Mode :character 129:Crosby : 1 Median :-0.08522   
## 212:Waterford: 1 Mean : 0.04814   
## 239:Cultus : 1 3rd Qu.: 0.57996   
## 237:Kingswood: 1 Max. : 2.34602   
## (Other) :62   
## condsd   
## Min. :0.4743   
## 1st Qu.:0.6105   
## Median :0.6594   
## Mean :0.7376   
## 3rd Qu.:0.8515   
## Max. :1.2216   
##

