

# Four principles for improved statistical ecology: a worked example of correlational animal data from a long-term monitoring project

2023-09-30

Setting up the workspace.

```
library(sjPlot)
library(ggplot2)
library(lme4)
library(lmerTest)
library(here)
library(emmeans)

options(digits = 3, scipen = 5)
```

The study used a six-year dataset from a wild nest-box population of Blue Tits to examine the relationship of hatching date with the expression of carotenoid-based colouration in nestlings. Considering decreasing caterpillar abundance in the second half of the breeding season as well as the predictions of the parental quality hypothesis [verhulst2007], the expectation was that the expression of nestlings carotenoid-based colouration should be negatively correlated with hatching date. Moreover, Blue Tit nestlings, contrary to most bird species, including the Great Tit [isaksson2008], are sexually dichromatic in both breast and tail feathers, with greater elaboration in males. This might suggest the signalling role of juvenile male colouration and hence its greater condition dependence. Therefore, the study also predicted a sex-specific pattern, with this association between hatching date and colouration more pronounced in males.

Data for this study were collected during six consecutive breeding seasons, beginning in 2011. During this period, several experiments were conducted in the study population, but to avoid their potential influence on this study we included only data from non-manipulated nests, and (in some nests in 2011-2013) cross-fostered nests in which the brood size was not manipulated. Each year, nest-boxes were regularly inspected from mid-April and the laying date, number of eggs and hatching date were recorded. Considering the potential signalling role of carotenoid-based colouration, we used the avian tetrahedral colour space (TCS) model [stoddard2008] allowing for the incorporation of the visual system of the signal receiver. The model expresses each reflectance measurement as a point in a tetrahedral space, in which the tetrahedral vertices correspond to maximum relative stimulations of four bird retinal cones [stoddard2008]. The variables of interest here are those most related to Blue tit signalling: UV chroma and feather brightness.

The example below is a simplified portion of the full analysis in [janas2019].

## Principle 1. First, define a focused research question, then plan sampling and analysis to answer it

### 1A – Define your research question

#### FINER

- **Feasible** - There are resources and logistics in place to allow for nestlings' sampling within the long-term Gotland project.

- **Interesting** - Colour signalling is attracting ample interest in evolutionary biology. Mechanistic studies that probe possible physiological mechanisms (such as dependence of colour on nutritional/environmental variability) are the most promising in finding the relevant underlying processes.
- **Novel** - The study is the first offering such a substantial sample size and data collected in a wild population.
- **Ethical** - Tiny feather samples used in subsequent measurements can be collected without exerting too much stress on the birds.
- **Relevant** - The biology of the Blue tit is still poorly understood - even though it's a model species. We also have little understanding of the mechanisms generating colour variation in animals.

## Predictions

- Birds hatched later in the season will have access to food less abundant in carotenoids, and therefore will have darker feathers with higher UV chroma.
- Predicted relationship will be more pronounced in males.

## PECO

- **Population** - Wild population of Blue tits on the Swedish island of Gotland.
- **Exposure** - Correlative study - response measured relative to environmental variation linked to date of hatching.
- **Comparison** - Response variables evaluated in a continuous gradient of variation linked to the natural range of hatching dates.
- **Outcome** - Feather colouration (UV chroma, brightness).

## 1B – Match data collection to research aims

The sampling is summarised below.

### Field Sampling

- Nestlings were sampled in all identified broods - all nestlings in each nest were sampled for 3-5 breast feathers 14 days after hatching, which were then mounted on blackened paper for further measurements.
- Natural distribution of hatching dates led to the majority of nests being sampled within medium hatching dates.

## 1C – Plan analysis and consider registration

### Model

- (Generalised) linear mixed model with colour variables as outcome, fixed effects for nestling sex and body weight on day 14 post-hatching to control for these, and random effect for year, nest of origin and nest of rearing (the two are the same in years without cross-fostering).
- Outcome type (normal, log transformed etc.), and relationship type (linear, quadratic) to be determined with reference to residual plots.

### Primary Effects of interest (with confidence intervals)

- Correlation of **hatching date** with the measured colour variables, controlling for nuisance variables (body weight) and random effects.
- Effect of **sex** on the direction of the relationship between colour and **hatching date** (interaction), controlling for nuisance variables and random effects.

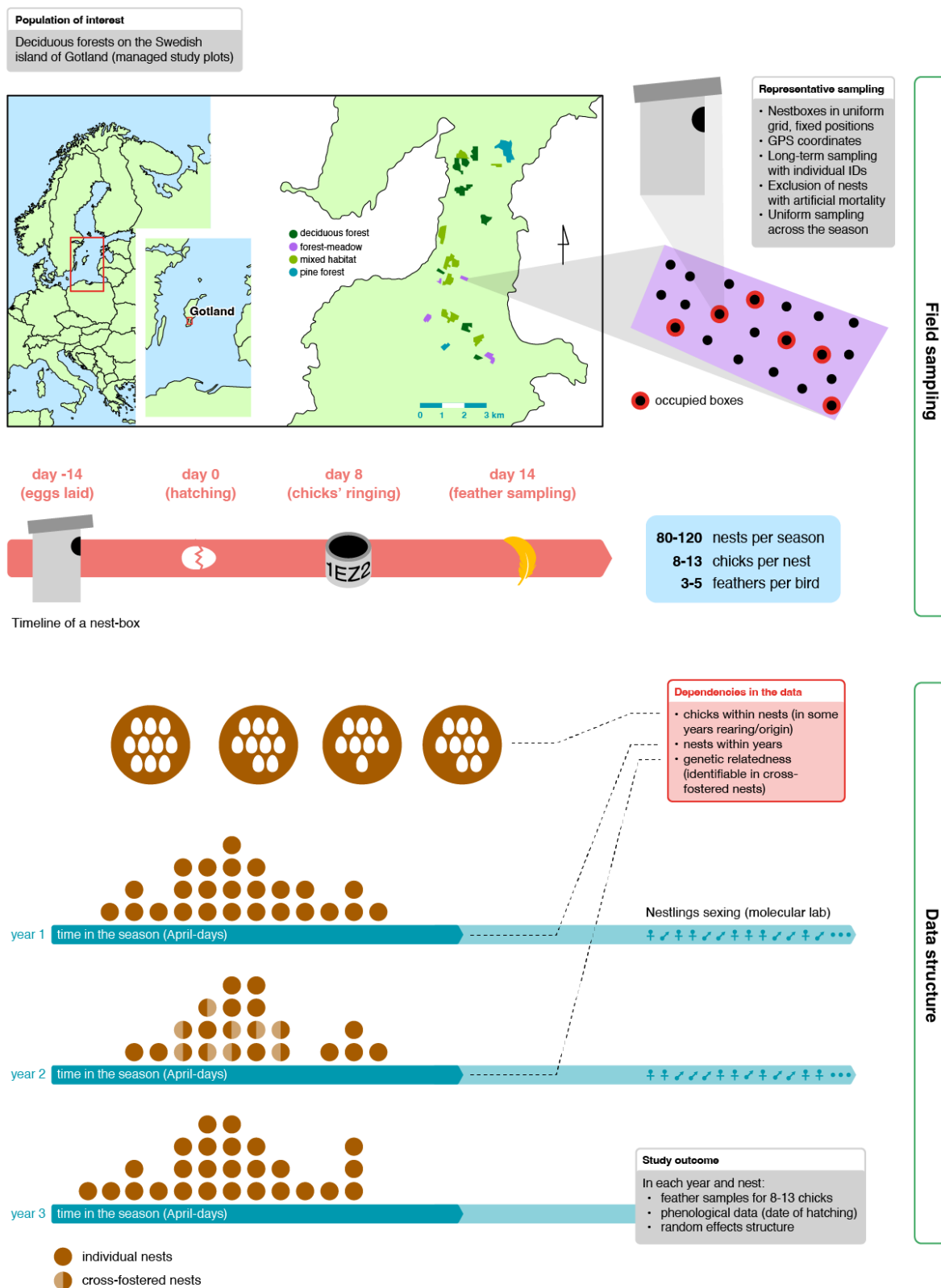


Figure 1: Schematic illustration of the blue tit study from Gotland.

## Multiple testing

No multiple-testing issues requiring additional steps (i.e., not captured by random effects and their hierarchical structure) in case of basic analyses. However, when performing pairwise comparison of predicted effects we adjust p-values using the multivariate t-distribution.

## Principle 2. Develop a model that accounts for the distribution and dependence of your data

### 2A – Model dependence

We have two main sources of dependence

- Nestlings come from different breeding seasons that have more or less unique environmental conditions. We will account for this by including a random effect for year.
- Nestlings come from the same nest, and in some years, from the same nest of rearing (half of brood exchanged in pairs of nests). We will account for this by including a random effect for nest of origin and nest of rearing.

### Read in blue tit data

```
caro_data <- read.csv(here("data","carotenoids.csv"), sep = ";")

# turn sex into a factor
caro_data$SEX <- as.factor(caro_data$SEX)

# for ease of interpretation - scale the responses
caro_data$UV_CHROMA <- scale(caro_data$UV_CHROMA)
caro_data$BRIGHTNESS <- scale(caro_data$BRIGHTNESS)
```

### Mixed models

```
uv_chroma1 <- lmer(
  UV_CHROMA ~ OH * SEX +
  MASS_14 +
  (1 | YEAR) + (1 | BOX) + (1 | BOX_OR),
  data = caro_data
)

uv_chroma1 <- lmer(
  UV_CHROMA ~ OH + SEX +
  MASS_14 +
  (1 | YEAR) + (1 | BOX) + (1 | BOX_OR),
  data = caro_data
)

brightness1 <- lmer(
  BRIGHTNESS ~ OH * SEX +
  MASS_14 +
  (1 | YEAR) + (1 | BOX) + (1 | BOX_OR),
  data = caro_data
)
```

The warning about model singularity is clear after inspecting the output - it arises due to one of the random

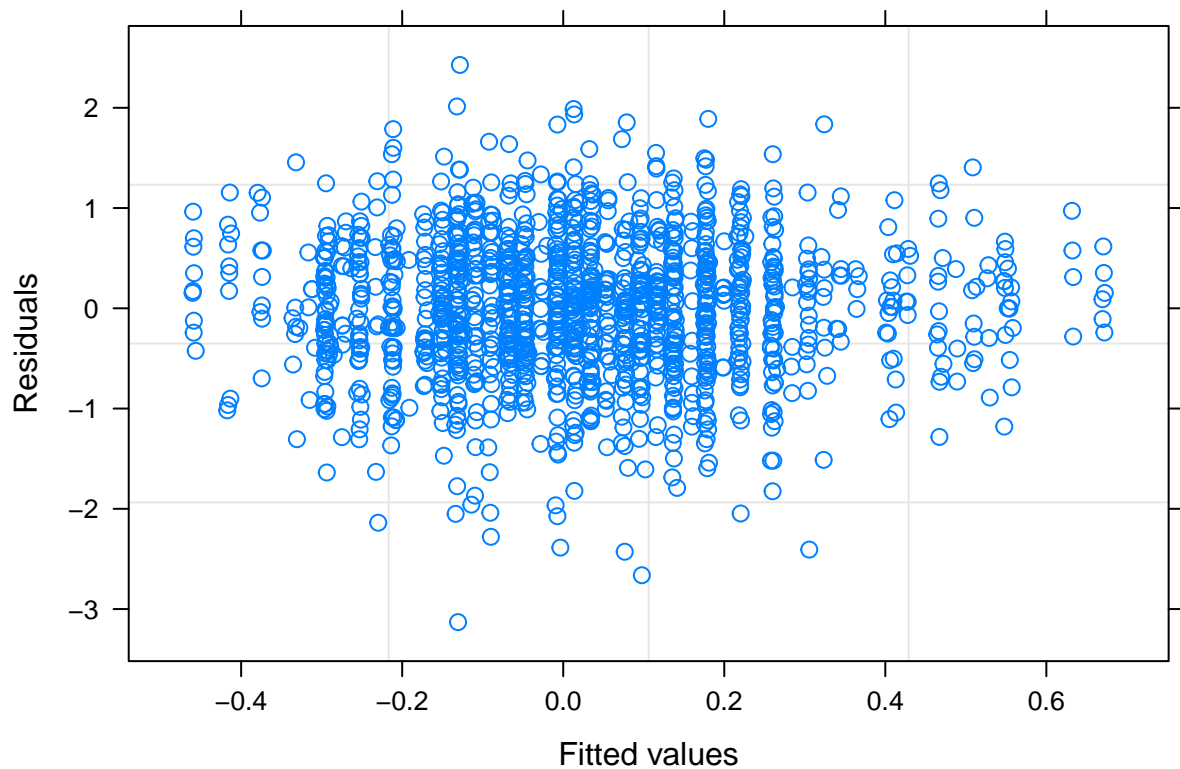
effects contributing little to no variance. We will keep random effects structure constant - but follow-up analysis could involve dropping the random effects yielding zero variance estimates.

## 2B – Check assumptions

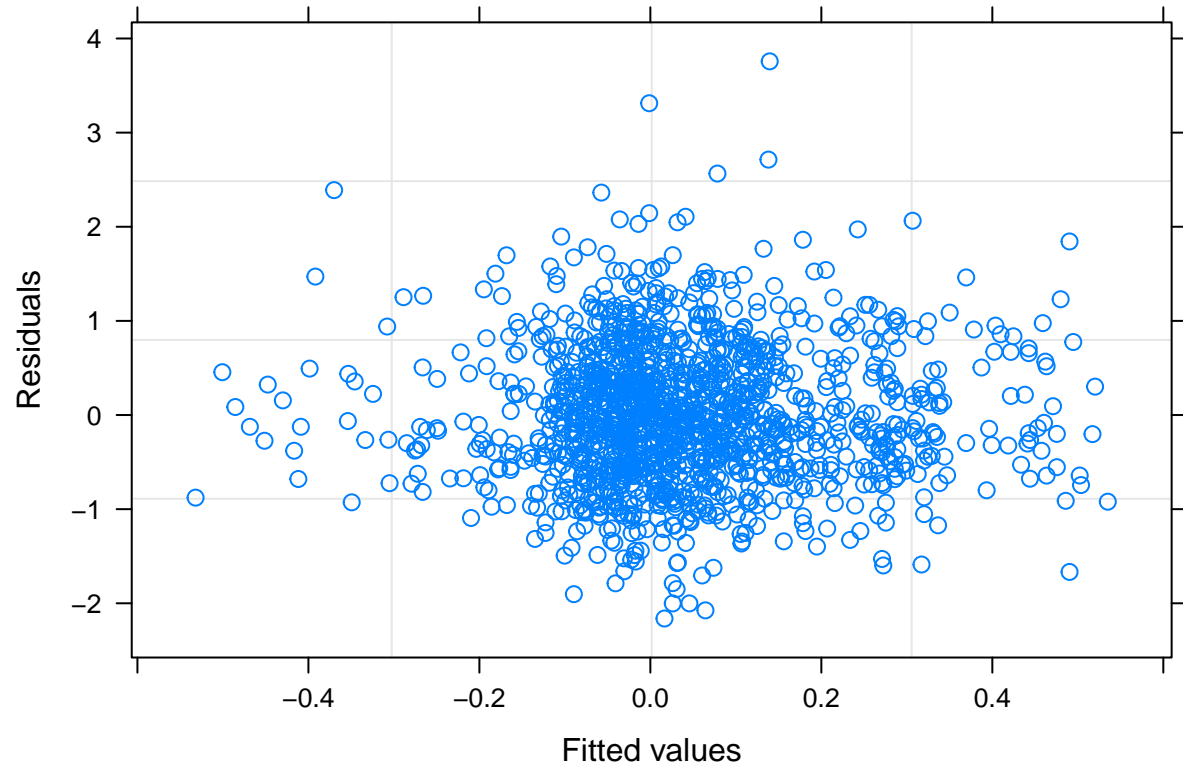
### Residual v.s. fitted plot (marginal)

Note - `re.form = NA` gives marginal residuals.

```
plot(uv_chroma1, resid(.) ~ predict(., re.form = NA),  
     xlab = "Fitted values", ylab = "Residuals")
```

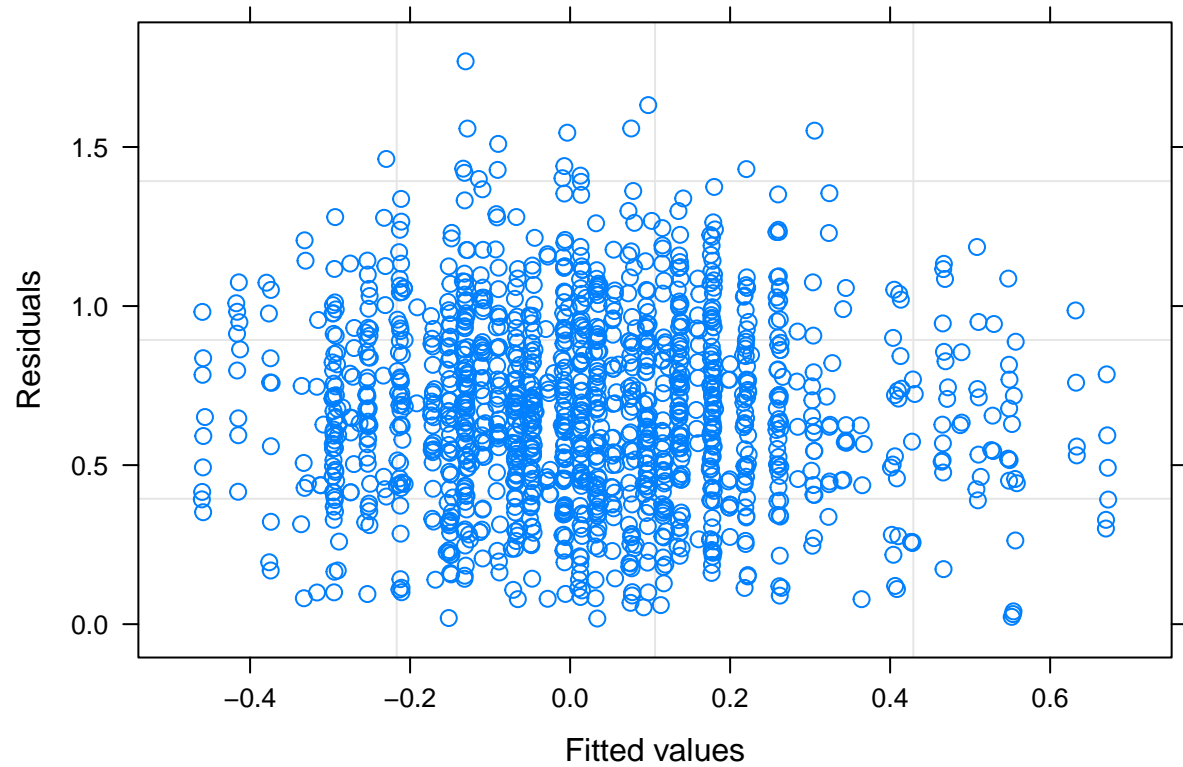


```
plot(brightness1, resid(.) ~ predict(., re.form = NA),  
     xlab = "Fitted values", ylab = "Residuals")
```

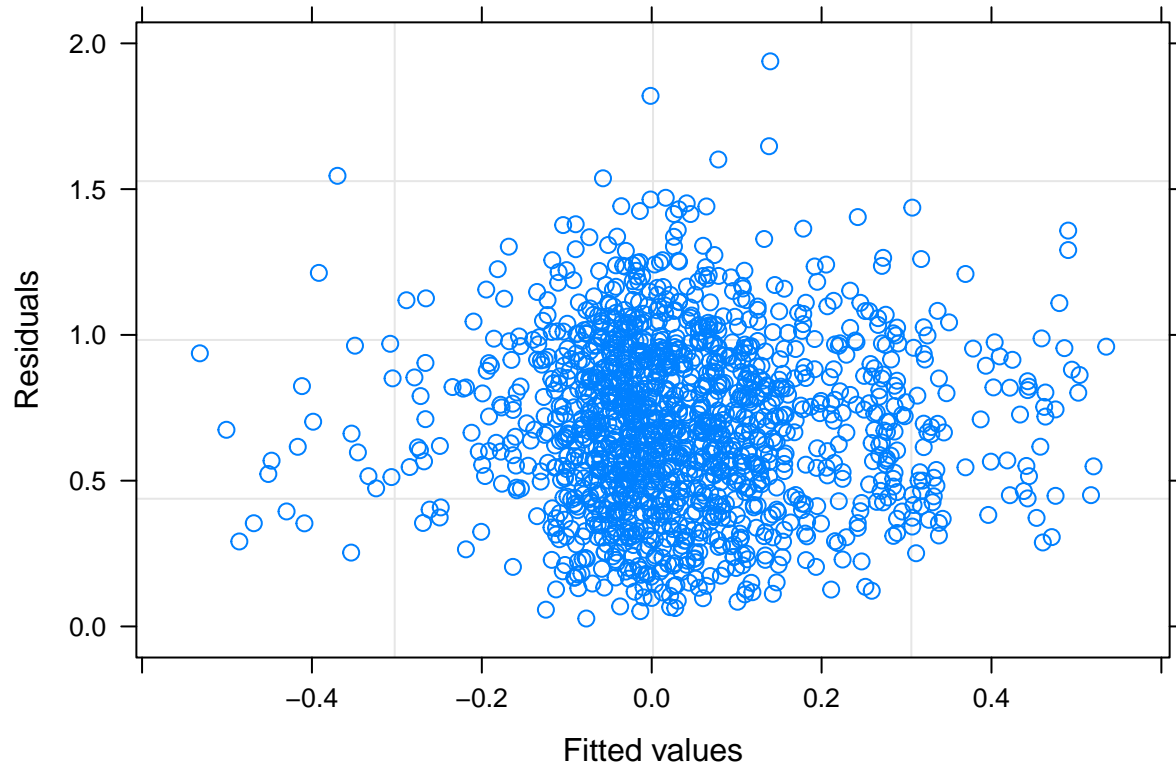


Scale-location plot

```
plot(uv_chroma1, sqrt(abs(resid(.))) ~ predict(., re.form = NA),  
     xlab = "Fitted values", ylab = "Residuals")
```



```
plot(brightness1, sqrt(abs(resid(.))) ~ predict(., re.form = NA),  
     xlab = "Fitted values", ylab = "Residuals")
```



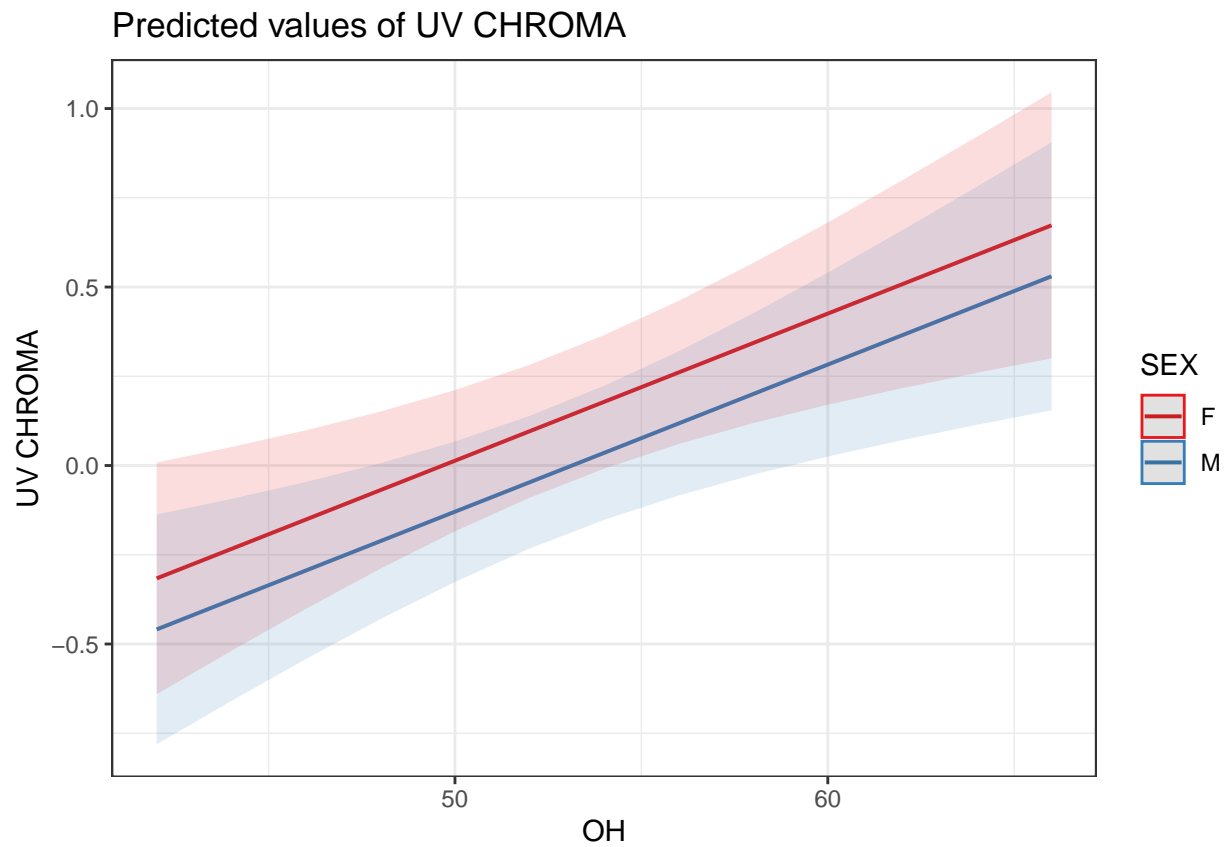
Upon visual inspection we can conclude, that assumptions of linearity and constant variance are satisfied.

### 3A – Replace statistical significance with ecological relevance by emphasising effect sizes

Start by plotting model estimates. We can use the `sjPlot` package to plot both the estimates (with their CIs) and the predicted (sex-specific) correlation with hatching date.

```
pred_plot1 <- plot_model(uv_chroma1,
  type = "pred",
  terms = c("OH", "SEX")
)
# always include confidence intervals
pred_plot1 + theme_bw()
```



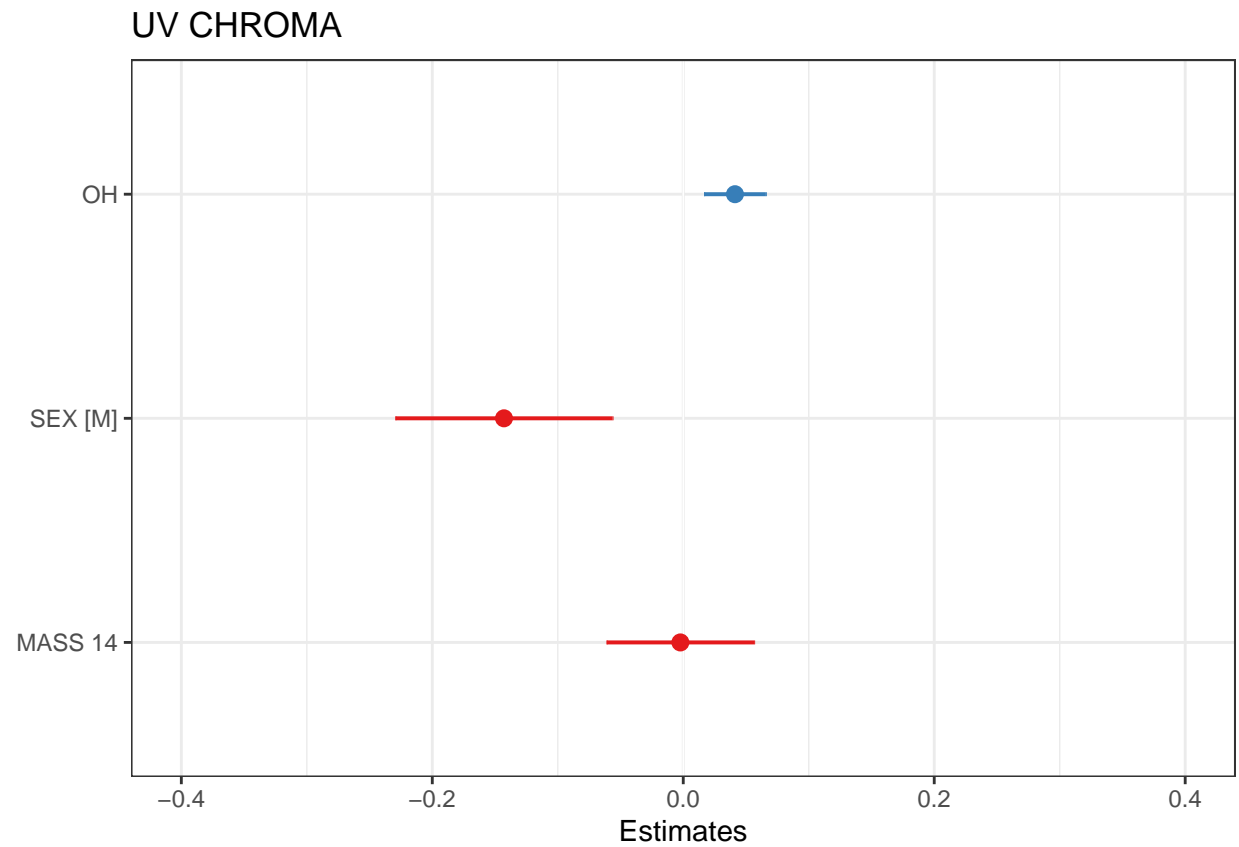


```
est_plot1 <- plot_model(uv_chroma1) + ylim(-0.4, 0.4)
```

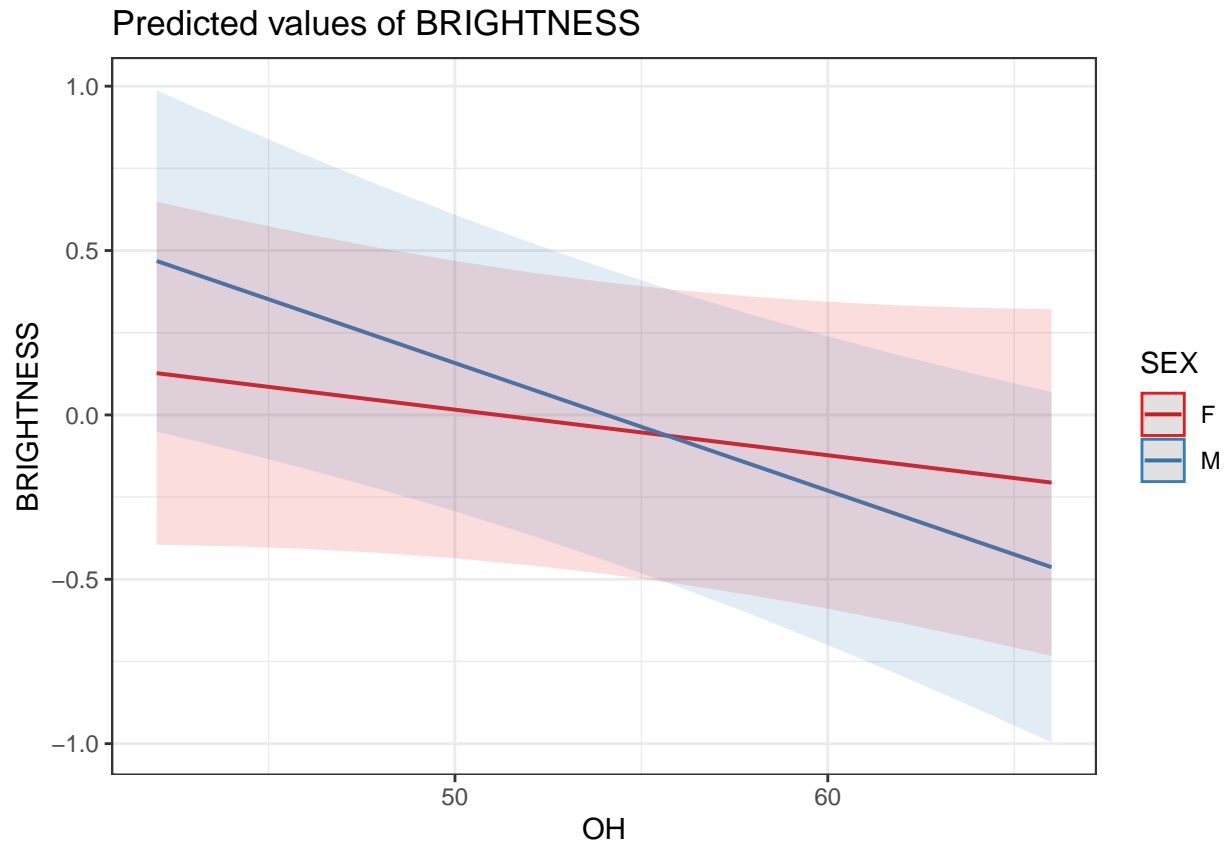
```
## Scale for y is already present.
```

```
## Adding another scale for y, which will replace the existing scale.
```

```
est_plot1 + theme_bw()
```



```
pred_plot2 <- plot_model(brightness1,  
  type = "pred",  
  terms = c("OH", "SEX")  
)  
# always include confidence intervals  
pred_plot2 + theme_bw()
```

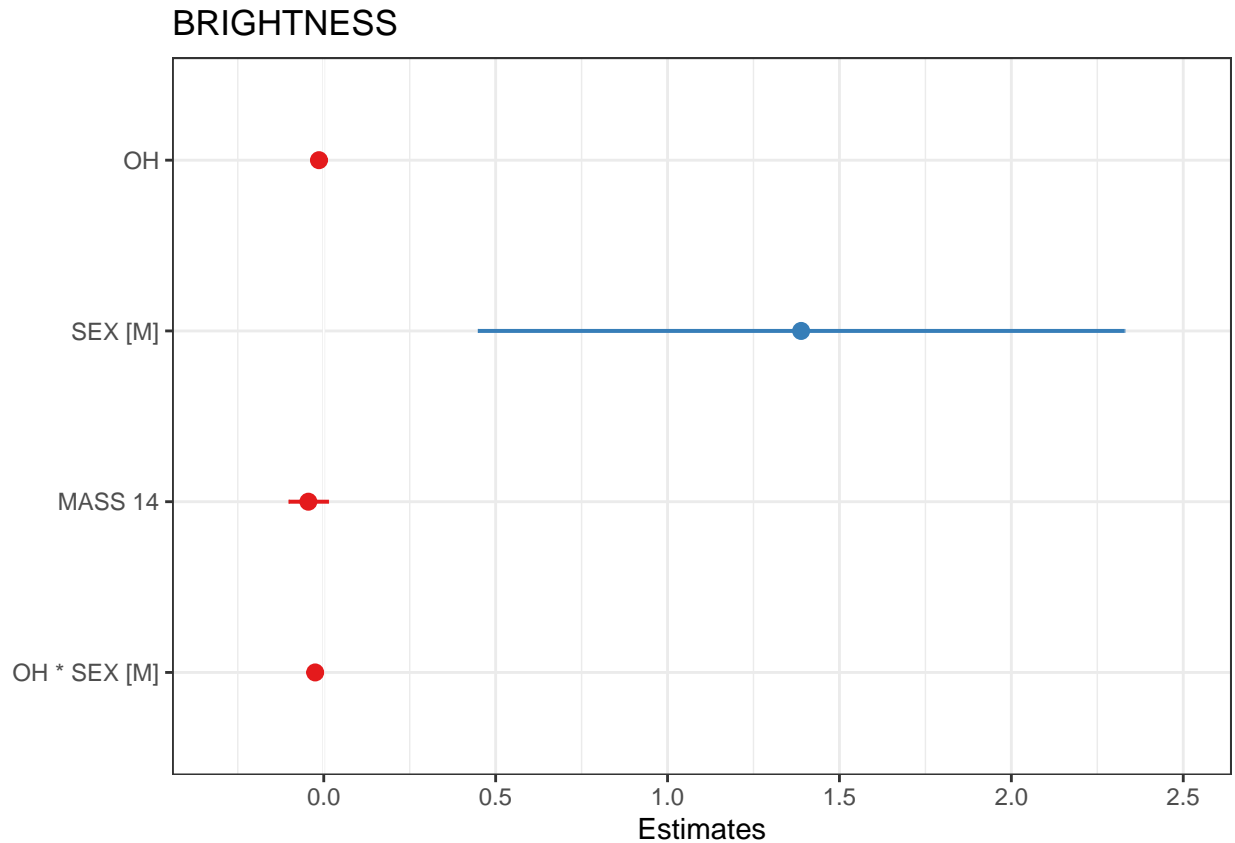


```
est_plot2 <- plot_model(brightness1) + ylim(-0.3, 2.5)
```

```
## Scale for y is already present.
```

```
## Adding another scale for y, which will replace the existing scale.
```

```
est_plot2 + theme_bw()
```



Using the profile method, we calculate 95% confidence intervals for the estimated coefficients.

```
ci_chroma1 <- confint(uv_chroma1)
```

```
## Computing profile confidence intervals ...
```

```
ci_bright1 <- confint(brightness1)
```

```
##Computing profile confidence intervals ...
```

Next we calculate the desired effects and their confidence intervals, using emmeans.

```
em_chroma1 <- emmeans(uv_chroma1, ~ OH,
                      at = list(OH = c(45, 65)))
```

```
em_chroma2 <- emmeans(uv_chroma1, ~ OH + SEX)
```

```
em_bright1 <- emmeans(brightness1, ~ OH + SEX,
                      at = list(OH = c(45, 65)))
```

```
et_bright1 <- emtrends(brightness1, var = "OH", pairwise ~ SEX)
et_bright_ci <- confint(et_bright1$contrasts)
```

Calculate all pairwise comparisons between the predicted effects in case of brightness (where hatching date and sex interact).

```
pairwise_bright <- contrast(em_bright1, method = "pairwise", adjust = "mvt")
pairwise_chroma <- contrast(em_chroma2, method = "pairwise", adjust = "mvt")
```

```
pairwise_bright_ci <- confint(pairwise_bright)
```

## Results

In case of UV chroma, hatching date was positively correlated with UV colouration (slope estimate:  $b = 0.041$ , 95% CI: 0.018 - 0.071,  $p = 0.002$ ). Males and females did not differ in the strength of this relationship (interaction removed from the model). Males had consistently lower UV chroma than females (predicted chroma at average hatching date; males: -0.051, 95% CI: -0.301 - 0.2; females: 0.092, 95% CI: -0.159 - 0.343).

Brightness of feathers decreased with hatching date and the rate of decrease was different in males and females (males:  $b = -0.039$ , 95% CI: -0.063 - -0.015; females:  $b = -0.014$ , 95% CI: -0.038 - 0.01; sex difference 0.025, 95% CI: 0.007 - 0.043,  $p = 0.007$ ; @fig-brightpred). Translating these estimates to predicted brightness levels: female feather brightness is predicted to drop between the beginning and end of the season (marginal predictions for April-days 45 and 65) by 0.277, 95% CI: -0.325 - 0.88 ( $p = 0.6$ ); in males the drop in brightness would be 0.776, 95% CI: 0.168 - 1.384 ( $p = 0.007$ ).

## Discussion

The analysis has demonstrated, according with the expectations, that Blue tit nestlings hatching later had darker feathers, and higher UV chroma. Both of these patterns are indicative of a lower amount of lutain in nestling diet, typical for seasonal decline in the quality and abundance of carotenoid-rich prey brought by the parents. These results suggest that environmental control of carotenoid-based colouration in this species is strong. Moreover, in case of feather brightness - often shown as a key parameter in mate choice [parker2013] - males responded more strongly than females, in line with the expectation that resource-based trade-offs will be stronger in males than in females.