

Hunting Effect on Individual Deer Stress Level

P15.2 Fortgeschrittenes Praxisprojekt

Nikolai German, Thomas Witzani, Ziqi Xu, Zhengchen Yuan, Baisu Zhou

Table of contents

1	Summary	4
2	Introduction	5
2.1	Background	5
2.2	Data Generating Process	5
2.3	Research Question	5
3	Data Analysis	6
3.1	Hunting Events	6
3.2	Movement Data	7
3.3	Faecal Sample Data	7
4	Data Preprocessing	8
4.1	Filtering	8
4.2	Assumptions	9
4.3	Uncertainties	9
5	Model Selection	10
5.1	Statistical modeling approach	10
5.2	Machine learning approach	10
6	Model Evaluation	11
6.1	subtopic 1	11
6.2	subtopic 2	11
7	Conclusion & Outlook	12

List of Figures

1	Location of Bavarian Forest National Park	5
2	FCM levels over time	5
3	Hunts - completeness of timestamp	6
4	Hunts - Daily Count	6
5	Hunts - Locations	6
6	Deers - Locations of four Deer	7
7	Samples - Daily Count	7
8	Samples - Locations	7

1 Summary

2 Introduction

2.1 Background

Apart from the change in populations size, the effects of hunting on wildlife have so far been studied mainly on the basis of behavioral changes. This project aims to investigate the physiological stress response in red deer using a non-invasive method, the measurement of faecal cortisol metabolites (FCMs).

2.2 Data Generating Process

The data for the project originates in the *Bavarian Forest National Park*. Its location is highlighted in green in Figure 1. Within and on the borders of this area, red deer roam freely. Some of these deer have been **collared with a GPS-device**, which helps to track the movement. At some time, a **hunting event** happens and the deer experiences some amount of stress. Later, the deer defecates (**"defecation event"**). Subsequently, researchers visit the defecation location and collect a **faecal sample**.



Figure 1: Location of Bavarian Forest National Park

Stress is expected to be higher in proximity¹ to hunting events. With higher stress, FCM values are expected to be higher. Huber et al. (2003) showed (Figure 2) that the FCM levels peak between 16 and 19 hours after a stress event (called "challenge"). Additionally we expect, that FCM levels are lower, the more time passes between defecation and sampling.

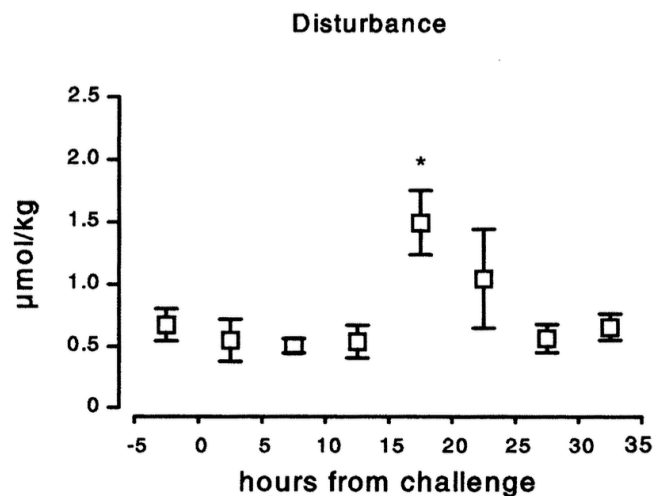


Figure 2: FCM levels over time

2.3 Research Question

Therefore our research question is two-fold:

- assess the effect of temporal and spatial distance on FCM level
- assess if the time between defecation event and sample collection affect the FCM levels

¹this includes temporal and spatial proximity

3 Data Analysis

We were provided with four distinct Datasets. In the following subchapters we are going to describe the main features of each data set, but the reproduction data (see Chapter 7) and address any anomalies.

3.1 Hunting Events

The dataset contains the location and time of roughly 700 individual hunting events, spanning from 2020 to 2022. There are three main challenges:

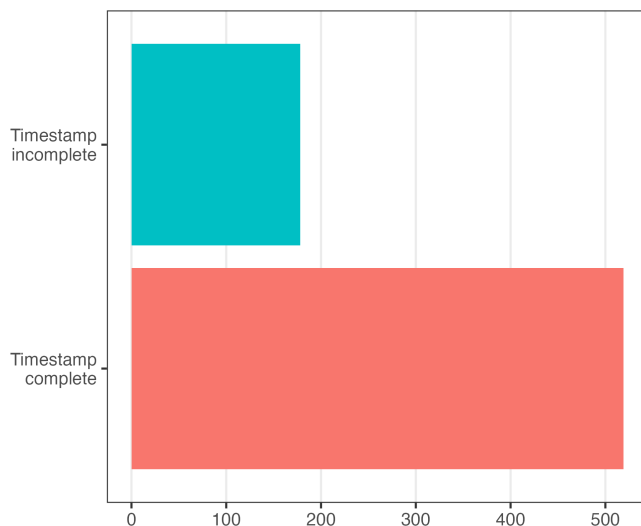


Figure 3: Hunts - completeness of timestamp

- i) Just a little over 500 of these events have a complete timestamp, consisting of date and time of day (Figure 3).
- ii) The events are not represented as a period of time, but as a single moment in time. Additionally, as shown in Figure 4, there appears to be seasonality in the occurrence of hunts.
- iii) Similarly to ii), the events are only associated with a single spatial point. The locations of the events with complete timestamps are illustrated in Figure 5.

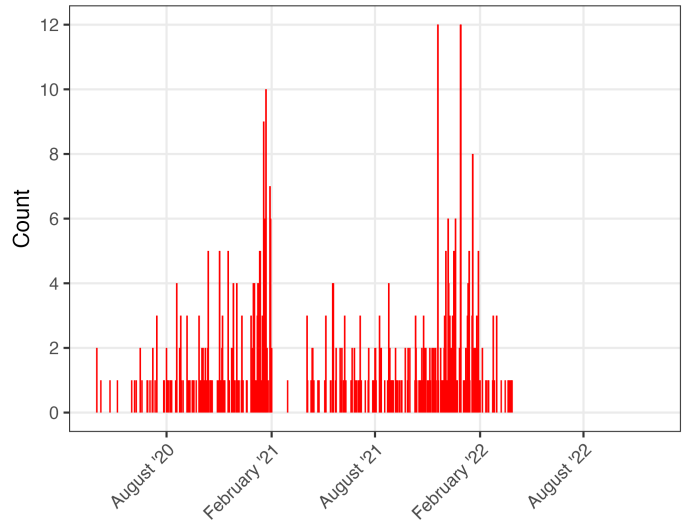


Figure 4: Hunts - Daily Count

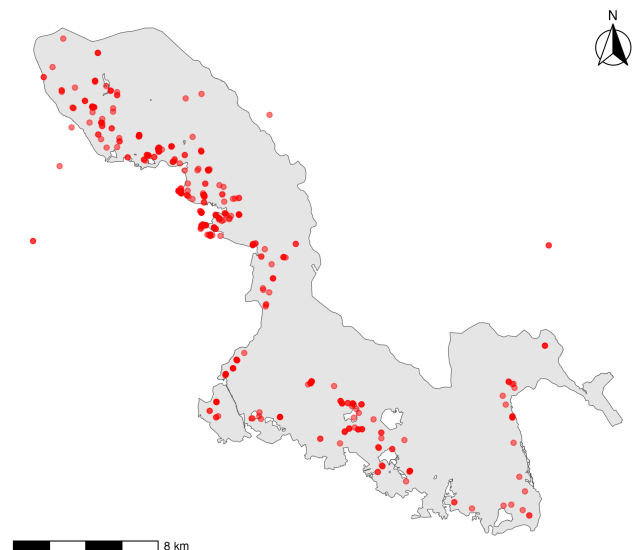


Figure 5: Hunts - Locations

3.2 Movement Data

There are **40 collared deers** which movements have been tracked completely or partially between February 2020 and February 2023. Some collars stopped working before the end-date, some deers got collared *within* the timespan of interest. The location of each individual deer is **tracked on an hourly basis**. The Movement of four randomly selected deer is visualised in Figure 6. During the winter months, the deers roam in one of four enclosures within the national park.

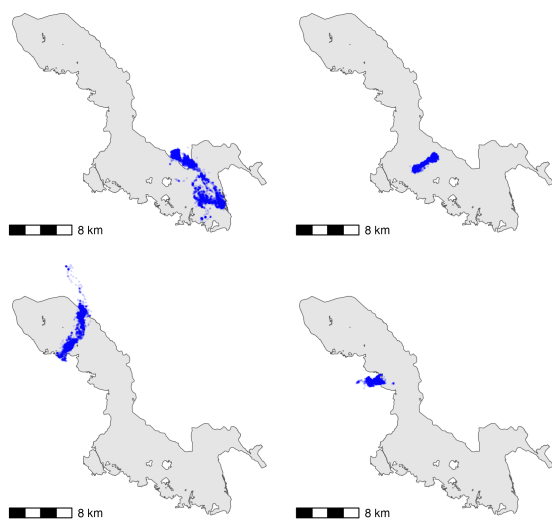


Figure 6: Deers - Locations of four Deer

3.3 Faecal Sample Data

The faecal sample dataset contains information on **809 faecal samples**. Most importantly, the FCM-level (in nanograms per gram [ng/g]), the location of the sample, as shown in Figure 8, the associated collared deer, the approximate time of defecation and the time of sampling. The samples were taken at irregular intervals, but with obvious seasonality (see Figure 7) from 2020 to 2022.

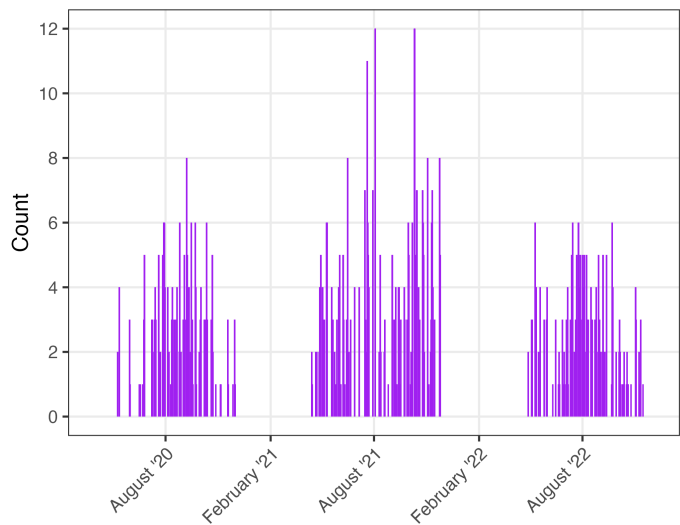


Figure 7: Samples - Daily Count

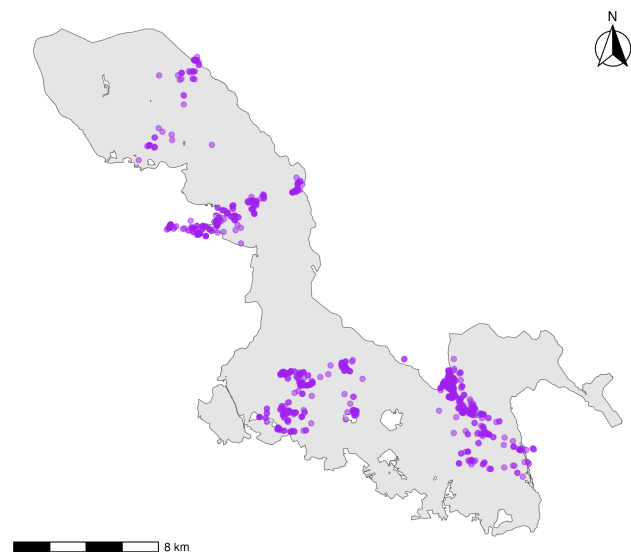


Figure 8: Samples - Locations

4 Data Preprocessing

The original datasets provided by Dr. Ferry are as follows: 1. “FCM Stress - Collared Deer - CRS=ETRS UTM 33N.csv” Contains all measured Faecal Cortisol Metabolite (FCM) values (stress levels) from faeces samples collected from GPS-monitored red deer. 2. “Movement - CRS=ETRS UTM 33N.csv” Includes all GPS locations for each red deer equipped with GPS collars. Most locations were recorded at 1-hour intervals, with some at 15-minute intervals. 3. “Hunt Events - CRS=ETRS UTM 33N.csv” Lists hunting events recorded between 2020 and 2022. A few weeks into the project, Dr. Ferry provided an updated Hunt Events dataset. This updated dataset was corrected by Dr. Ferry and his colleagues after we expressed concerns about data validity. Additionally, Dr. Ferry provided another dataset, “Reproduction Success Results.xlsx,” containing pregnancy status data for certain deer at specific times. The initial idea was to include this information in our analysis, as pregnancy might affect stress levels. However, we did not find a meaningful way to incorporate this data. Reasons included uncertainties about the pregnancy status of deer not listed, sparse data points and unclear entries. The three main datasets underwent several preprocessing steps to facilitate analysis: 1. Timestamps were parsed and standardized to a uniform format. 2. Columns were renamed for clarity; for example, in the FCM dataset, Sample_ID remained as Sample_ID, and X and Y coordinates were renamed to DefecX and DefecY. 3. Pregnancy data was initially merged into the FCM dataset but later omitted, as mentioned previously. 4. “Collar_day” and “Collar_time” were combined into a single “DefecTime” variable. 5. Coordinates were converted from geographic to UTM format. Entries with missing values, zeros, or other highly implausible data were filtered out. 6. The cleaned datasets were saved as .rds files in the ./Data directory. Since our primary goal is to investigate the relationship between deer stress levels and hunting events, we needed relevant variables. A hunting event and a deer are separated by both space and time, especially since stress levels are measured via faecal samples collected after the event. Essen-

tial variables include the spatial distance between a deer and the hunting event at the exact time of the event, as well as the time difference between the hunting event and the subsequent defecation. Given that movement data primarily have 1-hour intervals and hunting events occur at arbitrary times, we needed to approximate the exact deer positions at hunting event times. We first interpolated deer positions at the precise times of all hunting events: 1. Duplicate entries and records with missing timestamps were removed from hunting events. 2. For each hunting event, we identified the closest movement data point before and after the event. 3. Assuming constant velocity and linear movement between these two points, we calculated the exact position of the deer at the hunting event time. This interpolation involved computing the fraction of time elapsed between the first movement point and the hunting event relative to the total interval between the two movement points. This fraction was then multiplied by the movement vector, resulting in an interpolated position. Next, we paired FCM samples with hunting events. Knowing the exact deer associated with each sample allowed a many-to-many pairing. Each FCM sample was associated with all hunting events occurring before the sample was produced. Calculated distances became associated with these sample-event pairs, and the time difference between the defecation event and the hunting event was computed. The resulting dataset contained samples linked to all prior hunting events, creating multiple pairings per sample. This posed a problem: samples collected later had disproportionately many associated hunting events.

4.1 Filtering

To prevent collinearity between observations, we decided to retain at most one assignment per FCM sample. Three versions of the reduced dataset were proposed: 1. Retaining the observation with the time difference between hunting and defecation closest to 19 hours, as suggested by Dr. Ferry and literature (e.g.,

Huber et al.), since cortisol metabolite concentrations peak around 19 hours post-stress event. 2. Retaining the observation with the shortest spatial distance between the hunting event and the deer’s interpolated position. 3. Introducing a scoring system to consider both distance and time difference simultaneously. Recognizing the importance of both factors, we designed a score proportional to the inverse square of the distance and a “spike” function based on time difference. The inverse square component was inspired by physical dynamics (such as the inverse square law for sound intensity), while the spike function was inspired by pharmacokinetics. Multiplying these two components resulted in a combined score, and the observation with the highest score was selected. A minimum score threshold was also established to ensure plausibility regarding the stress level impact.

4.2 Assumptions

We had to make several assumptions about the data. First, we assumed that the movement of each deer between two points is approximately linear and at a constant speed, i.e. we used linear interpolation to determine the coordinates between two signals from the GPS collar. We also assumed that each hunting event was actually a sound event (i.e. a shot was fired). Therefore, we weighted the value of our score function by $\frac{1}{d^2}$, where d is the Euclidean distance between a given deer and a given hunting event.

4.3 Uncertainties

A major challenge in our analysis is the large uncertainty in data which propagates through the pre-processing steps into our models.

First, the uncertainty about spatial distances between deer and hunting events is large. While the movement of deer was tracked at hourly intervals, hunting events occurred irregularly. In our interpolation-based approximation of a deer’s location at the time of hunting event, the approximation error is unknown.

Second, we have very limited information for addressing the short-term stress response of the deer. On the

one hand, the available data contain only 500 hunting events with complete time and location information scattered across a 30km \times 30km area in the span of two years. The observed hunting events only populate the space-time scarcely. On the other hand, we observe in Figure 6 that the deer tended to roam around within a small area. As a result, each deer is likely to have only encountered very few hunting events at small spatial and temporal distances.

Third, there could be a large number of confounders for which we possess no information. Based on GPS coordinates, we can compute Euclidean distances between deer and hunting events. However, we cannot account for terrain and vegetation which could affect the propagation of sound and therefore contaminate the potential effect of hunting events. Furthermore, there could be unobserved stress stimuli, such as weather conditions, predators, and human activities other than hunting.

5 Model Selection

To address how the FCM level depends on time difference and spatial distance, we follow a statistical approach for interpretability and a machine learning approach for exploration. In this section, we present details on model specification.

5.1 Statistical modeling approach

For each dataset resulted from TBD, we consider the following variables as covariates:

- Time Difference between defaction and the most relevant hunting event;
- Distance (in space) between the deer and the most relevant hunting event;
- Sample Delay, i.e., the time difference between sample collection and defecation;
- Defecation Day, i.e., the day of year on which the defecation event occurred;
- Number of Other Relevant Hunting Events within the specified spatiotemporal frame.

The first two covariates are of main interest for our first research question. The effect of sample delay is the subject of our second research question. We consider the day of defecation as a potential confounder, as the deer's stress level exhibits seasonal patterns (Vilela et al. 2020). The number of other relevant hunting events reflects the intensity of recent hunting activity. In time periods of frequent hunting events, a deer might have a generally higher stress level, or adaptation might occur which dampens the stress response towards new hunting events.

We propose a generalized additive mixed model (GAMM) to account for potentially non-linear effects of covariates and potential individual differences between the observed deer. We use the R package `mgcv` (Wood 2011) to fit the models. The covariates Time Difference, Distance, Sample Delay, and Defecation Day are encoded by penalized cubic regression splines. We do not use a cyclical spline for Defecation

Day, because faecal samples were not collected during winter and therefore Defecation Day is not a cyclical variable. The Number of Other Relevant Hunting Events is adopted as a linear effect. We further introduce a random intercept for each deer.

For the distributional assumption, we use the Gamma family with a log link. Since the FCM level is non-negative, we expect the Gamma assumption to more appropriate than Gaussian. A full specification of our GAMM model is as follows. Let $i = 1, \dots, N$ be the indices of the deer and $j = 1, \dots, n_i$ be the indices of faecal samples for each deer

$$\begin{aligned} \text{FCM}_{ij} &\stackrel{\text{iid}}{\sim} \mathcal{Ga}\left(\nu, \frac{\nu}{\mu_{ij}}\right) \quad \text{for } j = 1, \dots, n_i, \\ \mu_{ij} &= \mathbb{E}(\text{FCM}_{ij}) = \exp(\eta_{ij}), \\ \eta_{ij} &= \beta_0 \\ &\quad + \beta_1 \text{Number of Other Relevant Hunting Events}_{ij} \\ &\quad + f_1(\text{Time Difference}_{ij}) + f_2(\text{Distance}_{ij}) \\ &\quad + f_3(\text{Sample Delay}_{ij}) + f_4(\text{Defecation Day}_{ij}) \\ &\quad + \gamma_i, \\ \gamma_i &\stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\gamma^2), \end{aligned}$$

where f_1, f_2, f_3, f_4 are penalized cubic regression splines, $\sigma_\gamma^2 > 0$, and $\nu > 0$.

Our model has two major limitations. First, it does not account for potential interactions between the covariates. Second, the random intercept term can only capture different base stress levels, but not differences in stress response. However, given the limited data, one must trade off between model complexity and stability in estimation. In fact, our current model already exhibits instability with respect to estimation procedures, which we discuss in Chapter 6.

5.2 Machine learning approach

6 Model Evaluation

6.1 subtopic 1

6.2 subtopic 2

7 Conclusion & Outlook

Due to the high level of uncertainty in the data, we were not able to show an effect of spatial and temporal proximity to a hunting event on the FCM levels. However, we were able to show to a certain extent that the FCM level decreases with increasing sample delay.

We deliberately chose not to include the pregnancy data in our model as it only contains information on some deer. As we are not able to label the remainder of the deer as “pregnant”/“not pregnant”, we are convinced that including this information would lead to a huge bias. However, we are strongly in favor of adding additional characteristics that could explain a shift in baseline stress levels.

We believe that a more rigorous recording of hunting events (i.e. timespans instead of single moments, enforcing complete timestamps) and an overall larger amount of data could lead to less uncertainty and thus a clearer result.

Huber, Susanne, Rupert Palme, Wolfgang Zenker, and Erich Möstl. 2003. “Non-Invasive Monitoring of the Adrenocortical Response in Red Deer.” *Journal of Wildlife Management* 67 (April): 258–66. <https://doi.org/10.2307/3802767>.

Vilela, Sofia, António Alves da Silva, Rupert Palme, Kathreen E. Ruckstuhl, José Paulo Sousa, and Joana Alves. 2020. “Physiological Stress Reactions in Red Deer Induced by Hunting Activities.” *Animals: An Open Access Journal from MDPI* 10 (6): 1003. <https://doi.org/10.3390/ani10061003>.

Wood, S. N. 2011. “Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models.” *Journal of the Royal Statistical Society (B)* 73 (1): 3–36.