Hunting Effect on Individual Deer Stress Level

P15.2 Fortgeschrittenes Praxisprojekt

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

Table of contents

Abstract

The experience of hunting events causes stress in red deer, resulting in elevated faecal cortisol metabolites (FCMs). Modelling FCM levels on the spatial and temporal distance of a given deer to hunting events did not show significant effects due to uncertainty within the data. However, we were able to show some relationship between *sampling delay* and FCM levels.

Figure 1: Location of Bavarian Forest National Park

1 Introduction

1.1 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 FCMs.

1.2 Data Generating Process

The data for the project origins 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**.

Stress is expected to be higher in proximity[1] 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.

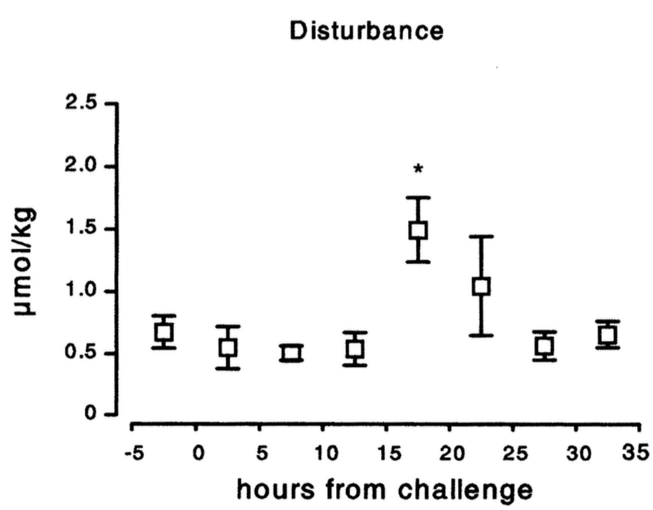


Figure 2: FCM levels over time

[1] **this includes temporal and spatial proximity**

1.3 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

2 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 Section 7.1) and address any anomalies.

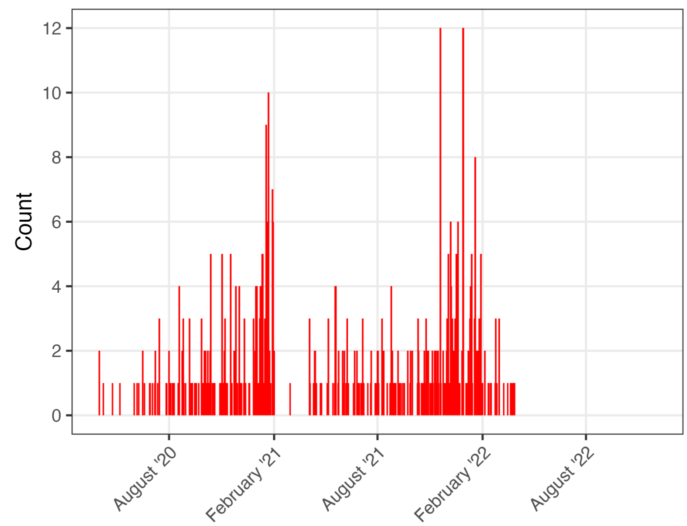


Figure 3: Hunts - Daily Count

2.1 Hunting Events

The dataset contains the location and time of 697 individual hunting events, spanning from 2020 to 2022. There are three main challenges:

i) Just **519 of these 697 events have a complete timestamp**, consisting of date and time of day. The remainder of 178 events only reported the date of the event.
ii) The events are not represented as a period of time, but as a as a **single moment in time**. Additionally, as shown in Figure 3, there appears to be seasonality in the occurrence of hunts.
iii) Similiarly to ii), the events are only associated with a **single spatial point**. The locations of the events with complete timestamps are illustrated in Figure 4.

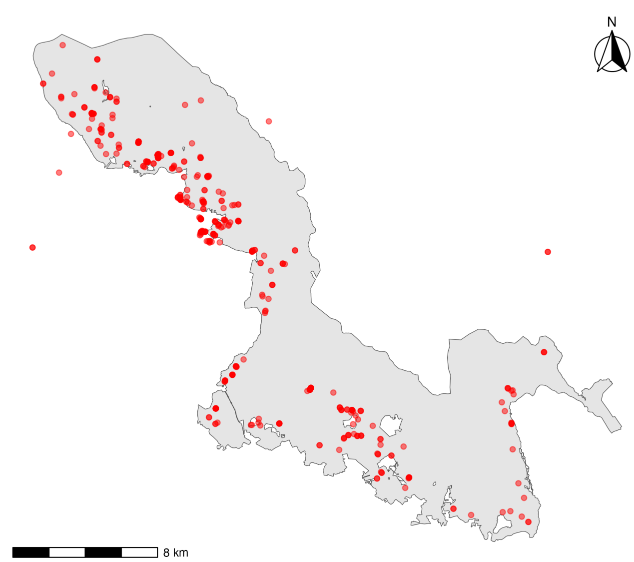


Figure 4: Hunts - Locations

2.2 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 5. During the winter months, the deers roam in one of four enclosures within the national park.

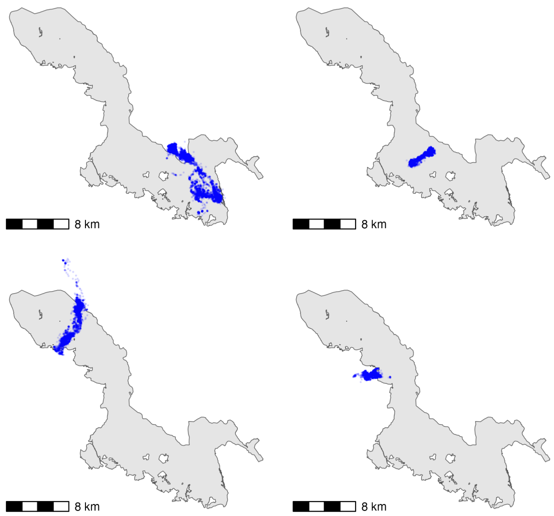


Figure 5: Deers - Locations of four Deer

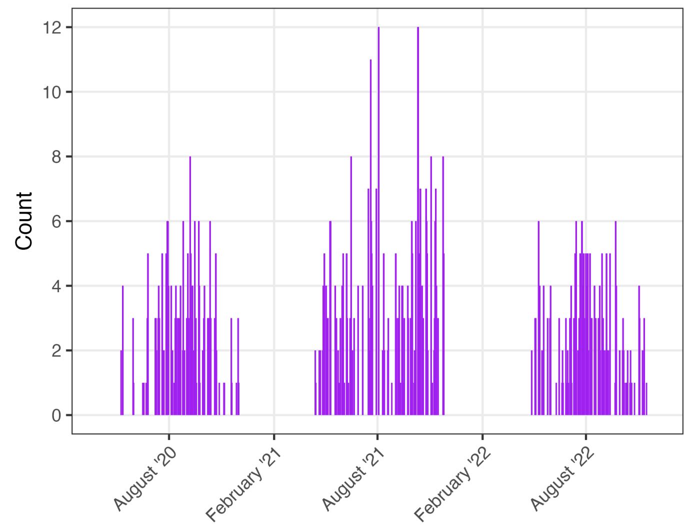


Figure 6: Samples - Daily Count

2.3 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 7, 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 6) from 2020 to 2022.

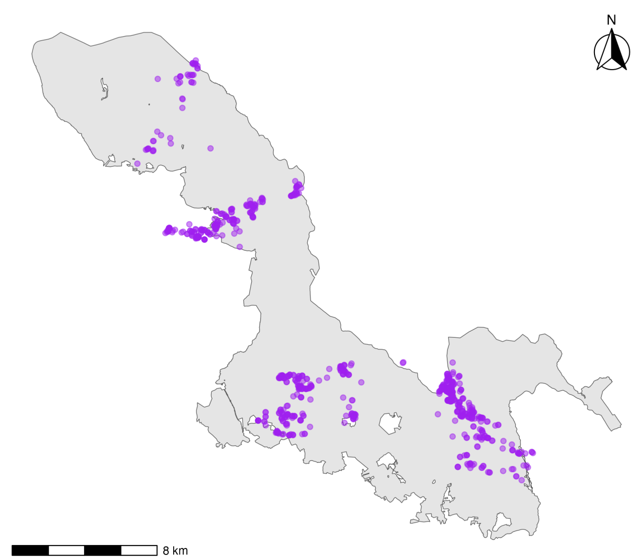


Figure 7: Samples - Locations

3 Data Preprocessing

The datasets undergo preprocessing steps to facilitate analysis and modelling:

- Convertion of Timestamps and Coordinates to a uniform format
- Removal of entries with missing values (especially timestamps), zeros, or highly implausible data

Each FCM sample gets associated with all hunting events occurring before the sample was produced. The resulting dataset containes samples linked to all prior hunting events, creating multiple pairings per sample.

3.1 Filtering

A hunting event and a deer are separated by both space and time. Therefore we define the **essential features** to be:

i) *time difference* between the hunting event and the subsequent defecation
ii) *spatial distance* between a deer and the hunting event at the exact time of the event

Additionally we identify **other features**:

iii) time difference between defecation and sampling (*Sample Delay*)
iv) day of year [1-365] (*Defecation Day*)
v) Number of hunts a deer experiences before defecation (*Number of other Hunts*)

As each sample is associated with one or more hunting event, the main challenge is to identify the most relevant hunting event. We propose three different criterions to identify which hunting event is the most relevant one:

1. **Closest in time**: Choosing the hunting event, which happened closest to 19 hours prior to the defecation event:

$$\underset{t_{Hunt}}{arg\,min} \quad |(t_{Defecation} - 19h) - t_{Hunt}|$$

   This takes the findings of Huber et al. (2003) into account, that cortisol metabolite concentrations peak around 19 hours post-stress event (see Figure 2).

2. **Nearest**: Choosing the observation with the shortest spatial distance between the hunting event and the deer's interpolated position at the time of the hunting event.

3. **Highest score**: Introducing a scoring function to consider both spatial and temporal distance. We designed a score proportional to the inverse square of the distance and a "spike" function based on time difference (see Figure 8). The inverse square component is inspired by physical dynamics[1], spike function was inspired by the general dynamics of stress responses, incorporating insights from Dr. Ferry and further supported by findings on cortisol fluctuations in brown bears following an ACTH challenge, as described in the study 'Blood cortisol and faecal cortisol metabolite concentrations following an ACTH challenge in unanaesthetized brown bears (Ursus arctos) (https://pmc.ncbi.nlm.nih.gov/articles/PMC11734623/)) (https://en.wikipedia.org/wiki/Pharmacokinetics). The observation with the highest score was selected.

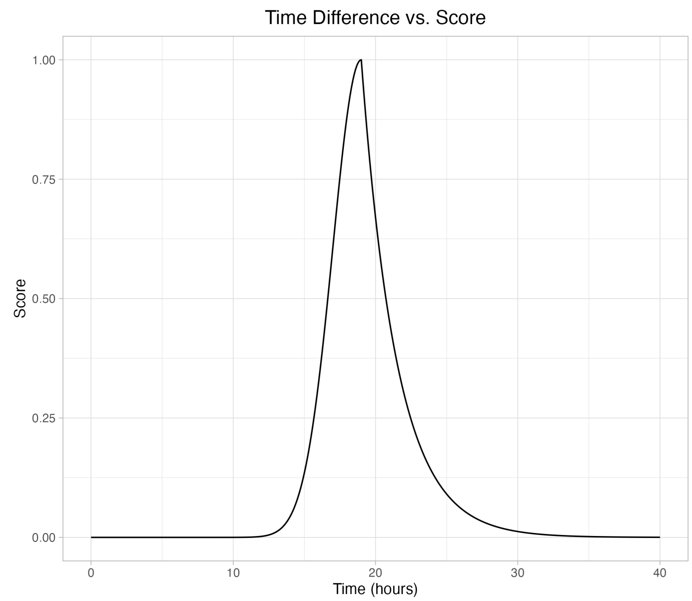


Figure 8: Score Function

$$S(d,t) \propto \frac{1}{d^2} \cdot \begin{cases} f_{\mathbf{t}}(t), \ \mathbf{t} \sim \mathcal{N}(\mu, \sigma^2) & |t \leq \mu \\ f_{\mathbf{t}}(t), \ \mathbf{t} \sim \mathcal{Laplace}(\mu, b) & |t > \mu \end{cases}$$

Where d is the spatial distance, t the time difference between hunting and defecation event, while $\mu = 19$, once again incoporates Huber et al. (2003) findings. $\sigma = 2$ and $b = 2.5$ are shape paramters.

Having identified the most relevant hunting event, feature i) can be calculated easily as the time difference of the hunting event and the defecation event. To obtain feature ii) we have to take in account, that the movement data is recorded in 1-hour intervals, while the

[1] such as the inverse square law for sound intensity (http://hyperphysics.phy-astr.gsu.edu/hbase/Acoustic/invsqs.html)

Table 1: Thresholds

criterion	temporal [h]	spatial [km]
closest in time	36	10
nearest	36	10
highest score	200	15

Table 2: Datasets

DataSet	Proximity Criterion	Deer	Observations
1	closest in time	35	149
2	nearest	35	147
3	score	36	223

hunting events occur at arbitrary times. Therefore we need to approximate the exact deer positions at hunting event times. We interpolate deer positions at the timestamp of each hunting event, assuming constant velocity and linear movement. Features iii) and iv) are again easy to obtain. For feature v), we need to define a cutoff point in time and space, to identify which other hunting events could have affected a deer as well (see Table 1).

3.2 Uncertainties

A major challenge in our analysis is the large uncertainty in data which propogates through the preprocessing 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 5 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.

Fourth, we 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 the inverse square distance.

4 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. Both approaches use all the datasets shown in Table 2.

4.1 Statistical modeling approach

For each dataset shown in Table 2, we consider the five features introduced in Section 3.1 as covariates.

The *time difference* and *spatial distance* 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 $i = 1, \ldots, N$ be the indices of the deer and $j = 1, \ldots, n_i$ be the indices of faecal samples for each deer

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

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 5.

4.2 Machine learning approach

Chen and Guestrin (2016) introduced XGBoost as a machine learning algorithm that improves prediction accuracy by building an ensemble of decision trees sequentially. Each new tree corrects the errors made by previous trees. XGBoost minimizes a loss function using gradient boosting (hence the name) by iteratively adding trees that best reduce the residual errors. Simultaneously, complexity gets penalized through regularization to prevent overfitting. A major feature of XGBoost is that it doesn't require specifying a parametric relationship between target and explanatory variables.

We propose a xgboost model using the R package `xgboost`, choosing *spatial distance* and *time difference* as covariates. To find the best fitting model for each dataset (see Table 2), we had to tune hyperparamters of the proposed xgboost model. We choose to tune the following paramters, using randomized grid search[1]:

- *maximum depth*: Controls the maximum depth of each decision tree, determining how complex each tree can become.
- η: The learning rate, controls how much each new tree contributes to the ensemble, affecting how quickly or cautiously the model learns.
- γ: The minimum loss reduction required to make a split, acting as a regularization parameter that prevents creating splits that don't significantly improve the model.
- *subsample*: Specifies the fraction of training data used to build each tree, introducing randomness to prevent overfitting.
- *features by tree*: Determines the fraction of features (columns) randomly sampled for each tree, increasing diversity among trees.
- *minimum weight of child*: Minimum sum of instance weight needed in a child node, helping control the complexity of the model by preventing overly specific partitions.

5 GAMM Model Evaluation and Results Summary

5.1 Overview

In this analysis, we fitted and evaluated GAMM models across three different matching strategies using two smoothing parameter estimation methods: REML and

[1] see: Bergstra and Bengio (2012)

GCV. The evaluation focused on assessing the interpretability of smooth effects and the validity of random effects.

5.2 Main Findings

5.2.1 Smooth Effect Analysis

- Using REML, the smooth effect curves remained relatively stable, showing a consistent pattern.
- Using GCV, the smooth effect curves exhibited strong fluctuations.
- Across all datasets, there was high uncertainty (large standard errors) in estimated effects, particularly for time difference and distance.
- A consistent sample delay effect was observed with REML: larger sample delay led to lower FCM levels, as expected.
- Instability across estimation methods: GCV tended to produce more wiggly smooth effects compared to REML.
- The estimation of random intercepts was sensitive to the choice of dataset.

5.3 Key Notes and Reflections

5.3.1 Model Performance Concerns

The overall performance of the GAMM model, both in terms of fitting and evaluation, was unsatisfactory. Key concerns include:

- The fluctuation of smooth effects and instability in parameter estimation indicate substantial risks in model reliability.
- These issues may stem from factors such as limited data availability, instability in data preprocessing strategies, or other underlying methodological constraints.

5.4 XGBoost

After comparing the models with those trained on randomized labels, we concluded that there is some relationship between the covariates and the response, as our models slightly outperformed those trained on randomized labels. However, this relationship was very weak and not obvious or intuitive. After reviewing the hyperparameters we concluded that this was likely not noise. Although the predictive power of the XGBoost models was better than that of all other models, it was still not particularly good.

6 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 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.

Acknowledgements

We would like to express our deepest gratitude to *Dr. Nicolas Ferry* for providing us with the opportunity to work on this project.

Daniel Schlichting's mentorship throughout the project was invaluable. We thank him for his guidance and patience.

References

Bergstra, James, and Yoshua Bengio. 2012. "Random Search for Hyper-Parameter Optimization." *J. Mach. Learn. Res.* 13 (null): 281–305.

Chen, Tianqi, and Carlos Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–94. KDD '16. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785.

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.

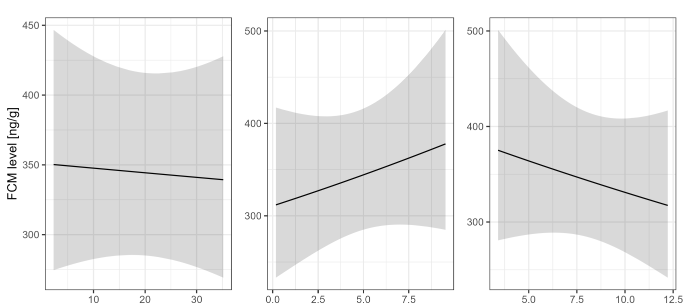
7 Appendix

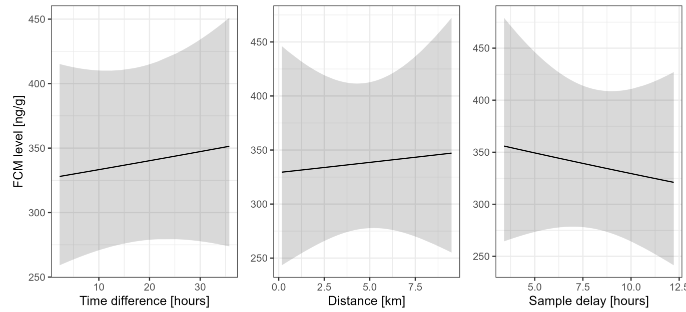
7.1 Feature selection

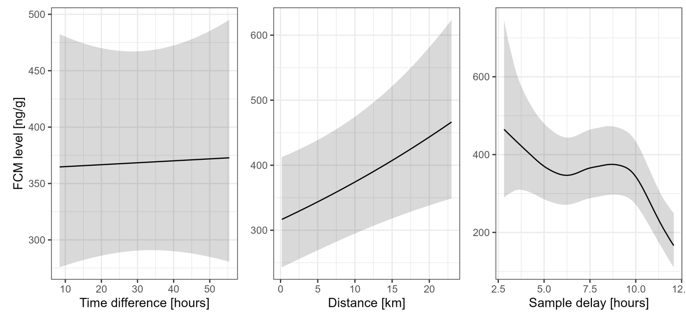
As the reproductive data only contain information on a subset of the deer, we were faced with the decision of how to label the rest of the deer (“pregnant”/“not pregnant”/NA). We are convinced that imputing this information would introduce a huge bias. For this reason, we deliberately chose not to include pregnancy data in our model. However, we are strongly in favour of including additional characteristics that could explain a shift in baseline stress levels.

7.2 REML vs. GCV estimation

7.2.1 REML Smooth Effects Analysis

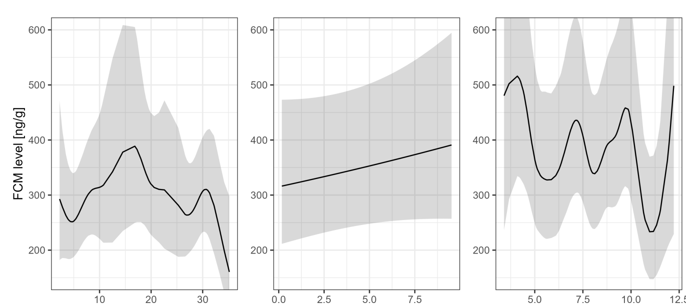


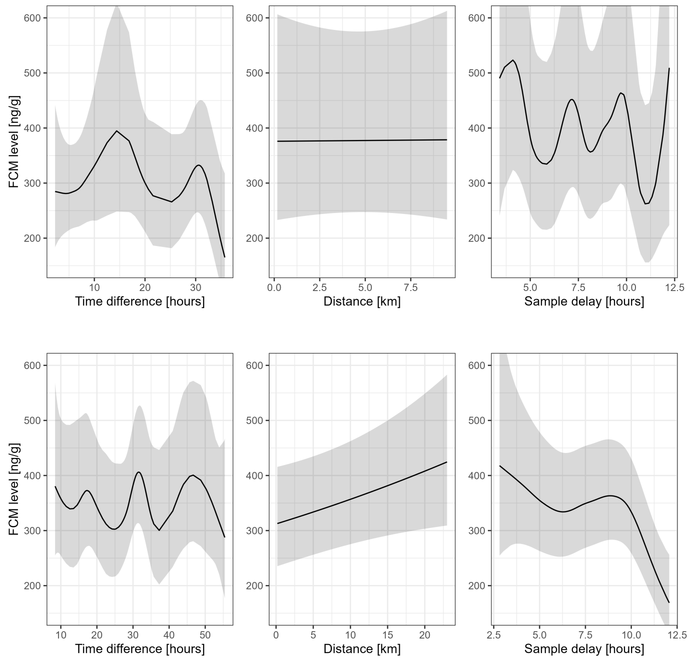




Adjusted predictions for the FCM level. Each plot shows the predicted values of the FCM level in dependence of one covariate. The other covariants are held constant at the mean. The grey area shows the approximate confidence band (mean $\pm$ 1.96 standard error).

7.2.2 GCV Smooth Effects Analysis





GCV (Generalized Cross-Validation) tends to produce more fluctuating smooth effects due to its weaker control over model complexity. Compared to REML (Restricted Maximum Likelihood), it is more likely to select overly complex models, leading to more wiggly estimated smooth curves.

7.3 Additional diagnostic plots

7.4 Deviance explained by random intercept