Chapter 2
Body Weight Prediction of Turkeys: From Walk to Mass

Alexandery Mey[1], Behrouz Raftari[1], Elizabeth Zúñiga[2], Javier Fernández[3], Jos Hageman[4], Martin Stefanov[4], Oliver Sheridan-Methven[5], Yang Zhou[6]

Abstract:
Force plates are useful for examining kinetic movements of turkeys. They provide information about the external forces when a turkey stands or walks on them. Such collected time series data can be used to calculate the weight force and body mass of a turkey. Obtaining data of a high quality and minimum error requires understanding a force plate, good control of data collection, as well as accurate process of transferring data and analysis.
The aim of this article is to investigate how some mathematical and statistical models, combined with machine learning procedures, can help to predict body weight of turkeys based on electrical signals from a force plate. The authors also provide a discussion about techniques for arranging force plates and a scientific method of data collection to assist with a few practical examples of how force plates and mathematical models can be used to reduce manual workload. This work was done jointly with Hendrix Genetics, who in particular provided the data.

Key words: *Time series; Signal filtering; Force plate; Machine learning; Bayesian machine learning, Bayesian hypothesis testing.*

[1]Delft University of Technology, The Netherlands
[2]Université d'Évry Val D'Essonne, France
[3]University of Mondragon, Spain
[4]Wageningen University, The Netherlands
[5]University of Oxford, UK
[6]University of Bath, UK

2.1 Introduction

The problem of predicting the body weight of turkeys by using a force plate was raised by Hendrix Genetics. This introduction section will briefly explain the business interest of the company and why solving the problem could be valuable for the sector and the commercial values of Hendrix Genetics. The main focus of our report is to demonstrate the potential use of statistical models and machine learning techniques to be applied to the problem. However, we aim to present the report in a way that it is also accessible for non-experts in the field.

2.1.1 Background

Hendrix Genetics is a Dutch multispecies animal breeding, genetics and technology company and one of the world's leading breeders and distributors of turkeys and laying hens (fig. 2.1).

Figure 2.1: Hendrix Genetics has breeding programs in turkeys, layers, traditional poultry, swine, salmon, trout and shrimp. They look for innovative, sustainable solutions, together with the entire animal protein chain. At the start of the chain, through better breeding, they have a big influence on the outcome at the consumer level.

The core business consists of the following four steps:

- data collection is carried out for one generation of one breeding species;

- genetic evaluation methods are used for understanding the physical features of that generation;
- animals' selection strategies are outlined based on the genetic evaluation of the generation;
- strategies are considered for breeding the future offspring of the considered generation. Selected individuals will become the new generation to be studied and evaluated through the whole process again.

An important feature of turkeys is their body weight as it can be used to monitor their development. The current practice for weighing these animals is by picking them up and standing with them on a scale. As these animals can weigh up to 25 kilos, this is a slow and labour intensive task which can potentially be stressful for the animals. As a consequence, the animals are not weighed as often as people would want to.

One way of weighing the animals more often and at the same time improving animal welfare is the use of an automated setup for weighing the turkeys. One way of automating weight readings is to measure the turkeys when they walk across a force plate. This way, measurements can be taken each time a turkey crosses the force plate while the animals remain unhindered. A force plate gives of a set of varying electrical signals as long as the animal walks on the force plate. The idea is that the signals are linked in some way to the weight of the turkey. In this study, we will investigate if and how we can use the electrical signals from the force plate to predict an animal's body weight. For this purpose we have a set of roughly 200 animals that have walked across a force plate and also have their body weight manually determined. We will use different mathematical, statistical and machine learning approaches to this end.

2.1.2 Outline

In our report, we will examine the problem in four main sections:

- In Section 2.2 we approach the problem by explaining how the problem at hand can be linked and translated into mathematical language. We discuss in particular how physical values of forces can be calculated from the output signals of force plates.
- In Section 2.3 we present the mathematical and machine learning models we used to approach the problem. We introduce our four models: Bayesian hypothesis testing, sparse Bayesian generalized linear model, Linear Regression and Random Forests.

- The Results section presents and compares the results from all four models. Each model has its advantages and disadvantages in dealing with different types of input. At the same time, the outputs of the models presents insights of our problem of determining the weight of turkeys from different perspectives.
- In the last section, limitations and applications of the models, practical advice on improving data collection and modifying models with new data, are discussed.

2.2 Theoretical Background

2.2.1 Problem Description

We want to develop a model that can accurately predict turkey body weights using a force plate, removing the need to manually weigh the birds on a scale. The challenge of this project is to develop an innovative model to predict body weight and maybe other features from the plate measurements. A model is considered accurate enough when the mean average error is 50 grams.

A force plate is simply a metal plate with one or more sensors to produce an electrical output proportional to the force on the plate. One of the applications of the force plate is to measure the reaction force of the ground on each foot while walking. Since here we are studying an animal, not a human being, and the fact that we can not tell an animal to walk properly, it is better to replace ”walking” with ”motion”. So apart from walking, the movement can include standing, running, jumping and whirling. Accordingly, the time-series records that we have for the turkeys probably are related to a combination of all types of movements. As a result, although this information shows some interesting details of the motion process, it cannot be fully understood and analyzed. Now, in order to estimate the body mass through the dynamic force measurements, it will definitely be useful to study each type of movement separately.

Let's start with the easiest one which is ”standing”. If a turkey stands on the force plate, a force roughly equal to Mg is recorded. Here, because we are interested in body weight, when we say ”force”, it means the vertical component of force (F_z). An accurate Mg can be recorded, when the turkey is completely stable and maybe with the center of mass on the vertical line passing through the turkey legs. Any imbalance and consequently a change in the center of mass eliminates this accuracy and causes the force to be greater/less than Mg. During walking, F_z initially rises from zero to the maximum Mg, then goes below Mg and repeats a similar process and eventually arrives at zero again at the end. This is definitely related to the 3D motion of the center of the mass. Considering a 2D picture in the vertical axis and the direction of movement, the center of the mass follows a swinging curve path. The same analysis is expected to be at least qualitatively valid for "running", since running can be seen as walking with high speed. Let us stop here and assume that motion only involves standing, walking and running, so it can be argued that the maximum F_z/g can be an acceptable, though rough, estimate for mass. This conclusion is supported by our result shown in Fig. 2.2, which compares actual weights with the max and mean values of F_z/g. However, this argument is challenged if we add jumping to move. This is due to fact that by jumping a force much greater than Mg will be recorded. The force generated by jumping is pretty much the same as that of a bouncing ball. The last type of motion considered here, whirling, is perhaps the most complex one. With that in mind, the only thing we can easily understand is that the force will be lower than that of jumping. Further details on the force plate, along with a comprehensive analysis of various types of motion, can be found in the well-cited study of [2]. The complications mentioned are part of the reason we prefer machine learning techniques to a model based on physics.

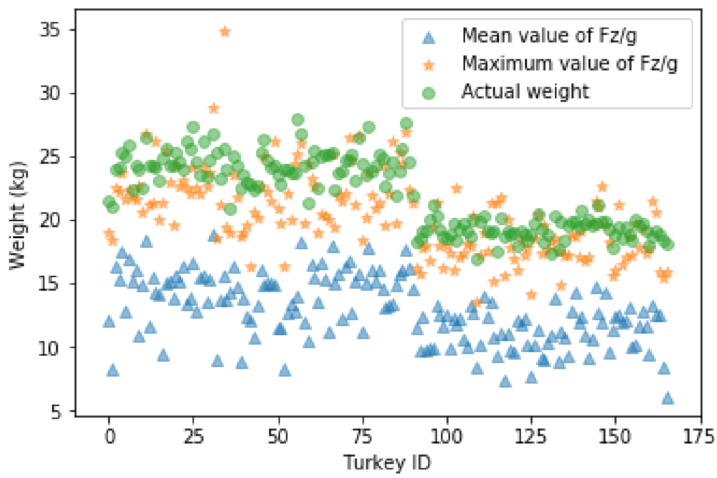


Figure 2.2: Actual weights of the turkeys, max and mean values of F_z/g. Individual turkeys are indexed on the x-axis.

Kistler force plates are useful tools in many areas as clinical research, performance diagnostics and motion analysis. Hendrix Genectics used this force plate to perform a gait analysis in 200 turkeys. The force plate output voltages (eight channels) are used to compute ground reaction forces, moments, center of pressure and coefficients of friction as seen in Appendix 2.A.

We are interested to find a model that can accurately predict turkey body weights using the force plate outputs. The turkeys walks over the force plate approximately for 15 seconds while voltages are measured per centisecond. This information results in time series of 1500 measurements for each turkey in each channel. This data together with the computed forces, the blood line (identified as 1 or 2) and the actual weight of the turkey constitutes the database we work with.

2.2.2 Theories

The authors of [4] measure the body weight with smart insoles. Their methodology is based on Newton's second law, which states that the net force is equated to the product of the mass times the acceleration. When integrating all forces in the vertical direction (plane-z) over time, it leads to the next equation:

$$\int_{HSL}^{HSR} (F_{ZL} + F_{ZR} + W)dt = m \int_{HSL}^{HSR} a_z dt, \tag{2.1}$$

where F_{ZL} and F_{ZR} are the vertical forces from the left and right foot respectively. Furthermore a_z is the gravitational acceleration and W is the body weight. In their experiment, participants start with their left foot and finish with the right one, therefore the measurement starts from the first left heel strike HSL until the last right heel strike HSR. They stated that mean vertical velocity is zero in level walking, thus $\int_{HSL}^{HSR} a_z dt = 0$. Assuming the body weight does not change over the short period of the test, it leads to the following equation:

$$-W = \frac{\int_{HSL}^{HSR} (F_{ZL} + F_{ZR})dt}{t_{HSR} - t_{HSL}} \tag{2.2}$$

The integral (2.2) will be used as a variable in some of our models, computed taking the total force F_z exerted by the two foot. Figure 2.15 from [1] shows the typical form of the curve of the vertical ground reaction force F_z over time when a person is in the stance phase (foot in contact with the ground) of the gait cycle. The curve has two peaks that surpass the body weight and in the middle of these two peaks, a minimum is attained. The curve starts once the foot contacts the ground and finishes once the heel lifts away from the ground.

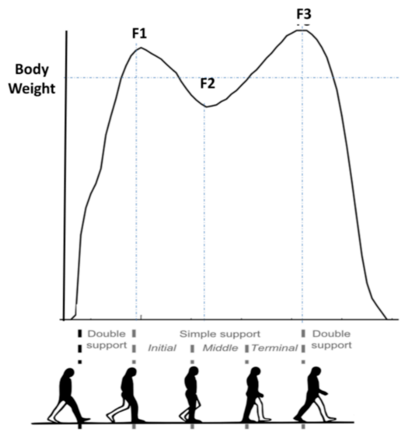


Figure 2.3: Vertical ground reaction force F_z.

2.3 Methodology

2.3.1 Bayesian Hypothesis Testing

One approach which suggests itself for tackling this problem would be Bayesian hypothesis testing. To see this we note that before any turkey walked across the force plate, the farmer/breeder of the turkeys will likely have an initial guess at what the weight should be. After a measurement from the force plate has been recorded, the farmer would have a better estimate of the turkey's weight. However, given the measurements are noisy, the farmer has an estimate for the turkey's weight after the measurement, but with a degree of uncertainty. Assuming the farmer could force the turkey to repeatedly walk over the force plate, with more measurements, the farmer would hopefully become more certain about the birds weight if the readings were consistent, and less certain if the readings were very noisy or contradictory. The process of updating some initial estimate of the birds weight as data becomes available is naturally modelled using a Bayesian updating scheme and can be posed as a hypothesis testing problem. Furthermore, if a farmer wished to know a turkey's weight to a given confidence, this approach would predict when the farmer can remove a turkey from the force plate, and when he requires more readings and longer measurement periods. For a more detailed treatment of the mathematics underlying this section we recommend the reader to Grindrod [3, Chapter 5].

To begin by establishing the mathematical framework, let us denote $\mathbb{P}(A)$ as the probability of an event A happening, $\mathbb{P}(A \cap B)$ as the probability of both A and B happening, and $\mathbb{P}(A \mid B)$ as the probability of A happening given that event B has happened (called a *conditional* probability). We can relate these three quantities using

Bayes' theorem [3, pages 219–223]

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \tag{2.3}$$

We can denote the set of previously recorded/known observations as D, and a new observation d. Before observing the data we may have a hypothesis H for what the turkey's weight is, and the alternative hypothesis H^c. Based on the newly observed data we would like to update our belief that H is true or false. This can be achieved by trivially manipulating eq. (2.3) into the Bayesian updating scheme

$$\underbrace{\mathbb{P}(H \mid D, d)}_{\text{Posterior}} = \underbrace{\frac{1}{\mathbb{P}(d)}}_{\text{Normalisation}} \underbrace{\mathbb{P}(d \mid H, D)}_{\text{Model}} \underbrace{\mathbb{P}(H \mid D)}_{\text{Prior}}. \tag{2.4}$$

The question is then, how can we apply Equation (2.4) to the problem of weighing a turkey? This is easily done, where the first step is to realise that the weight of the turkeys is a continuous property, and hence we must discretise the turkey's weight into certain levels. These could be very coarse discretisations, or arbitrarily fine discretisations, where for mathematical and computational purposes the latter will be more convenient. Our hypothesis then is that the weight w of a turkey is in a weight class W_i. In this case we can write Equation (2.4) in the form

$$\mathbb{P}(w \in W_i \mid D, d) \propto \mathbb{P}(d \mid w \in W_i, D)\mathbb{P}(w \in W_i \mid D) \tag{2.5}$$

where we have ignored the normalisation factor, (which is trivially implemented).

The next question is, what are the data which we observe? The observations ultimately come from the force plate in the form of a force profile as in Figure 2.4. From this we can extract various possible features to act as the new observation d and we show some possibilities in Figure 2.5. For the sake of ease and simplicity to demonstrate the method we will present the following results using the value of the maximal force difference as our observed feature d, which is depicted in the bottom left of Figure 2.5, which we call the variation in F_z.

The first and easiest part of implementing this scheme is coming up with a prior estimate for a turkey's weight. In the presence of prior knowledge, a good choice is to use the empirically observed distribution of weights for the turkeys. This prior distribution is shown in Figure 2.6. (An even better prior would be the last posterior distribution for an individual turkey on a previous occasion, provided it is available and the turkey can be identified).

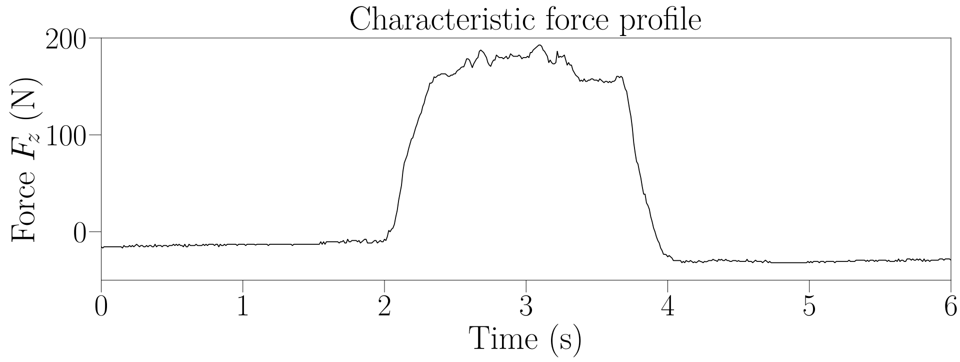


Figure 2.4: Characteristic force profile as a turkey walks over a force plate.

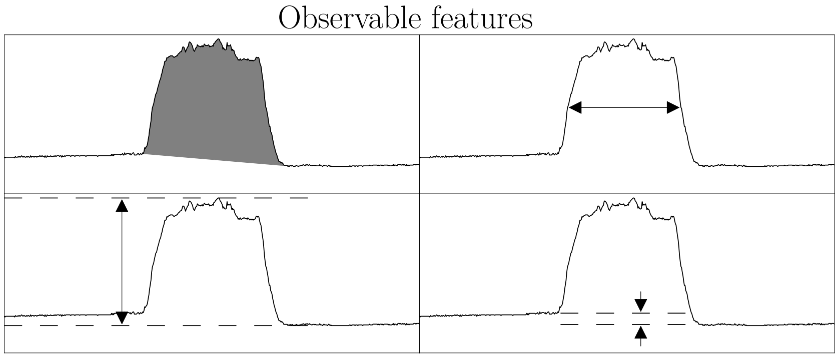


Figure 2.5: Some possible observable features which can be extracted from Figure 2.4. For much of our discussion we use the maximal force difference, which is the bottom left feature.

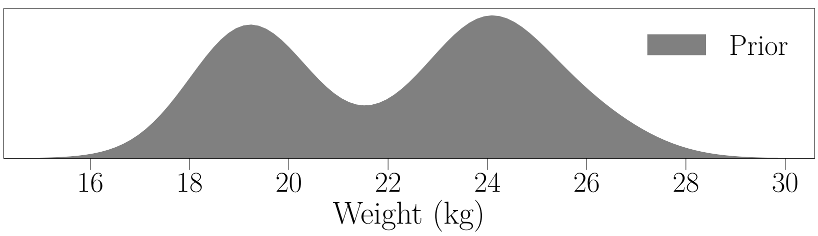


Figure 2.6: A prior model for the turkey's weight.

Having a prior, we can construct a model. The easiest way to do this is to empirically see if a feature is non-uniform across weight classes. This can be readily done by seeing the empirical distribution of a feature and the corresponding weights. The model can be constructed from this by a simple *kernel density estimate* (KDE), where for our purposes we use a Gaussian KDE. An example of such a Gaussian KDE is shown in Figure 2.7. Having constructed a model, we can evaluate the model update in Equation (2.5) for a given weight class by using the value of the normalised margin for the given weight class (depicted by the horizontal region in Figure 2.7). For a different weight class we update similarly, ensuring the marginals are normalised. After this update we only have to normalise the posterior distribution is normalised.

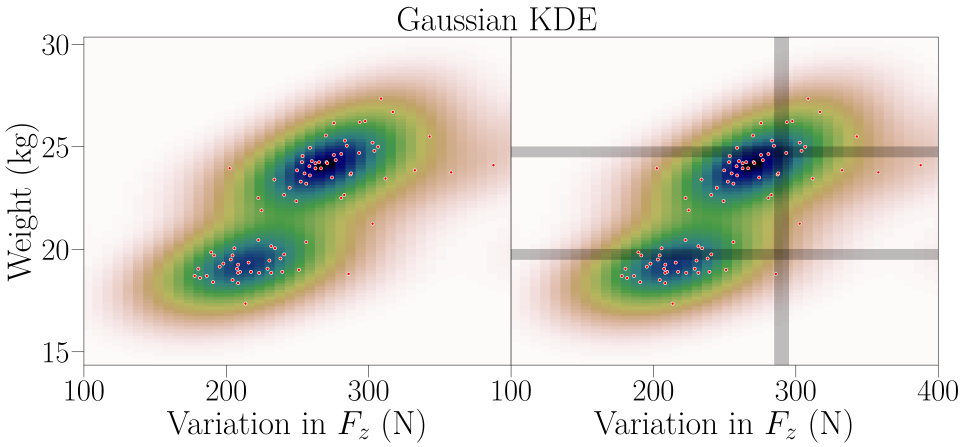


Figure 2.7: The Gaussian KDE formed by considering the variation in F_z. On the right we show two marginal cross sections for two different weight classes and a value for a new observation.

Having outlined how we construct the Bayesian model and apply Equation (2.5), it is informative to see how the prior changes for a given observation. A demonstration showing the difference between the prior and posterior is shown in Figure 2.8.

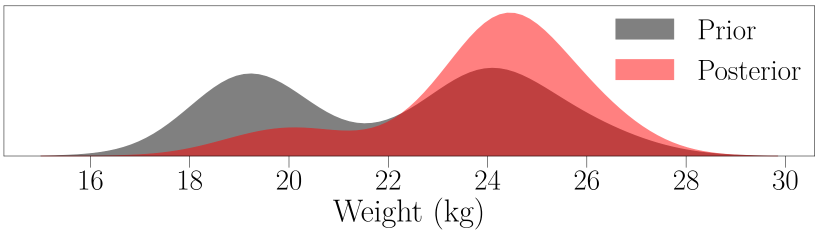


Figure 2.8: A posterior model of the turkey's weight.

It is then a matter of discretion about how to measure the accuracy of such a scheme. Should the predicted weight be the most likely weight, the expected weight, the median, etc. Similarly, how should the error be measured? As we have a probability distribution we could now measure the classification performance using either the RMSE, or a weighed RMSE, where we use the posteriors inverse variance as a weighting. There are several possibilities that can be explored here, and this is left as an avenue for further

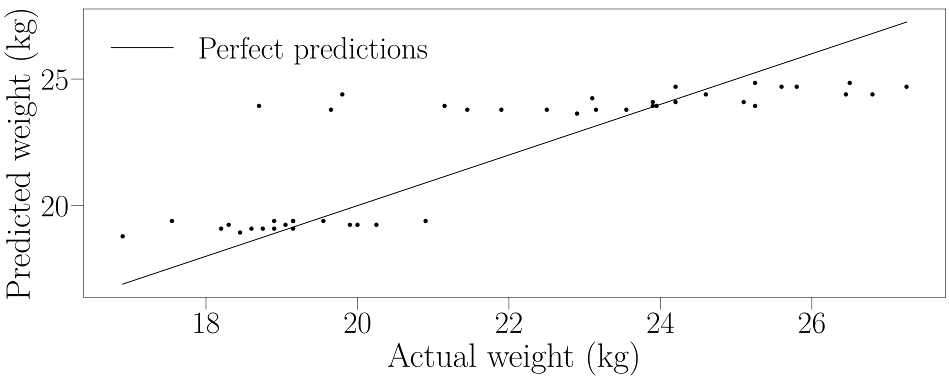


Figure 2.9: The points represent for each considered individual its predicted weight from a simple Bayesian hypothesis testing scheme.

investigation. If we take the most likely weight as our predicted weight, then we can compare the effectiveness of such a classification scheme, as shown in Figure 2.9.

To see the improvement achieved from only using a single feature (recall that we suggested several in Figure 2.5) we can compare the RMSE from using only the prior to the posterior. Using just the maximal value of the prior the RMSE was 4.3 kg, and using the mean of the prior gives a RMSE of 2.8 kg. Using just the maximal value of the posterior gives an RMSE of 2.1 kg, and using the mean gave 2.0 kg, although this latter improvement in going from the maximum likelihood estimate to the mean is likely not statistically significant.

2.3.2 Sparse Bayesian Generalized Linear Model

The first step in this work was to clean up and visualize the data in aid of the analysis. First, we removed duplicate data. Second, we inspected the time series of the 3D forces per turkey. We classified the time series manually in terms of their shape and baseline offset. Furthermore, we eliminated incomplete or dubious time series.

We extracted the main peaks from the time series to filter out the signals before and after a turkey walked over it. We extracted the beginning and end-points from the (vertical) z-force time series as they are the more consistent in shape than the x-, and y- forces. We performed this selection with the help of two lines defined by the initial, final, maximum and mean observations (Figure 2.10). The horizontal coordinate of the maximum observation and the mean of the time series formed one point of both lines. The initial and final points defined the second points of both lines. The beginning and end of the measurements were defined by the maximum distance between these liness and the time series. The reasonable location of the starting and end points were manually inspected for all turkeys. The baseline was defined as the line between the extracted starting and end points and was corrected for. The baseline of the forces in the x-, and y- directions were corrected for in a similar fashion, but the signals themselves were transformed on absolute scale.

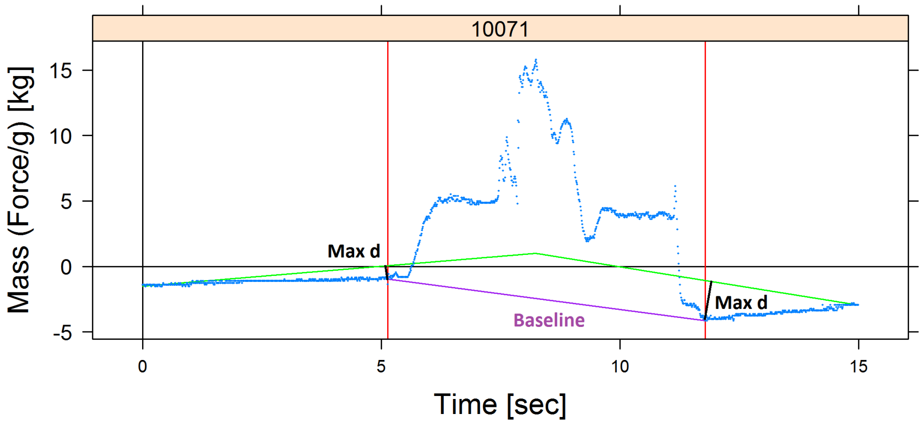


*Figure 2.10: Extracting the measurement portion of the z-force (vertical) time series for a turkey with Wing band number 10071. The vertical axis reflects the measured mass, and the horizontal axis the time. The blue points are the measurements. The green lines connect at the mean force during the entire measurement process at the horizontal location of maximally measured force. They start at the beginning and end of the time series. The maximum distance (**d**) below the light green lines and above the measurement curve determines the start and end of the turkey measurement process as visualized by the horizontal red lines. The purple line connects the start and end point of the measurement process and marks the adjusted linear baseline.*

The bodyweights Y of turkeys i were modelled according to eqs. Equation (2.6)-Equation (2.10), where *Normal* and *Cauchy+* indicate normal and positive Cauchy distributions; μ_i is the expected weight of turkey i; σ is the scale (standard deviation) of the measurement noise; α is the average weight of turkeys from line 1; $Line2_i$ is a logical variable that indicates if turkey i belongs to line 2; β is the difference between the average bodyweights of turkeys from lines 2 and 1; $\gamma_{j,k}$ indicates the effect of a force curve quantile k for the j force (mx, my, or mz); $X_{j,k}$ indicates the quantiles k for the force j at 1% intervals (i.e. 0% , 1% , 2% , ..., 100%) for each turkey i; τ is the global scale parameter of a horseshoe prior, whereas λs are the local scale parameters; p_0 are the expected number of non-zero parameters in the model, out of all D parameters; n is the total number of observations (turkeys); and τ_0 is the scale of the parameters. The normal distribution is used to capture the measurement noise of turkey weight estimation. We estimated the average bodyweight per genetic line of turkeys and refined the estimation per turkey with the help of the x-,y-, and z-force quantiles. It should be noted that the quantiles of the x-, and y- forces were taken from their absolute measurements. In addition, the quantile predictors were standardized by subtracting the mean and dividing by the standard deviation. The effects of the quantiles were modelled with a horseshoe prior. The horseshoe prior induces sparsity (i.e. near zero effects) if the effects of the predictors were too small to be distinguished from the measurement noise. It models the effects as normal distributions with Cauchy distributed scales. The global scale parameter τ is related to the scale of the measurement noise σ and determines the effects (parameters) that are large enough to be explained by the measurement noise alone.

$$Y_i = normal\left(\mu_i, \sigma\right) \tag{2.6}$$

$$\mu_i = \alpha + \beta \cdot Line2_i + \sum_{j=1}^{n_{forces}} \sum_{k=1}^{n_{quantiles}} \gamma_{j,k} \cdot X_{j,k} \tag{2.7}$$

$$\gamma_{j,k} \sim Normal\left(0, \tau \cdot \lambda_{j,k}\right) \tag{2.8}$$

$$\lambda_{j,k} \sim Cauchy\left(0, 1\right)^{+} \tag{2.9}$$

$$\tau = \frac{p_0}{D - p_0} \cdot \frac{\sigma}{\sqrt{n}} \cdot \tau_0 \tag{2.10}$$

In the spirit of Bayesian statistics, we provided prior information for the parameters. We assigned a standard normal prior to the measurement error since the force plates are expected to provide accurate measurements within 1 kg accuracy. Furthermore, we put a normal prior with mean of 20 kg and standard deviation of 2.5 kg since the average weight of adult turkeys is expected to be in the range of 15-25 kg on the basis of the information provided in the presentation from Hendriks Genetics. The scale of the difference between lines should be the same as the scale of the average turkey weight. Also, the standardized quantiles are expected to improve the accuracy of turkey estimates beyond the turkey line average by no more than the scale of the average turkey weight. Thus, the global scale of the parameters was set to 2.5 kg.

Lastly, we estimated the leave-one-out root-mean-square-error (LOO-RMSE) by pareto smoothed importance sampling (PSIS). This is a fast approximation to the LOO-RMSE obtained from leave-one-out cross validation.

Table 2.1: Prior distributions for the parameters in the model.

Parameter	***Designation***	***Prior***
σ	Scale of measurement error	Normal(0,1)
α	Average weight of turkeys from line 1	Normal(20,2.5)
p_o	Expected non-zero parameters in the model	10.5 out of 21 (i.e. 50%)
$\tau\ _o$	Scale of parameters	Cauchy+(0,2.5)

All analyses were carried out with the program *R*, and the packages *gdata*, *lattice*, *ggplot*, *rstan*, rstanarm, *loo*, *brms*, *plyr*, *dplyr*, and q*uantmod.*

2.3.3 Appliance of Machine Learning algorithms

In this approach, the aim was to discover the capacity of the regressor models to predict the target value, based on the attributes that complement it. For this, tests have been carried out with Machine Learning regression algorithms: *Random Forests* (RF) and *Multiple Linear Regression* (MLR). In the same way, we wanted to know its performance using all the attributes (a.k.a. variables), or using only those that were more correlated with the target variable.

MLR has been used because it is the most basic regression algorithm. With this, it was intended to demonstrate the usefulness of the attributes and their selection to solve the problem. On the other hand, RF has been used to check if with a more sophisticated regression algorithm the results can vary.

First the data has been pre-processed. As can be seen in Figure 9, there are entries in which the peaks are not clearly defined. In some cases there is more than one peak for the same entry (5), and in others the capture of the signal has not been completed (7). Therefore, we decided to eliminate those data that contain incomplete or duplicate entries. For the rest of the cases, the signal has been cut to work only with the part in which the peak is represented. Once the signal relative to the peak has been cut, the equivalent cut-off signal is obtained for the rest of the signals.

All these analyses were carried out with the programming language *Python*, and the libraries *matplotlib, pandas, scikit-learn* and *numpy.*

2.4 Results

2.4.1 Sparse Bayesian Generalized Linear Model

Characteristic measurement time series can be seen in Fig. 2. The typical time series (Figure 2.11.1) constitutes of two peaks in the vertical z-force that likely reflect the two steps of a turkey. However, the second peak is always lower than the first. What is more, the baseline does not return to 0 after a bird steps of the force plate but goes to negative values. The direction, number of peaks, and shape of the x- and y- force time series are highly variable and unpredictable. Many turkeys have been measured before the z-force measurement returned to the null baseline (Figure 2.11.2). In some time series it looks like that if a turkey walked fast and that the two steps appear as 1 (Figure 2.11.3), whereas others seem to show a turkey making many slow steps (Figure 2.11.4). Unfortunately, there are time series with two distinct signals (Figure 2.11.5), though sometimes it is possible to filter out the likely walk of the turkey from the (smaller) sensor disturbance (Figure 2.11.6). Finally, there are incomplete time series that suggest that the measurement process started too late (Figure 2.11.7). Turkey measurements of the later type were removed from further analysis.

We extracted the portion of the time series of all forces when the turkey walked by filtering the time series of the z- force (Figure 2.12.1). We then adjusted for the dynamically sinking or raising baseline of all forces (Figure 2.12.2-3). This step did not seem to fix the consistently higher z-force(vertical) peak of the first turkey step, but it did centre the time series around the null baseline. The algorithm that we employed seemed to work on manual inspection of the time series of all turkeys (not shown). It should be noted that the estimated baseline for the x-, and y- forces can be positive because the signal can be positive or negative depending on which leg the turkeys make the first step with.

Fitting the model on the full dataset yielded a measurement error with a scale (standard deviation) of 0.75 kg with a standard error of 0.05 kg. This implies that for 95 (%) power, the same turkey needs to be measured ca. 5848 times to estimate its weight with the desired within 0.05 kg accuracy. This method is comparable to the others as the correlation between the fitted values and the data is 0.93 (Figure 2.13.1), and cross validation of the model resulted in an LOO-RMSE of 0.85 (Figure 2.13.2). The residuals appear homogeneous, although there are a number of outliers.

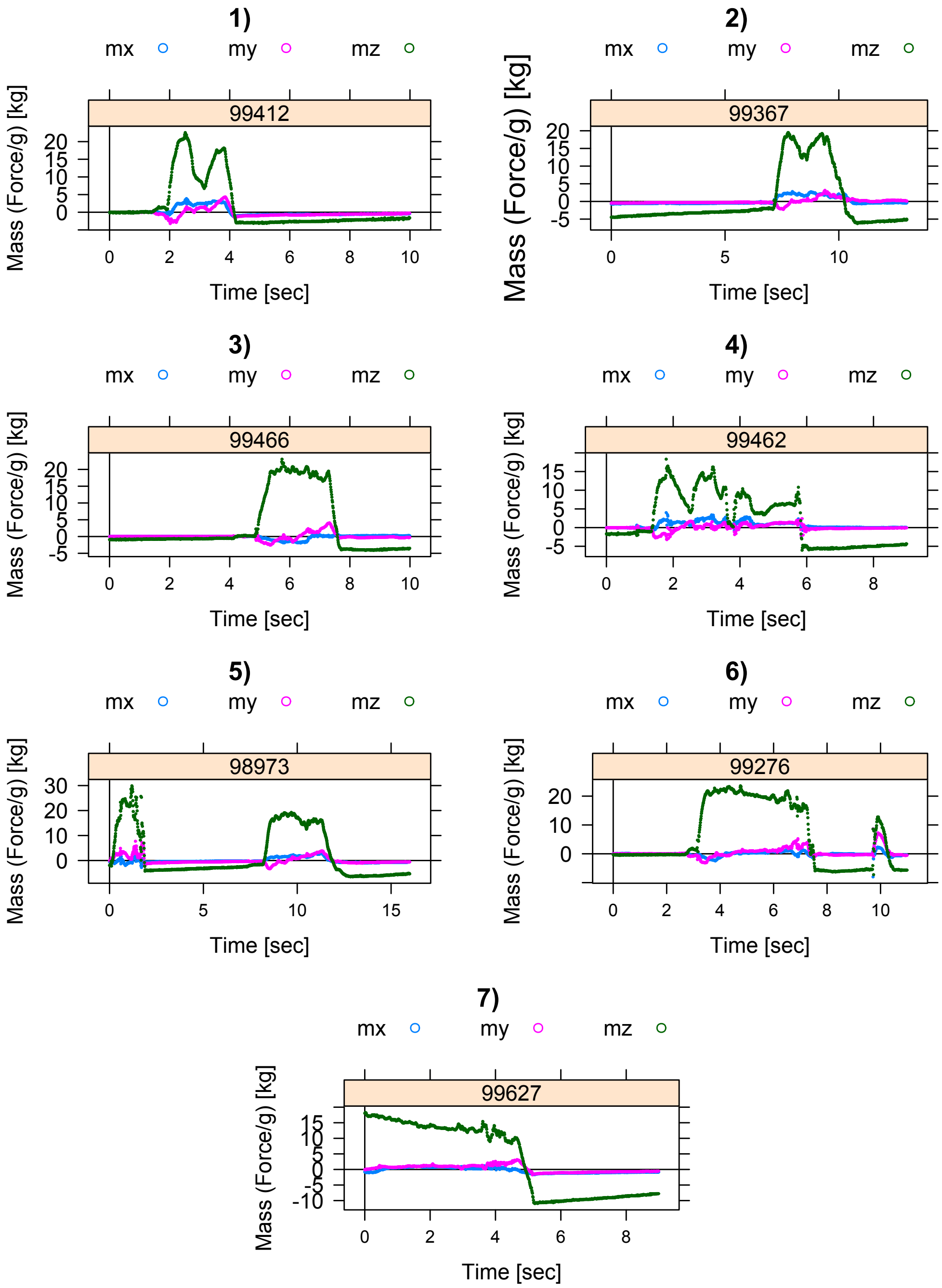


Figure 2.11: Typical time series of the 3D forces in the x-, y, and z- directions, or noted as mx (blue), my (red), and mz (green). The vertical axis reflects the measured mass, and the horizontal axis the time. 1) Typical time series with peaks for 2 steps, 2) Time series with offset, 3) Time series with unified 2 steps, 4) Time series with multiple steps, 5) Time series with 2 possible measurements, 6) time series with disturbances after the measurement, 7) incomplete time series.

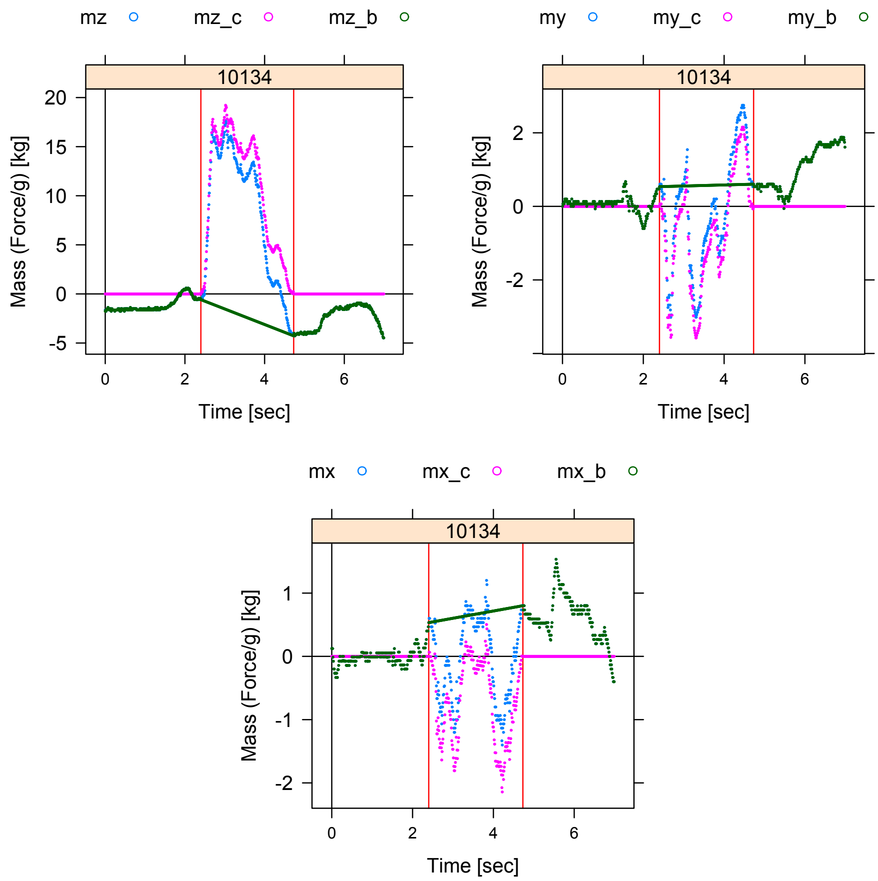


Figure 2.12: A sample force time series (blue) with corrected baseline (green), and adjusted measurements (red). The subplots indicate the 3D forces (mz, my, and mx) with suffices indicating the baseline corrected (_c) time series, and the baseline (-b).

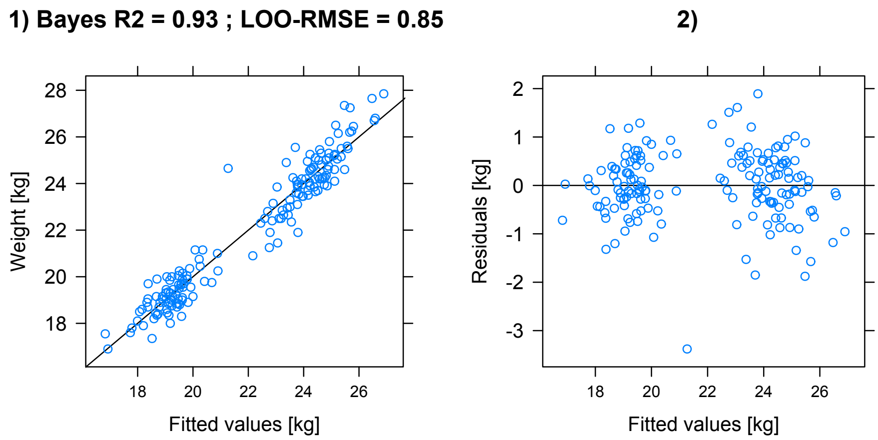


Figure 2.13: Quality of the predictions. 1) Fitted vs measured values, 2) residuals.

2.4.2 Learning Curves with Different Features

In this subsection we demonstrate the influence of the selected features for the prediction. We essentially demonstrate that most information is already captured by the integral of F_z (see also other sections) and the bloodline. First we describe the different types of features we test.

Bloodline As there are two bloodlines of turkeys, the bloodline takes either the value 1 or 2. More information on that is in earlier sections.

Integral F_z This is indeed the integral of the force F_z, so the force in the z-direction, from the cropped signal. So essentially the value calculated in Equation (2.1).

Summary With summary we captured a summary of the raw cropped signal from each of the 8 channels. For that we divided each cropped channel in 5 intervals and stored the mean value of each interval. This leads to 40 features per Turkey.

To compare the different features we show corresponding learning curves in Figure 2.14 as well as the leave-one-out error on the full dataset as shown in Table 2.2. For all curves we used a simple linear ridge regression scheme with regularization set to 0.1. The bloodline alone gives already a very good baseline with a leave-one-out root-mean-sqaured-error (LOO-RMSE) of 1.25, while the integral alone only achieves a LOO-RMSE of 2. Nevertheless, there is still valuable information in the integral, as when we add it to the bloodline we achieve a LOO-RMSE of 1.14. Although it does not seem much of an improvement, it is still significant with a p-value of 0.0082.[7] Through the learning curves we see that more data will not help when using those simple features, at least when using a simple linear regression scheme.

The summary features are able to outperform the integral feature alone, which indicates that there is more to learn from the channels than the force in the z-direction alone. Unfortunately though, the summary features together with the bloodline cannot outperform the simple integral and bloodline features.

For the summary feature we also observe that there is still a gap between the test and train error in the learning curve, so more data could help in this case. We note, however, that even in this case we expect a LOO-RMSE around 1 as this is the value that the

[7]The significance test was done with a paired two-tailed t-test.

training error approaches. So in case we have more data and we would use the summary feature we need a model that is more flexible than a simple linear regressor.

Features	Summary	Summary+Bloodline	Bloodline	Integral	Integral+Bloodline
RMSE	1.59	1.45	1.25	2	1.14

Table 2.2: The RMSE based on leave-one-out on the complete dataset.

2.4.3 Application of Machine Learning algorithms

There are three different tests to compare the performance of the regressors: (i) All: using all variables (the raw signals offered by the company and the calculated weights); (ii) Only Weights: using the weights calculated from the original variables; (iii) Correlated: using only the initial and transformed variables that were most correlated with the target label.

The results are reflected in table 2.3. The data are relative to the Root Mean Squared Error (RMSE):

Table 2.3: Results obtained by the application of Machine Learning algorithms.

Algorithm	**Only Weights**	**All**	**Correlated**
Multiple Linear Regression	1.167	1.125	1.034
Random Forests	1.142	1.058	0.706

Two conclusions can be obtained from these results: (i) applying the *Random Forests* algorithm with a configuration of 10 trees for the training, a lower RMSE is obtained than in the case of the *Multiple Linear Regression* in any case. (ii) Significantly better results are obtained using only the correlated features than all at the same time.

This leads to the consideration that the weight of the turkey is actually correlated with specific attributes, and that there are signals that are not relevant when generating the prediction. On the other hand, another of the main conclusions is that there have not been large amounts of data. This is due to the fact that a quantity of the captured data was incomplete or noisy, and it has been decided to eliminate from the historical data for the learning process. Possibly, if more clean and valid data were available, the prediction would be better through Machine Learning algorithms like those shown.

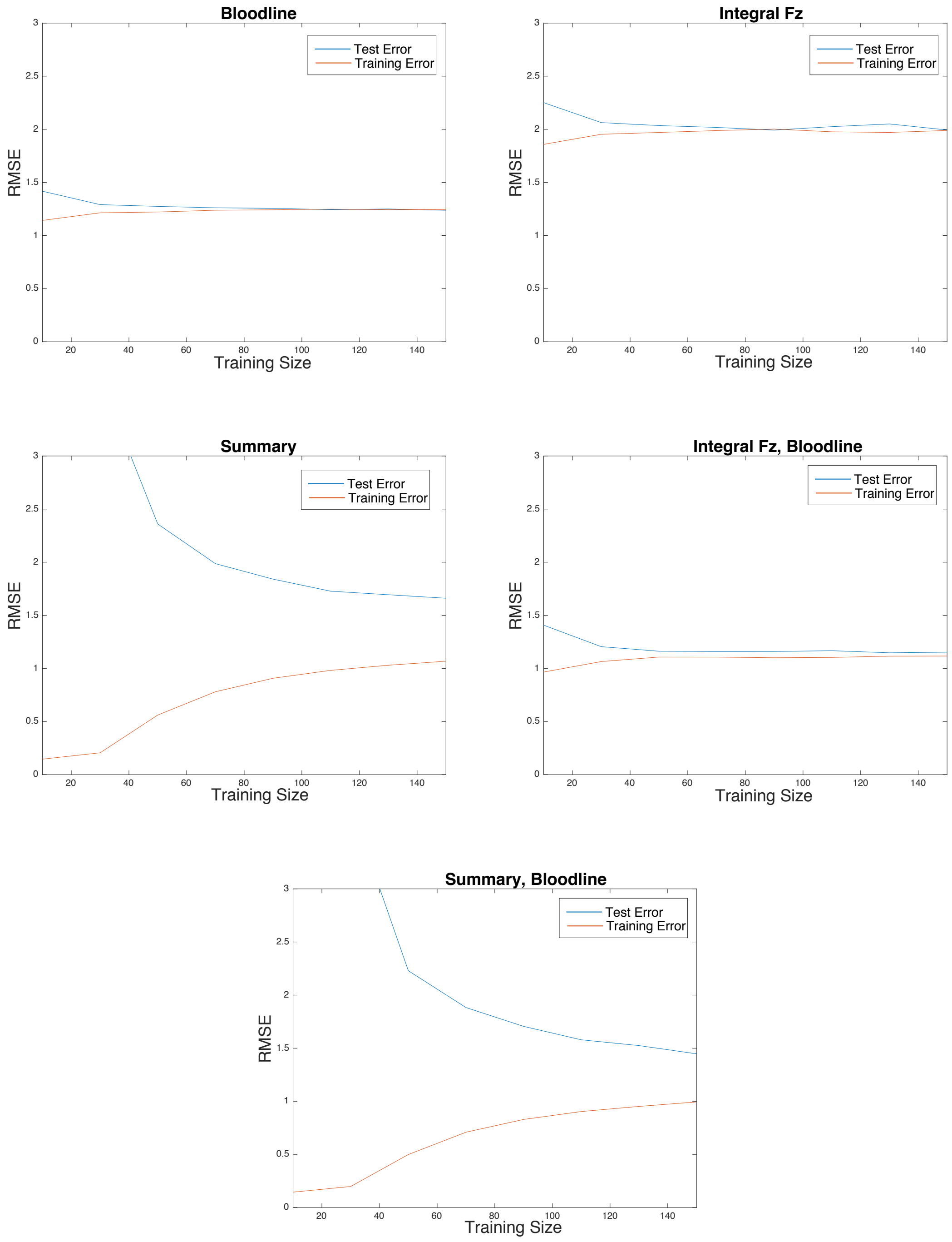


Figure 2.14: The five different learning curves when using different sets of features. We always use ridge regression with a linear Kernel and the regularization parameter set to 0.1. On the top of each plot we indicate which features we used, while we plot the root mean squared error (RMSE) vs the training set size. We show the RMSE as measured on the training set as well on the test set.

2.5 Conclusion, Discussion and Further Research

2.5.1 Conclusion

Our Bayesian hypothesis testing model requires more than one step from the turkeys. With every step from the same bird, the model can calculate and update the prior knowledge and produce much accurate results. The current model first give a guess that the body mass of a turkey is equal to the sample mean value.This guess produces initial RMSE of more than 4 kg. By taking one update of one step from the turkey, the posterior reduces the RMSE to around 2 kg. Therefore, the advantage of the first model is accuracy when there are more walking steps from the turkeys. From theoretical calculation, the accuracy could be good with information of 5 to 10 steps. Whereas, with the current limited data, this model gives the highest error.

Our second model - Sparse Bayesian generalized linear model - take the offset of signals into account. It applied Bayesian theory on learning the information from the time series. This method produces the best result by far and the RMSE is 1.14 kg. This is relatively good, but the the speed of increasing the accuracy is relatively slow compare to our first model. However, when more data are available, the model can be updated and produce more accurate results.

Our last two models uses basic machine learning: linear regression and random forest. These two are the simplest models but produces the best results regardless of the simplicity. The RMSE from Linear Regression is 1.034 kg and the RMSE from Random Forest is 0.706 kg. The input training data used the original noisy data without any baseline offset nor correction. The advantage of these two models are that they are really easy to build and gives really good results even with strong influence of noise. Disadvantage is that the improvement of these models heavily relies on the cleaning of noise. When we have clean data, the potential of increasing of accuracy will be quite limited.

Regarding the feature selection process we come to the following conclusion. Section 2.14 shows the integral feature (as defined in that section) together with the bloodline captures the most interesting features from the dataset. The summary feature suggests, however, that if we have more data available we could extract more information from the channels, possibly with more flexible regression schemes than linear regression.

2.5.2 Discussion

A few things need to be discussed before arguing how good our models are. The goal of this project is to explore the possibility of making prediction of turkeys' body mass within the error of 0.05 kg. However the trouble to verify the accuracy of our models is caused by the missing knowledge of the force plate. For example, if the equipment itself could cause errors of more than 0.5 kg, then such measurement mistakes can only damage the accuracy of our models which directly relays on the accuracy of measured data. Such systematical errors could be only reduced from the measurement side not from the modeling.

Another question is the reliability of 0.05 kg as an accuracy indicator, because body weight of either human or fowls has a range of values to move within a day. Especially after meals or drinking water, body mass measured could easily have a increase or decrease of roughly 0.5 kg. Therefore, in order to scientifically justify the comparison, the measurements should all be collected within a specially chosen time period, e.g. before meal time.

A third thing that caused our attention is the noise in the given data. Most of the noise could be avoided if certain methods being considered when recording those data. Here we suggest a few practical methods for helping to collect data:

1) One way to take the most of our Bayesian Hypothesis model is by collecting more steps. This can be achieved by placing three more plates along the current one. So each plate can collect one to two steps and four together can give a really good number of step information for our first model to get well updated.

2) To improve the overall accuracy, we need to reduce the noise of collected data, especially the problem of negative baselines. This can be avoided by allowing a short break between turkeys walking on the plate. This will allow the voltage signal to come back to the zero value.

3) We need to carry out a few simple tests to help to understand the force plate. Staff can step on or walk over the plate with different break between them. This could show the accuracy and performance of the equipment between measurements. People who will develop further models based on our study should have some knowledge of those force plates. A discussion with the company who build the plates can be helpful.

4) In general research experiment, force plates are mainly used in labs. The surface of plates are clean and directly touched or connected to the experimented subjects. However in order to make sure the turkeys are behaving in a natural way, the force plate used here is covered with mud and feathers. How this coverage is influencing the plate should be understood as well.

2.5.3 Further Research

Our models presented in this report are relatively basic due to time limitations. We have achieved good results but more sophisticated models should be developed based on our conclusions and results.

When clean data are available, we strongly suggest to run the models on the new data to see the accuracy level. Then, based on the improvement suggestions, new data should be collected and models can use those information to update themselves.

The time limit during the project also prevented us develop other models to make the whole process automatic. We went through the plots of time series of each turkey manually, which is not the most ideal and efficient method. Advanced machine learning models can be developed for pattern recognition - automatic working to speed up the data process for the company. Another direction to improve our machine learning model is to develop deep learning algorithms to understand the data and situation further.

Bibliography

[1] A. Bonnefoy and S. Armand. Normal gait. *Orthopedic Management of Children With Cerebral Palsy: A Comprehensive Approach*, (3):567, 2015.

[2] Rod Cross. Standing, walking, running, and jumping on a force plate. *American Journal of Physics*, 67(4):304–309, 1999.

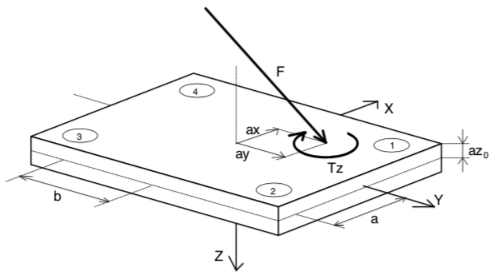
[3] P. Grindrod. *Mathematical Underpinnings of Analytics: Theory and Applications (1 edition)*. Oxford University Press, Oxford, January 27, 2015.

[4] Ming Tan, Yehuda Weizman, and Konstantin Fuss. Measurement accuracy of the body weight with smart insoles. *Proceedings*, 2(6), 2018. ISSN 2504-3900. doi: 10.3390/proceedings2060274. URL `http://www.mdpi.com/2504-3900/2/6/274`.

Appendix

2.A Force plate formulae

Kistler Force Plate Formulae



Force plate output signals

Output signal	Channel	Description
fx12	1	Force in X-direction measured by sensor 1 + sensor 2
fx34	2	Force in X-direction measured by sensor 3 + sensor 4
fy14	3	Force in Y-direction measured by sensor 1 + sensor 4
fy23	4	Force in Y-direction measured by sensor 2 + sensor 3
fz1 ... fz4	5 ... 8	Force in Z direction measured by sensor 1 ... 4

Calculated parameters

Parameter	Calculation	Description
Fx	= fx12 + fx34	Medio-lateral force 1)
Fy	= fy14 + fy23	Anterior-posterior force 1)
Fz	= fz1 + fz2 + fz3 + fz4	Vertical force
Mx	= b * (fz1 + fz2 - fz3 - fz4)	Plate moment about X-axis 3)
My	= a * (-fz1 + fz2 + fz3 - fz4)	Plate moment about Y-axis 3)
Mz	= b * (-fx12 + fx34) + a * (fy14 - fy23)	Plate moment about Z-axis 3)
Mx'	= Mx + Fy*az0	Plate moment about top plate surface 2)
My'	= My - Fx*az0	Plate moment about top plate surface 2)
ax	= -My' / Fz	X-Coordinate of force application point (COP) 2)
ay	= Mx' / Fz	Y-Coordinate of force application point (COP) 2)
Tz	= Mz - Fy * ax + Fx * ay	Free moment, Vertical torque, „Frictional“ torque
COFx	= Fx/Fz	Coefficient of Friction x-component
COFy	= Fy/Fz	Coefficient of Friction y-component
COFxy	= sqrt (COFx^2 + COFy^2)	Coefficient of Friction absolute

All formulae are in Kistler coordinate system

1) Walking direction is positive Y-axis

2) az0 = top plane offset (negative value)

3) a, b = sensor offset (positive values)

Figure 2.15: Force plate formulae.