# Method: Using generalized additive models in the animal sciences

G. L. Simpson[a]

[a] *Aarhus University, Department of Animal and Veterinary Sciences, Blichers Allé 20, Tjele, Denmark, 8830*

---

**Abstract**

Nonlinear relationships between covariates and a response variable of interest are frequently encountered in animal science research. Within statistical models, these nonlinear effects have, traditionally been handled using a range of approaches, including transformation of the response, parametric nonlinear models based on theory or phenomenological grounds (e.g., lactation curves), or through fixed spline or polynomial terms. If it is desirable to learn the shape of the relationship from the data directly, then generalized additive models (GAMs) are an excellent alternative to the traditional approaches. GAMs extend the generalized linear model such that the linear predictor can include one or more smooth functions, that are parameterised using penalised splines. A wiggliness penalty on each function is used to avoid over fitting while estimating parameters of the spline basis functions to maximise fit to the data without producing an overly complex function. Modern GAMs now include automatic smoothness selection to find an optimal balance between fit and complexity of the estimated functions. Because GAMs learn the shapes of functions from the data, the user can avoid forcing a particular model to the data. Here, I provide a brief description of GAMs visually illustrate how they work. I then demonstrate the utility of GAMs on three example data sets of increasing complexity, to show i) how learning from data can produce a better fit to data than parametric models, ii) how hierarchical GAMs can be used to estimate growth data from multiple animals in a single model, and iii) how hierarchical GAMs can be used for

formal statistical inference in a designed experiment of the effects of exposure to maternal hormones on subsequent growth in Japanese quail. The examples are supported by R code that demonstrates how to fit each of the models considered, and reproduces the results of the statistical analyses reported here. Ultimately, I shown that GAMs are a modern, flexible, and highly usable statistical model that is amenable to many research problems in animal science, and deserve a place in the statistical toolbox.

*Keywords:* Generalized additive model, Penalized spline, Hierarchical model, Conditional effects, Basis function

---

## Implications

Nonlinear relationships between covariates and a response variable of interest are frequently encountered in animal science research. Generalized additive models and automatic smoothness selection via penalised splines provide an attractive, flexible, data-driven statistical model that is capable of estimating these relationships. I provide a description of the Generalized additive model and demonstrate their use on three typical data examples; i) a lactation curve, ii) growth curves in commercial pigs, and iii) a experiment on the effects of maternal hormones in growth rates in Japanese quail.

*Corresponding author
*Email address:* gavin@anivet.au.dk (G. L. Simpson)

## Specification table

| | |
|---|---|
| Subject | Livestock farming systems |
| Specific subject area | Todo |
| Type of data | Table, graph, code |
| How data were acquired | Lactation curve: unstated in original source.<br>Pig growth data: weight measurements derived from a depth camera (iDOL65, dol-sensors a/s, Aarhus, Denmark) and a YOLO (you only look once) algorithm.<br>Quail growth experiment: body mass recorded using a digital balance. |
| Data format | Lactation curve: processed (averaged).<br>Pig growth data: processed (averages of multiple depth camera-based weight measurements).<br>Quail growth experiment: Raw. |
| Parameters for data collection | Lactation curve: daily fat content of milk for a single animal (cow 7450).<br>Pig growth data: Pigs raised under conventional husbandry conditions at two farms.<br>Quail growth experiment: 158 eggs from adult Japanese quails provided by Finnish private local breeders were injected with one or a combination of maternal hormones or a saline solution control, and those eggs that hatched successfully were monitored for 78 days for a range of parameters including body mass. |
| Description of data collection | Lactation curve: fat content of daily milk production was measured for a single cow.<br>Pig growth data: a depth camera observed the pigs *in situ* and a computer algorithm converted the digital imagery of individual animals into weight estimates.<br>Quail growth experiment: the body mass of hatched quail was recorded at 12 hours post hatching, once every three days for days three to 15, and weekly thereafter until day 78. |
| Data source location | Lactation curve: unknown.<br>Pig growth data: Data from two farms were reported in the original study: A commercial farm in Gronau, Germany, and the experimental farm of the Department of Animal and Veterinary Sciences, Aarhus University, Viborg, Denmark.<br>Quail growth experiment: Finland |
| Data accessibility | Repository name: Zenodo<br>Data identification number: 10.5281/zenodo.15777270<br>Direct URL to data: https://doi.org/10.5281/zenodo.15777270 |

## Introduction

Many research questions in the animal sciences involve nonlinear relationships between covariates and the response. A classic example, with a long history of statistically-based and mathematically-based research, is the plethora of models that have been described for the estimation of lactation curves from test day data or automatic milking machines. Another, is the estimation of growth curves in the context of breeding and genetics. Despite the frequency with which such nonlinear relationships are encountered, surprisingly little use of splines and generalized additive models (GAMs) in particular has been seen in animal science to date. Notable exceptions are…

In part, this lack of uptake reflects a traditional statistics workflow grounded in mixed effects modelling. Statistical training rarely includes more advanced models like GAMs, and GAMs are sufficiently different an approach that many academics may be wary of using them because they are unfamiliar with the nomenclature used to describe the models or software used to fit them. Where GAMs have been used in animal science, best practice is often not followed, for example in the choice of smoothness selection method, or users fail to adequately specify the conditional distribution of the response or transform the response to better meet the distributional assumptions of a Gaussian model (e.g. van Lingen et al., 2023).

More attention has been paid to the use of splines, frequently in comparisons against some form of polynomial-based model (e.g., Legendre polynomials). There, the focus has been on the choice of the number of knots in the spline basis expansion and on the placement of those knots. Modern GAMs largely make such choices redundant; with penalized splines, a wiggliness penalty is used to avoid over-fitting, and low-rank eigen bases, such as the low-rank thin plate regression spline basis of Wood (2003) avoid the knot-placement problem for most problems.

A guide to the use of GAMs in the animal science setting, written with those users in

mind, is needed, to raise awareness of the utility of these flexible models and to promote best practice in fitting GAMs. Below, I describe GAMs and demonstrate visually how they work. Then I apply GAMs to three different examples representing typical data encountered in animal science. The examples below are supported by tutorials containing the computer code needed to fit the models in the R statistical software (R Core Team, 2025), which are available along side the source code for the manuscript itself.

## Materials and methods

### *Generalized additive models*

A basic GAM has the following form

$$
\begin{aligned}
y_i &\sim \mathcal{D}(\mu_i, \phi) \\
\mathbb{E}(y_i) &= \mu_i \\
g(\mu_i) &= \mathbf{X}_i\boldsymbol{\gamma} + \sum_j f_j(\bullet),\ i = 1, \ldots, n;\ j = 1, \ldots, J
\end{aligned}
$$

where $y_i$ is a univariate response variable of interest that is modelled on the scale of a link function $g()$, $\mathcal{D}(\mu_i, \phi)$ is a distribution, typically from, though not limited to, the exponential family of distributions, with mean $\mu_i$ and scale parameter $\phi$. $\mathbf{X}_i$ is the $i$th row of the model matrix of any parametric terms (including the model intercept or constant term), and $\gamma$ the associated regression parameters. There $f_j$ are $J$ smooth functions of one or more covariates; I use $\bullet$ as a placeholder for this definition, but in the simplest case of a smooth of a single continuous covariate we have $f_j(x_{ij})$, where $x_{ij}$ is a univariate covariate.

In the remainder of this section, I aim to present *just enough* detail about GAMs to afford the reader a general understanding of what is involved in estimating such a model, so

that they can appreciate how GAMs work, and how we aim to avoid overfitting or having to choose how complex each smooth function should be. The original approach for fitting GAMs known as *backfitting* required the analyst to specify how many degrees of freedom each function in the model should take *a priori*, which was viewed by many as being too subjective for more than exploratory analysis. With modern automatic smoothness selection methods, this problem has largely been resolved, through the use of penalized splines and fast algorithms for smoothness selection.

*Penalized splines*

In a GAM, the smooth functions $f_j()$ are typically represented in the model using splines, although other functions fit into the framework, most notably iid Gaussian random effects. A spline is composed of $K$ basis functions, $b_k()$, and their associated coefficients, $\beta_k$

$$f_j(x_{ij}) = \sum_{k=1}^{K} \beta_k b_k(x_{ij}).$$

For identifiability reasons, the basis is subject to a sum-to-zero constraint to remove the constant function from the span of the basis. This allows a separate constant term or intercept in the model, which is desirable if we also want to include categorical (factor) terms in the model. Figure 1a shows a B spline basis ($K$ = 16 for a covariate $x$) after the absorbing the sum-to-zero constraint into the basis.

Fitting the spline then involves finding estimates for the coefficients of basis functions, $\beta_k$. These coefficients weight the individual basis functions as shown in Figure 1b. To find the value of the fitted spline at each value of $x$, we sum up the values of the weighted basis functions evaluated at each value of $x$. This yields the light blue curve in Figure 1b. The coefficients for the basis functions, $\beta_k$, are determined by forcing the fitted function to go as close to the data as possible. As shown in Figure 1b, we largely recover the true function from which the data were simulated.

Expanding $\mathbf{x}_j$ in this way into many basis functions means we must be cognizant of the

risks of over fitting the sample of data we have to hand. Taken to the extreme, we could obtain an arbitrarily close fit to the data by taking as many basis functions as there are data (i.e., $K = n$), but all this would achieve in practice is the replacement of the data with a set of coefficients.

This raises the question of how many basis functions should be used? One option, if there are $n$ unique values of the covariate $x$, is to use $n$ basis functions, leading to a situation where our model would have as many coefficients as data. However, we do not gain anything by using $n$ basis functions from a statistical viewpoint. In practice we typically use $\mathcal{N} \ll n$ basis functions. Yet, even with just $\mathcal{N}$ basis functions, we face the very real situation that our model may overfit the data. To address this, modern approaches to fitting GAMs use penalized splines.

If our aim is to avoid overfitting, we need to penalise highly complex fitted functions, and hence need to define what we mean by "complex". A complex function would be one that "wiggles" about markedly as we move from low to high values of the covariate, $x$. Such wiggling around implies that the function has a high amount of curvature, which we measure as the second derivative of the fitted function. While there are other definitions for complexity that we use, this the typical measure used for splines in statistical models. Therefore, to measure the wiggliness of a fitted function we need to integrate, or sum up, the second derivative of the function $f$ over the range of $x$

$$\int_x f''(x)^2 \, dx = \boldsymbol{\beta}^{\mathsf{T}} \mathbf{S} \boldsymbol{\beta} \tag{1}$$

where $f''$ indicates the second derivative of $f$, and note that we are integrating the *square* of the second derivative because we need to allow for both negative and positive curvature as the functions wiggles. Conveniently, we can compute this integral as a function of the model coefficients $\beta$ as shown on the right hand side of Equation 1. $\mathbf{S}$ is a known penalty matrix, which encodes the complexity of the basis functions. The penalty matrix for the basis expansion shown in Figure 1a is displayed in Figure 1c.

Figure 2 shows three different estimated smooths for the simulated shown in Figure 1 and their wiggliness or integrated squared second derivative. Fitting a GAM, therefore, requires us to balance the fit to the data *and* the complexity of the resulting model; we wish to avoid both overfitting the data (Figure 2a) and over smoothing (Figure 2c) the data. We want the fitted functions to be just "wiggly enough" to approximate the true, but unknown relationships (Figure 2b).

We find estimates of the basis function coefficients, $\beta_k$, that make the fitted spline go as close to the data as possible; in practice, this is done by maximising the penalised likelihood

$$\ell_p(\boldsymbol{\beta}) = \ell(\boldsymbol{\beta}) - \frac{1}{2\phi}\sum_j \lambda_j \boldsymbol{\beta}_j^{\mathsf{T}} \mathbf{S}_j \boldsymbol{\beta}_j \ ,$$

where $\ell_p(\boldsymbol{\beta})$ is the penalised log likelihood and $\ell(\boldsymbol{\beta})$ the log likelihood of the data, given the $\boldsymbol{\beta}$, and the remainder is the wiggliness penalty for the model. The log likelihood ($\ell(\boldsymbol{\beta})$) measures how well our model fits the data, while the wiggliness penalties $\boldsymbol{\beta}_j^{\mathsf{T}} \mathbf{S}_j \boldsymbol{\beta}_j$ measures how complex the model is. Note that any parametric coefficients, $\boldsymbol{\gamma}$, including the intercept have been absorbed into $\boldsymbol{\beta}$ for convenience. The $\lambda_j$ are known as the smoothing parameters of the model, and it is these smoothing parameters that actually control how much the wiggliness penalty affects the $\ell_p(\boldsymbol{\beta})$. We can think of the $\lambda_j$ as tuning or hyper-parameters of the model.

As mentioned in the introduction, the knot placement problem can largely be averted through the use of a low rank thin plate regression spline basis (Wood, 2003). The spline basis used in the above explanation was a B spline basis, and for that basis the knots were placed at evenly-spaced quantiles of the covariate. In general, the exact positioning of the knots is unimportant, so long as they are spread out over the range of the covariate. If we wish to avoid the knot placement altogether, then we could replace the B spline basis with a low rank thin plate regression spline basis (TPRS). We start with a

basis function at each unique value of the covariate. This rich basis is then transformed and truncated through an eigen decomposition to retain the $k$ eigenvectors with the smallest eigenvalues. The decomposition removes much of the excessive wiggliness that $n$ basis functions would provide, while retaining many of the good properties of the original basis (Wood, 2003). The main downside of the TPRS basis is that it is computationally expensive to form the basis when setting up the model; for larger data problems, with many thousands of data, simpler bases such as the cubic regression spline or B spline may be preferred. In the *mgcv* software used here, the TPRS basis is the default basis used for univariate smooths (Wood, 2025, Wood (2011)).

The algorithms used to fit GAMs need to estimate the model coefficients, $\boldsymbol{\beta}$, and choose appropriate values of the smoothing parameters, $\lambda_j$. This process is known as smoothness selection, and there are several approaches to smoothness selection that can be taken. One way is treat the problem as one of prediction, and choose $\lambda_j$ in such a way as to minimise the cross-validated prediction error of the model. In practice it would be computationally costly to actually cross-validate the model for fitting, so we approximate the prediction error by minimising the generalised cross validation (GCV) error, AIC, or similar measure. An alternative means of smoothness selection is to take a Bayesian view of the smoothing process (see Miller, 2025, for an accessible introduction to this viewpoint); in doing so, the $\boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{S}_j\boldsymbol{\beta}_j$ should be viewed as multivariate normal priors on the $\boldsymbol{\beta}$. From this Bayesian view of smoothing, we find that the criterion we wish to minimise is that of a mixed effects model. Hence, we can think of the wiggly parts of the $f_j$ as being fancy random effects, while the smooth parts of the $f_j$ are fixed effects, and the $\lambda_j$ are inversely proportional to the random effect variances one would observe if the model were fitted as a mixed effects model. The Bayesian approach to smoothing can involve fully bayesian estimation using Markov chain Monte Carlo (MCMC) or simulation free estimation via the integrated nested Laplace approximation (INLA), or, using the equivalence of splines and random effects, we can take an empirical bayesian approach, which yields the posterior modes of the $\boldsymbol{\beta}_j$, or the maximum a posteriori (MAP)

estimates.

*Examples*

*Lactation curves*

As a simple illustration of the benefits of GAMs to learn the functional form of a relationship between a response variable and a covariate, rather then impose one through a parametric model, I reanalyse a small data set of average daily fat content per week of milk from a single cow (Figure 3). The data were reported in Henderson and McCulloch (1990). Initially, I followed their (Henderson and McCulloch, 1990) analysis and fitted a Gamma generalized linear model (GLM) with log link function. However, subsequent analysis of this model and the GAM alternative described below showed that the average daily fat data are under dispersed relative to the assumed conditional distribution. Instead, a Tweedie GLM (log link) was fitted, which has the same response shape as the gamma GLM but is more easily compared with the Tweedie GAM described below. The Tweedie GLM fitted had the following form

$$\texttt{fat}_i \sim \mathcal{T}(\mu_i, \phi)$$

$$\log(\mu_i) = \beta_0 + \beta_1 \log(\texttt{week}_i) + \beta_2 \texttt{week}_i \; .$$

Henderson and McCulloch (1990) compared the fit of the GLM model with several other formulations and models, including the model of Wood (1967)

$$\texttt{fat}_i = \alpha \, \texttt{week}_i^{\delta} \exp(\kappa \, \texttt{week}_i) + \varepsilon$$

where $\alpha$, $\delta$, and $\kappa$ are parameters whose values are to be estimated, and $\varepsilon$ is a Gaussian error term. The linear predictors in the GLM and Wood's model are equivalent

because of the log link function; $\beta_0 = \log \alpha$, $\beta_1 = \delta$, and $\beta_2 = \kappa$. However, we would not expect both models to produce the same fitted lactation curve as different distributional assumptions are being made; in the GLM version we assume the average daily fat values are conditionally Tweedie distributed, while in Wood's model they are assumed to be conditionally Gaussian distributed.

These models were compared with a GAM version of the log-link, Tweedie GLM, where I replaced the fixed functional form for the lactation curve with a smooth function to be estimated from the data

$$\texttt{fat}_i \sim \mathcal{T}(\mu_i, \phi)$$

$$\log(\mu_i) = \beta_0 + f(\texttt{week}_i)$$

Wood's model (Wood, 1967) was estimated via a nonlinear least squares model using the `nls()` function, in R (version 4.5.0, R Core Team, 2025). The Tweedie GLM and GAM were fitted using the `gam()` function of the *mgcv* package (version 1.9.3, Wood, 2025, Wood (2011)) for R. $k = 9$ basis functions were used for $f(\texttt{week}_i)$ after application of the identifiability constraint.

*Pig growth*

In this second example, I illustrate how individual growth curves can be estimated using a hierarchical GAM. A hierarchical GAM (HGAM, Pedersen et al., 2019) is a GAM that includes smooths are different hierarchical levels of the data structure; HGAMs are the equivalent of hierarchical or mixed effects models. In this example, different parameterisations of the model lead to either direct estimation of each individual animals growth curve, or a decomposition into an average growth curve plus animal-specific deviations from this average curve. Although previsouly covered by Pedersen et al. (2019) in detail,

a new addition here is the use of a constrained factor smooth to produce animal-specific deviations that are truly orthogonal to the average curve. This basis type was not available to Pedersen et al. (2019), and helps avoid the with to a first derivative penalty for the animal specific curves that would be required to make the model identifiable if a "factor by" was used.

Automated estimation of body weights is a useful on-farm method for continuously monitoring growth of commercial pigs. I analyse a subset of the weight data reported by Franchi et al. (2023) from a study by Bus et al. (2025), using data from a single pen of 18 pigs (Figure 4). The body weight data were obtained using a depth camera (iDOL65, dolsensors A/S, Aarhus, Denmark), and the each weight observation is the daily average of mutliple measurements made by the camera. As the number of weight measurements made each day varied per animal and per day, each weight observation in the data is the average of a variable number of measurements. To accomodate this, the model included observation weights, where the precision of each response data is proportional to the number of measurements averaged.

The data exhibit common and animal-specific variation. To model these features, four different GAMs were fitted to the pig weight data. All of the models ultimately provide animal-specific estimates of weight over time, but they decompose the growth curves in different ways. The weight data were assumed to be conditionally distributed gamma, with log link $\texttt{weight}_i \sim \mathcal{G}(\mu_i, \phi)$, with $\log(\mu_i) = \eta_i$, where $\boldsymbol{\eta}$ is the linear predictor. The linear predictors for each of the four models were:

$$\text{P1: } \eta_i = \beta_{a(i)} + f_{a(i)}(\texttt{day}_i)$$

$$\text{P2: } \eta_i = \beta_0 + f_1(\texttt{day}_i) + f_{a(i)}(\texttt{day}_i)$$

$$\text{P3: } \eta_i = \beta_0 + f^*_{a(i)}(\texttt{day}_i)$$

$$\text{P4: } \eta_i = \beta_0 + f_1(\texttt{day}_i) + f^*_{a(i)}(\texttt{day}_i)$$

where $a(i)$ indicates to which animal the $i$th observation belongs. Model 1 includes the mean weight of each animal through a parametric factor term, $\beta_{a(i)}$, plus a smooth of observation day $\texttt{day}_i$ *per* animal. The parametric factor term is required, because the animal-specific smooths in this model are each subject to the sum-to-zero-constraint and as such do not contain the constant functions that are needed to model average weight of each animal. These smooths are known informally as "factor by smooths". Model 2: decomposes the data into an average growth curve, $f_1(\texttt{day}_i)$, plus smooths, one per animal, that represent deviations from the average smooth, $f_{a(i)}(\texttt{day}_i)$. Unlike the *factor by smooths*, these deviation smooths do include constant terms for each animal's average weight, hence only a constant term, $\beta_0$ is include in the parametric part of this model. In both Model 1 and Model 2, the $f_{a(i)}(\texttt{day}_i)$ smooths each have their own smoothing parameters, allowing the wiggliness of each animal's growth curve to vary, if supported by the data.

Model 3 is similar to Model 1, except each animal's growth curve, $f^*_{a(i)}(\texttt{day}_i)$ shares a single smoothing parameter and hence assumes that the wiggliness of each curve is similar. These smooths are denoted with a superscript $*$ to indicate the shared smoothing parameter, and can be thought of as the smooth equivalent of random slopes and intercepts; informally, we refer to these and *random smooths*. These smooths are fully penalized and contain constant terms to model the average weight of each animal, hence only the intercept, $\beta 0$ is included in the parametric part of the model. Model 4 is similar

to Model 2 in terms of the decomposition into an *average* smooth and animal-specific smooth deviations, but like Model 3, the animal-specific smooths share a smoothing parameter and therefore assume they have similar wiggliness.

For further details on the different approaches used in the models described above, see Pedersen et al. (2019). Each of the smooths in the models used $k = 9$ basis functions, after application of identifiability constraints.

*Japanese quail*

In the third example, I demonstrate how to fit GAMs in the context of a designed experiment to obtain estimated treatment effects.

The data [Sarraude2020-cw] are from an experiment into the short- and long-term effects on quailof elevated exposure to different types of maternal thyroid hormone in Japanese quail, *Coturnix japonica* (Sarraude et al., 2020b). Briefly, the yolks of $n = 57$ eggs were experimentally manipulated by injection with the prohormone, $T_4$, its active metabolite, $T_3$, or both $T_4$ and $T_3$ ($T_3T_4$), or a saline solution that acted as a control (CO). Body mass was initially measured 12 hours after hatching. Between day 3–15, body mass was recorded every three days, then, between day 15–78, once per week using a digital balance.

The quail weight data were provisionally assumed to be conditionally gamma distributed. Despite this being a reasonable working assumption, model diagnostics identified deviations from this assumption, and instead the data were modelled as being conditionally Tweedie distributed. The Tweedie family of distributions contains the Poisson (power, $p = 1$) and gamma distributions ($p = 2$) as special cases, but is more flexible than gamma, and allows for a range of mean-variance relationships through the power parameter, $p$, of the distribution. In the Tweedie GAMs that were fitted, the power parameter $p$ was allowed to vary between 1–2 and was estimated as an additional model constant term. Models fitted were:

$$\text{Q1: } \eta_i = \beta_0 + f(\texttt{day}_i) + f_{\text{sex}(i)}(\texttt{day}_i) + f^*_{\text{egg}(i)}(\texttt{day}_i) + \xi_{\text{mother}(i)}$$

$$\text{Q2: } \eta_i = \beta_0 + f(\texttt{day}_i) + f_{\text{treat}(i)}(\texttt{day}_i) + f_{\text{sex}(i)}(\texttt{day}_i) + f^*_{\text{egg}(i)}(\texttt{day}_i) + \xi_{\text{mother}(i)}$$

$$\text{Q3: } \eta_i = \beta_0 + f(\texttt{day}_i) + f_{\text{treat}(i)}(\texttt{day}_i) + f_{\text{sex}(i)}(\texttt{day}_i) + f_{\text{treat}(i),\text{sex}(i)}(\texttt{day}_i) + f^*_{\text{egg}(i)}(\texttt{day}_i) + \xi_{\text{mother}(i)}$$

$$\text{Q4: } \eta_i = \beta_0 + f(\texttt{day}_i) + f_{\text{treat}(i)}(\texttt{day}_i) + f_{\text{sex}(i)}(\texttt{day}_i) + f_{\text{treat}(i),\text{sex}(i)}(\texttt{day}_i) + \psi_{\text{egg}(i)} + \xi_{\text{mother}(i)}$$

where $\texttt{day}_i$ is the number of days since hatching, treat$(i)$, sex$(i)$, and egg$(i)$ indicate to which treatment group, sex, and animal the $i$th observation belongs. Smooth functions with subscript treat$(i)$, sex$(i)$, or egg$(i)$ are factor-smooth interactions representing deviations from the "average" smooth, $f(\texttt{day}_i)$, for the indicated factor. $f_{\text{treat}(i),\text{sex}(i)}(\texttt{day}_i)$ represents a higher order factor-smooth interaction, which allows the time varying treatment effects to also vary between male or female quail. $\psi_{\text{egg}(i)}$ and $\xi_{\text{mother}(i)}$ are iid Gaussian random intercepts for individual quail and their mother respectively. A $f^*$ represents a random smooth, where each smooth in the set shares a smoothing parameter. The factor-smooth interactions in this model were fitted using the constrained factor-smooth interaction basis in the *mgcv* package; this basis excludes the main effects (and low-order terms in the case if higher-order interactions) to insure that the basis functions are orthogonal to those lower order terms.

Model Q1 represents a null model containing no treatment effects but models the remaining features of the data, decomposing the growth curves into an average effect, a sex-specific effect, and individual quail-specific effects, plus a maternal effect. Model Q2 extends Q1 by adding the treatment effect, $f_{\text{treat}(i)}(\texttt{day}_i)$, which models deviations from the average curve for each of the four treatment levels. Model Q3 further extends the model to allow different treatment-specific curves for male and female quail. Model Q4

is a variant of Q3, replacing the quail-specific growth curve deviations with an individual level random intercept. *A priori*, model Q3 represents the complete set of hypotheses an analyst might expect to consider for these data.

Each smooth in the quail models used $k = 9$ basis functions after application of identifiability constraints, except the quail-specific smooths, $f^*_{\text{egg}(i)}(\texttt{day}_i)$, which used $k = 6$ basis functions per individual. This reduced number of basis functions per smooth was a reflection that after accounting for the average shape of the growth curves, plus sex and treatment deviations, any remaining individual-specific variation from these other effects would be smaller in magnitude and less complex (wiggly).

*Smoothness selection & inference*

Restricted marginal likelihood (REML) smoothness selection (Wood, 2011) was used to estimate each of the GAMs fitted to the example data sets. REML was used because it has slightly better performance in terms of estimating smoothing parameters compared to marginal likelihood (ML) smoothness selection. We did not use GCV smoothness selection for these examples because i) prediction error is a less important consideration here where we are interested in estimation of effects and statistical inference, and ii) GCV is prone to undersmoothing (Reiss and Ogden, 2009; Wood, 2011).

In the lactation curve example I use standard model appraisal teachniques to identify the most useful model. In the pig growth example, AIC was used to identify which of the decompositions of time provided the best fit, but the specific choice of model was made on other grounds, using domain knowledge.

In the quail hormone example, AIC values are reported, but *a priori* I chose to report results from model Q3, the "full" model from the point of view of the potential hypotheses under consideration; as the treatment effect was allowed to vary between sexes, I take an estimation-based approach and quantify the magnitudes of any treatment by sex interactions using the most complex model. The remaining models are fitted and reported

on briefly for illustrative purposes; performing model term selection (for example, testing the higher order $f_{\text{treat}(i)}(\texttt{day}_i)$ interaction, and deciding on the basis of a *p* value or AIC whether to retain it in the model or not), would introduce biases into the inference process. As we currently lack good post selection inference methods for handling this kind of selection, it is prudent to simply not enter into such selection procedures.

Using model Q3 allows there to be treatment differences between male and female quail; subsequent estimation of values of interest and comparison among these estimates is instead the inference route followed for the quail example. To illustrate, I estimate two quantities of interest: i) the estimated growth rate (slope) of an average quail at $\texttt{day} = 20$ in all combinations of treatment and sex, and ii) the estimated weight of an average quail at the end of the experiment ($\texttt{day} = 78$), again for all combaination of treatment and sex. All pairwise comparisons among treatment levels within sex were performed, with adjustment of $p$ values to control the false discovery rate using the Benjamini–Yekutieli (Benjamini and Yekutieli, 2001) procedure. Estimates and pairwise comparisons were computed using the `slopes` and `predictions()` functions of the *marginaleffects* package for R (version 0.27.0, Arel-Bundock et al., 2024). Estimates of the expected growth curves for average quail in all combinations of treatment and sex were produced using the `conditional_values()` function in the R package *gratia* (version 0.10.0.9018, Simpson, 2024).

All figures were produced using the R packages *gratia* and *ggplot2* (version 3.5.2, Wickham, 2016).

## Results

### *Lactation curves*

Figure 3a shows the three lactation curves estimated using Wood's model, the Tweeide GLM, and the Tweedie GAM. While all three models capture the general shape of the lactation data, the Tweedie GLM and, to a lesser extent, Wood's model, overestimate the

peak yield, with the data exhibiting a broader period of peak fat content. Wood's modela and the Tweedie GLM also fail to capture the features of the mid–late lactation decline in fat content, other than the decline itself. Conversely, the GAM, as anticipated, estimates a lactation curve that more faithfully tracks the observed data. The model response residuals ($y_i - \hat{y}_i$) for Wood's model (Figure 3b) and the Tweedie GLM (Figure 3c) show a significant amount of unmodelled signal, while the response residuals for the Tweedie GAM (Figure 3d) are much smaller and do not show a residual pattern. Despite using roughly twice as many degrees of freedom as the other models (Table 1), the Tweedie GAM was clearly favoured in terms of AIC and root mean squared error of the fitted values.

*Pig growth*

The estimated growth curves for the 18 pigs in the pig growth example are shown in Figure 4, which were produced by model P2. Of the four models fitted, the two models that decomposed the growth effects into an average curve plus animal specific deviations from the average curve, models P2 and P4 resulted in the most parsimonious fits from the point of view of AIC Table 2. This is due to these two model forms using many fewer degrees of freedom (~64) than either models P1 (EDF = 83.54) or P3 (EDF = 90.57). Models P1 and P3 do not include the average curve and as a result expend many more degrees of freedom modelling the same general shape for each animal. Interestingly, for these data at least, allowing for a different smoothing parameter (P2) for each animals' deviation smooth seems to be preferred over using a single smoothing parameter (P4). In part, this is due to the somewhat idiosyncratic nature of each animal's growth curve.

Although the deviance explained is very high (~96%), much of this is due to use of random effects modelling differences between animals, and should not be taken as a sign that the model can effectively perfectly predict the weight of a pig under similar conditions; it is clear from Figure 4 that there is much unmodelled variation aorund the estimated growth curves. One feature of the data that I do not address here is the clear

variation among animals in the variance of the depth camera-based weight measurements; animals 5, 6, and 9 in particular, exhibit substantially greater variation than the other animals in the data set.

Table 3 provides an overview of the two terms in model P2. Although the average curve ($f(\texttt{day}_i)$) was allowed to use $k = 9$ basis functions, the wiggliness penalty has shrunk this back to 7.332 effective degrees of freedom (EDF). The animal-specific deviation smooths, $f_{\texttt{a}(i)}(\texttt{day}_i)$, were fully penalised and as such used $k = 10$ basis functions per animal. Here we clearly see the effect of the wiggliness penalty, which ahs resulted in a reduction from a potential 180 EDF for the set of animal-specific smooths to 56.054 EDF. The test of the null hypothesis for the average growth curve and the omnibus test for the pig-specific deviation smooths indicate both are statistically interesting.

*Quail hormone experiment*

As assessed by AIC, models containing treatment effects resulted in no improvement in the model fit (Table 4), a result that is consistent with the findings of Sarraude et al. (2020a). However, the results of model Q3 are reported below as this model is consistent with the *a priori* assumption of treatment effects, possibly varying by sex, that underlay the original experiment. Using AIC to perform model selection would invalidate subsequent statistical inference we might wish to conduct on the fitted model. Figure 5 shows the original data and the estimated growth curves for each individual quail that were obtained from Model Q3. The need for quail-specific random smooths is clear; model Q4 had the same model structure as that of Q3, except for the replacement of the quail-specific random smooths with quail-specific intercepts, which resulted in an increase in AIC of over 500 units (Table 4).

To focus on the estimated treatment effects, model Q3 was evaluated at 100 evenly-spaced values over the time covariate, the number of evaluattion points chosen to obtain a visually smooth representation of the estimated functions, plus all combinations of treatment level and sex. The effects of the quail specific random smooths and the mother

random effects were excluded from these estimates. This results in estimated treatment effects for the average quial, which, strictly speaking should not be interpretted as population level effects due to the non-identity link function. The resulting estimated effects are shown in Figure 6. While there are clear differences in the growth curves based on sex, there appears to be little qualitative difference in the treatment levels on quail growth.

The statistical summary of model Q3 is shown in Table 5. The omnibus tests of the treatment specific deviations, and the treatment by sex deviations, from the average curve, both have $p > 0.05$, which further reinforces the previously made observations that the effects of exposure to maternal hormone, if any, are small relative to the uncertainty in the model itself. The $p$ value for the omnibus test of the quail-specific deviation smooths ($f_{\text{egg}(i)}(\texttt{day}_i)$) is also greater than 0.05. This is somewhat surprising, given the magnitude of the difference in AIC that is observed when these quail-specific random smooths are replace by a simple random intercept term ($\Delta\text{AIC} = 525.548$). Despite their lacking "statistical significance", removing these terms on the basis of $p$ values would lead to invalid inference.

The largest contribution to the overall EDF of this model (EDF = 274.43) is from the quail-specific deviation smooths (EDF = 221.359), although as there are 57 individual animals and the random smooths include random intercepts to acocunt for differences between the average weight of each quail, this is unsurprising. Again, we see the effect of the wiggliness penalty, which has penalised the smooths back to EDF = 221.359 from a maximum of EDF of 342.

To formally assess these differences, pairwise comparisons of treatment levels within sex were conducted for the slope of the average growth curve at day 20. The estimated differences in the slopes of the growth curves are shown in Table 6. For all comparisons, despite some observed differences in the growth rates at day 20 among the treatment levels, the 95% credible intervals included zero for all comparisons. Therefore, if these

results reflect the broader population of quail, we should expect that any differences in growth rate due to maternal hormones are small, and largely indistinguishable from zero.

Finally, pairwise comparisons of quail weight at the end of the experiment (day 78) among treatment levels within sex were used to examine differences in estimated weight of quail due to maternal hormones. The results of these comparisons are shown in Table 7. The largest observed difference (~10.1 g) in weight of an average quail was between the $T_4$ hormone treatment and controls in female quail. This represents approximately a 5% decrease in the weight of an average female quail when exposed to the $T_4$ hormone compared with the untreated controls. Yet given the uncertainty in the estimates of the model coefficients, the 95% credible interval includes 0 for this, and all other, comparisons.

## Author' Points of View

The results presented above, clearly demonstrate the utility of GAMs for modelling the kinds of data commonly encountered in research involving animals. GAMs are often viewed unfavourably as being subjective, requiring the user to specify how complex they want the estimated smooth functions to be, data hungry, and more of a data visualisation tool ill-suited to formal statistical inference. These views, while perhaps valid in the case of the subjectivity critique, are largely out-dated with modern GAMs as I have presented above. The subjectivity critique has been addressed through developments in automated smoothness selection. None of the data sets analysed here are especially large, certainly not given today's standards, and much research has been done (e.g., Li and Wood, 2020) and continues to be done to adapt the algorithms to ever higher dimensional problems. The `bam()` function in the *mgcv* package, for example can handle data on the order of millions of rows, with tens of thousands of model parameters, given sufficient availablity of computer memory.

As demonstrated in the quail maternal hormone example, formal statistical inference

is entirely feasible. If we ignore the selection of smoothness parameters, the models described here are little more than generalized linear (mixed) models (GL(M)Ms) once the basis expansions of covariates have been performed. Modern software tools like the *marginaleffects* package for R allow a consistent interface to statistical inference across many disparate statistical models, and GAMs are no different in this regard.

The advantage of GAMs over GL(M)Ms is their ability to learn the shapes of non-linear relationships between covariates and the repsonse from the data themselves. This relieves the analyst from having to force data to fit a particular theoretical model, unless they have a good justification to use that model of course. GAMs also avoid the model selection problems inherent to modelling with polynomial basis expansions (e.g. $x + x^2 + x^3 + \cdots + x^p$).

The main disadvantage of GAMs is that they are more complex to fit than GL(M)Ms; the analyst has some additional data modelling choices to make when using GAMs. The main choice is the number of basis functions, $k$, that should be used by each smooth in the model. The general advice here is for the analyst to imagine the largest amount of wiggliness that they would expect and then set $k$ a little larger than this. Of course, the novice user of GAMs will lack the experience required to do this easily, but for the sorts of data problems exemplified above, we would not expect highly complex smooth functions, and the default of $k$ = 10 for univariate smooths in *mgcv* is usually sufficient for many problems. The key requirement is that the initial number of basis functions used should be large enough such that the span of functions representable with that basis will contain the true but unknown function or a close approximation to it. The basis dimension must be checked of course, and this adds another step to the model appraisal or checking procedure. With *mgcv* for example, the `k.check()` function provides a test for sufficiency of the basis size used to fit the model (Pya and Wood, 2016).

A further disadvantage is that statistical inference is somewhat more approximate that with GL(M)Ms. While GAMs share with GL(M)Ms the property of having asymptotically

correct *p* values, the *p* values for smooths are more approximate than for terms in a GL(M)M because the current tests do not account for the selection of smoothing parameters; for the purposes of the tests, the smoothing parameters are treated as being fixed and known, but instead they are estimated from the data. While there has been some progress in adapting the theory to include this additional source of uncertainty (e.g., Wood et al., 2016), as yet this has not been used to correct the *p* values of tests of smooths.

Finally, GAMs can require more substantial amounts of computing resources to fit than GL(M)Ms once data sets get above ten thousand observations or where models include several or complex random effect terms, using the algorithms provided by *mgcv*. The main requirement is computer memory, although the `bam()` function can help with this using algorithmic improvements from Li and Wood (2020).

GAMs are a very broad and general class of models. In the case of the pig growth data, several animals exhibited visibly more variation in their weight measurements than the majority of the animals in the data. Such heteroscedasticity (non-constant variance) can be modelled using distributional GAMs (or Generalized additive models for location, scale, shape or GAMLSS) (e.g. Rigby and Stasinopoulos, 2005; Kneib et al., 2021; Klein, 2024), which include linear predictors for all of the parameters of a distribution. In the examples above, the additional parameters (scale in the case of the Gamma, or the Tweedie power parameter) were estimated as constants for the entire data set. A distributional GAM would allow those parameters to potentially vary with individual animals, or as as smooth functions of the covariates, just as was done for the mean of the distribution in this study.

In conclusion, GAMs are a modern, flexible, and highly usable statistical model that is amenable to many research problems in animal science, and deserve a place in the statistical toolbox.

## Ethics approval

Not applicable. See the original sources of the data used for ethical approvement.

## Declaration of Generative AI and AI-assisted technologies in the writing process

The author did not use any artifical intelligence technologies in the writing process.

## Author ORCIDs

**Gavin L. Simpson**: 0000-0002-9084-8413

## Author contributions

GLS: Conceptualization, Methodology, Software, Formal analysis, Writing, Visulaization.

## Declaration of interest

None.

## Acknowledgements

The author would to like to express their appreciation to colleagues at the Department of Animal and Veterinary Sciences, Aarhus University for making the pig growth data available, and in particular to Dr. Mona Larsen for supplying the data and arranging with their coauthors to enable the subset analysed in this manuscript to be made available as open data.

## Financial support statement

This work was supported by an Aarhus Universitets Forskningsfond (Aarhus University Research Foundation; AUFF) starting grant awarded to the author.

## References


Arel-Bundock, V., Greifer, N., Heiss, A., 2024. How to interpret statistical models using marginaleffects for *R* and *Python*. J. Stat. Softw. 111, 1–32. URL: https://www.jstatsoft.org/index.php/jss/article/view/v111i09, doi:10.18637/jss.v111.i09.

Benjamini, Y., Yekutieli, D., 2001. The control of the false discovery rate in multiple testing under dependency. Ann. Stat. 29, 1165–1188. URL: http://projecteuclid.org/download/pdf_1/euclid.aos/1013699998.

Bus, J.D., Franchi, G.A., Boumans, I.J.M.M., Te Beest, D.E., Webb, L.E., Jensen, M.B., Pedersen, L.J., Bokkers, E.A.M., 2025. Short-term associations between ambient ammonia concentrations and growing-finishing pig performance and health. Prev. Vet. Med. 242, 106555. URL: http://dx.doi.org/10.1016/j.prevetmed.2025.106555, doi:10.1016/j.prevetmed.2025.106555.

Franchi, G.A., Bus, J.D., Boumans, I.J.M.M., Bokkers, E.A.M., Jensen, M.B., Pedersen, L.J., 2023. Estimating body weight in conventional growing pigs using a depth camera. Smart Agric. Technol. 3, 100117. URL: http://dx.doi.org/10.1016/j.atech.2022.100117, doi:10.1016/j.atech.2022.100117.

Henderson, H.V., McCulloch, C., 1990. Transform or link. URL: https://ecommons.cornell.edu/bitstream/1813/31620/1/BU-1049-MA.pdf.

Klein, N., 2024. Distributional regression for data analysis. Annu. Rev. Stat. Appl. 11. URL: https://www.annualreviews.org/content/journals/10.1146/annurev-statistics-040722-053607, doi:10.1146/annurev-statistics-040722-053607.

Kneib, T., Silbersdorff, A., Säfken, B., 2021. Rage against the mean – a review of distributional regression approaches. Econometrics and Statistics URL: https://www.sciencedirect.com/science/article/pii/S2452306221000824, doi:10.1016/j.ecosta.2021.07.006.

Li, Z., Wood, S.N., 2020. Faster model matrix crossproducts for large generalized linear models with discretized covariates. Stat. Comput. 30, 19–25. URL: https://doi.org/10.1007/s11222-019-09864-2, doi:10.1007/s11222-019-09864-2.

van Lingen, H.J., Fadel, J.G., Kebreab, E., Bannink, A., Dijkstra, J., van Gastelen, S., 2023. Smoothing spline assessment of the accuracy of enteric hydrogen and methane production measurements from dairy cattle using various sampling schemes. J. Dairy Sci. 106, 6834–6848. URL: http://dx.doi.org/10.3168/jds.2022-23207, doi:10.3168/jds.2022-23207.

Miller, D.L., 2025. Bayesian views of generalized additive modelling. Methods Ecol. Evol. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.14498, doi:10.1111/2041-210x.14498.

Pedersen, E.J., Miller, D.L., Simpson, G.L., Ross, N., 2019. Hierarchical generalized additive models in ecology: an introduction with mgcv. PeerJ 7, e6876. URL: http://dx.doi.org/10.7717/peerj.6876, doi:10.7717/peerj.6876.

Pya, N., Wood, S.N., 2016. A note on basis dimension selection in generalized additive modelling. arXiv [stat.ME] URL: http://arxiv.org/abs/1602.06696, arXiv:1602.06696.

R Core Team, 2025. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Reiss, P.T., Ogden, R.T., 2009. Smoothing parameter selection for a class of semiparametric linear models. J. R. Stat. Soc. Series B Stat. Methodol. 71, 505–523. URL: http://doi.wiley.com/10.1111/j.1467-9868.2008.00695.x, doi:10.1111/j.1467-9868.2008.00695.x.

Rigby, R.A., Stasinopoulos, D.M., 2005. Generalized additive models for location, scale and shape. J. R. Stat. Soc. Ser. C Appl. Stat. 54, 507–554. URL: http://dx.doi.org/10.1111/j.1467-9876.2005.00510.x, doi:10.1111/j.1467-9876.2005.00510.x.

Sarraude, T., Hsu, B.Y., Groothuis, T., Ruuskanen, S., 2020a. Dataset of prenatal thyroid hormones manipulation in japanese quails. URL: https://zenodo.org/record/3741711, doi:10.5281/zenodo.3741711.

Sarraude, T., Hsu, B.Y., Groothuis, T., Ruuskanen, S., 2020b. Testing the short-and long-term effects of elevated prenatal exposure to different forms of thyroid hormones. PeerJ 8, e10175. URL: http://dx.doi.org/10.7717/peerj.10175, doi:10.7717/peerj.10175.

Simpson, G.L., 2024. gratia: An R package for exploring generalized additive models. J. Open Source Softw. 9, 6962. URL: https://joss.theoj.org/papers/10.21105/joss.06962, doi:10.21105/joss.06962.

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Use R!, Springer International Publishing. URL: https://link.springer.com/book/10.1007/978-3-319-24277-4, doi:10.1007/978-3-319-24277-4.

Wood, P.D.P., 1967. Algebraic model of the lactation curve in cattle. Nature 216, 164–165. URL: http://dx.doi.org/10.1038/216164a0, doi:10.1038/216164a0.

Wood, S.N., 2003. Thin plate regression splines. J. R. Stat. Soc. Series B Stat. Methodol. 65, 95–114. URL: http://dx.doi.org/10.1111/1467-9868.00374, doi:10.1111/1467-9868.00374.

Wood, S.N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semi-parametric generalized linear models. J. R. Stat. Soc. Series B Stat. Methodol. 73, 3–36. URL: http://dx.doi.org/10.1111/j.1467-9868.2010.00749.x, doi:10.1111/j.1467-9868.2010.00749.x.

Wood, S.N., 2025. mgcv: Mixed GAM computation vehicle with automatic smoothness estimation. URL: http://dx.doi.org/10.32614/cran.package.mgcv, doi:10.32614/cran.package.mgcv.

Wood, S.N., Pya, N., Säfken, B., 2016. Smoothing parameter and model selection for general smooth models. J. Am. Stat. Assoc. 111, 1548–1563. URL: https://doi.org/10.1080/01621459.2016.1180986, doi:10.1080/01621459.2016.1180986, arXiv:http://dx.doi.org/10.1080/01621459.2016.1180986. doi: 10.1080/01621459.2016.1180986.

Table 1: Comparison of models fitted to the lactation curve data set. Degrees of freedom and effective degrees of freedom are shown for the Wood model and Tweedie GLM, and for the Tweedie GAM, respectively, showing the complexity in terms of (effective) numbers of parameters in each model. Akaike's An information criterion (AIC) values are shown for each model, alongside an estimate of the root mean squared error of the model fit estimated from the response residuals of each model.

| Model | (E)DF | AIC | RMSE |
|---|---|---|---|
| Wood | 4 | -100.61 | 0.30 |
| Tweedie GLM | 5 | -82.76 | 0.32 |
| Tweedie GAM | 8.004 | -132.97 | 0.15 |

Table 2: Comparison of models fitted to the pig weight data set. The model label is shown (see text for the specific model formulations). Effective degrees of freedom represents model complexity in terms of the (effective) number of parameters in each model. Akaike's An information criterion (AIC) values, model deviance, and deviance explained as a proportion are also reported.

| Model | EDF | AIC | Deviance | Deviance expl. |
|---|---|---|---|---|
| P1 | 83.542 | 3838.987 | 2.074 | 0.964 |
| P2 | 64.386 | 3797.407 | 2.088 | 0.964 |
| P3 | 90.570 | 3823.751 | 2.050 | 0.964 |
| P4 | 64.754 | 3809.395 | 2.148 | 0.963 |

Table 3: Model summary for model P2 fitted to the pig growth data set. The model term for comparison with the descriptions in the text and the label reported by the software are both shown for clarity. $k$ is the number of basis functions *per smooth*, EDF is the effective degrees of freedom, a measure of the complexity of each term, F is the test statistic and $p$ the *p* value of the null hypothesis of a flat constant function or functions.

| Model term | Label | $k$ | EDF | F | $p$ |
|---|---|---|---|---|---|
| $f(\texttt{day})$ | s(day) | 9 | 7.332 | 1302.500 | <0.001 |
| $f_{a(i)}(\texttt{day})$ | s(day,animal) | 10 | 56.054 | 10.033 | <0.001 |

Table 4: Comparison of models fitted to the quail hormone experimental data set. The model label is shown (see text for the specific model formulations). Effective degrees of freedom represents model complexity in terms of the (effective) number of parameters in each model. Akaike's An information criterion (AIC) values, model deviance, and deviance explained as a proportion are also reported.

| Model | EDF | AIC | Deviance | Deviance expl. |
|---|---|---|---|---|
| Q1 | 275.028 | 4458.053 | 12.229 | 0.998 |
| Q2 | 274.839 | 4458.922 | 12.360 | 0.998 |
| Q3 | 274.430 | 4461.062 | 12.455 | 0.998 |
| Q4 | 77.115 | 4986.609 | 30.416 | 0.995 |

Table 5: Model summary for model Q3 fitted to the quail hormone experimental data set. The model term for comparison with the descriptions in the text and the label reported by the software are both shown for clarity. $k$ is the number of basis functions *per smooth*, EDF is the effective degrees of freedom, a measure of the complexity of each term, F is the test statistic and $p$ the *p* value of the null hypothesis of a flat constant function or functions.

| Model term | Label | $k$ | EDF | F | $p$ |
|---|---|---|---|---|---|
| $f(\texttt{day})$ | s(day) | 9 | 8.947 | 4785.756 | <0.001 |
| $f_{\text{treat(i)}}(\texttt{day}_i)$ | s(day,treat) | 10 | 8.307 | 1.501 | 0.1467 |
| $f_{\text{sex(i)}}(\texttt{day}_i)$ | s(day,sex) | 10 | 7.322 | 14.393 | <0.001 |
| $f_{\text{treat}(i),\text{sex}(i)}(\texttt{day}_i)$ | s(day,treat,sex) | 10 | 9.807 | 1.596 | 0.0984 |
| $f_{\text{egg}(i)}(\texttt{day}_i)$ | s(day,egg) | 6 | 221.359 | 67.155 | 0.2666 |
| $\xi_{\text{mother}(i)}$ | s(mother) | 21 | 17.688 | 9.055 | <0.001 |

Table 6: Pairwise comparisons of treatment effects by sex on the slopes of the growth curves at day 20 for an average quail of the indicated sex in the pair of treatments shown. The Hypothesis column lists the specific pairwise comparison, Diff. is the estimated difference in the slope of the growth curves compared, SE its standard error. Z is the Wald test statistic, $p$ its *p* value, and associated endpoints of a bayesian 95% credible interval on the estimated difference of slopes.

| Sex | Hypothesis | Diff. | SE | Z | $p$ | 2.5% | 97.5% |
|---|---|---|---|---|---|---|---|
| Female | $T_3 -$ Control | -0.339 | 0.290 | -1.168 | 1 | -1.148 | 0.471 |
| Female | $T_3$ $T_4 -$ Control | -0.395 | 0.298 | -1.326 | 1 | -1.227 | 0.437 |
| Female | $T_4 -$ Control | -0.495 | 0.356 | -1.389 | 1 | -1.490 | 0.500 |
| Female | $T_3$ $T_4 - T_3$ | -0.056 | 0.218 | -0.259 | 1 | -0.666 | 0.553 |
| Female | $T_4 - T_3$ | -0.156 | 0.231 | -0.676 | 1 | -0.802 | 0.490 |
| Female | $T_4 - T_3$ $T_4$ | -0.100 | 0.275 | -0.364 | 1 | -0.866 | 0.667 |
| Male | $T_3 -$ Control | -0.194 | 0.299 | -0.649 | 1 | -1.031 | 0.642 |
| Male | $T_3$ $T_4 -$ Control | -0.251 | 0.263 | -0.957 | 1 | -0.984 | 0.482 |
| Male | $T_4 -$ Control | -0.388 | 0.339 | -1.144 | 1 | -1.335 | 0.559 |
| Male | $T_3$ $T_4 - T_3$ | -0.057 | 0.262 | -0.217 | 1 | -0.788 | 0.674 |
| Male | $T_4 - T_3$ | -0.194 | 0.273 | -0.708 | 1 | -0.957 | 0.570 |
| Male | $T_4 - T_3$ $T_4$ | -0.137 | 0.294 | -0.465 | 1 | -0.957 | 0.684 |

Table 7: Pairwise comparisons of treatment effects by sex on the estimated mean of the quail growth curves at day 78 for an average quail of the indicated sex in each pair of treatments. The Hypothesis column lists the specific pairwise comparison, Diff. is the estimated difference of mean body mass (g) of the growth curves compared, SE its standard error. Z is the Wald test statistic, $p$ its *p* value, and associated endpoints of a bayesian 95% credible interval on the estimated difference of means.

| Sex | Hypothesis | Diff. | SE | Z | $p$ | 2.5% | 97.5% |
|---|---|---|---|---|---|---|---|
| Female | $T_3 - \text{Control}$ | -6.332 | 5.838 | -1.085 | 1 | -22.682 | 10.017 |
| Female | $T_3T_4 - \text{Control}$ | -5.577 | 5.614 | -0.993 | 1 | -21.298 | 10.144 |
| Female | $T_4 - \text{Control}$ | -10.087 | 5.762 | -1.751 | 1 | -26.224 | 6.050 |
| Female | $T_3T_4 - T_3$ | 0.756 | 4.344 | 0.174 | 1 | -11.411 | 12.923 |
| Female | $T_4 - T_3$ | -3.755 | 4.038 | -0.930 | 1 | -15.064 | 7.555 |
| Female | $T_4 - T_3T_4$ | -4.510 | 3.648 | -1.236 | 1 | -14.727 | 5.707 |
| Male | $T_3 - \text{Control}$ | 2.493 | 5.391 | 0.462 | 1 | -12.605 | 17.592 |
| Male | $T_3T_4 - \text{Control}$ | -2.992 | 4.035 | -0.742 | 1 | -14.292 | 8.308 |
| Male | $T_4 - \text{Control}$ | -0.808 | 4.579 | -0.176 | 1 | -13.632 | 12.017 |
| Male | $T_3T_4 - T_3$ | -5.485 | 4.811 | -1.140 | 1 | -18.960 | 7.990 |
| Male | $T_4 - T_3$ | -3.301 | 4.709 | -0.701 | 1 | -16.489 | 9.887 |
| Male | $T_4 - T_3T_4$ | 2.184 | 3.740 | 0.584 | 1 | -8.291 | 12.659 |

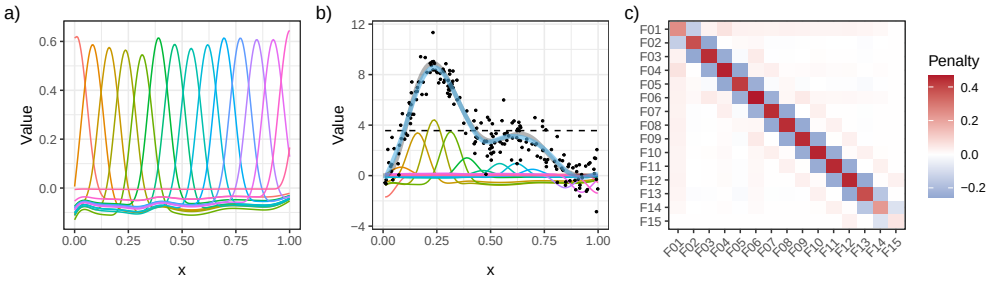


Fig. 1: Illustration of how penalised splines work. A spline basis expansion (a) and associated penalty matrix $\mathbf{S}$ (c) are formed for a covariate $x$. Model fitting involves finding estimates for the coefficients of the basis functions that make the fitted spline (thick, blue curve) go as close to the data (black points) as possible, without over fitting (b). In (a) and (b) the basis functions are shown as thin coloured lines and are from a B spline basis. The sum-to-zero identifiability constraint needed so that an intercept can be included in the model has been absorbed into the basis shown. The dashed horizontal line in (b) is the estimated value of the intercept. The penalty matrix (c) encodes how wiggly each basis function is in terms of its second derivative.

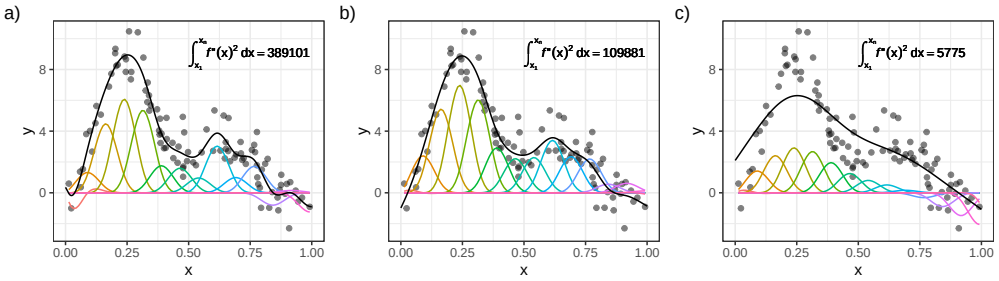


Fig. 2: Illustration of how the wiggliness penalty controls the resulting fit of a penalised spline. The weighted basis functions are shown as thin coloured lines. In each panel a penalised spline is shown by the solid black line, which has been fitted to the data points shown. The wiggliness value of the spline, the integrated squared derivative of the fitted spline over $x$ is given in the upper right of each panel. The spline in (a) is over fitted to the data, resulting in a very wiggly function with a large wiggliness value. The spline in (c) is over smoothed, resulting in a a simple fitted function with low wiggliness, but which does not fit the data well. The spline in (b) represents a balance between fit to the data and complexity of fitted function. The smoothing parameter for the spline, $\lambda$, is used as a tuning parameter in the model, which ultimately controls this balance between fit and complexity.

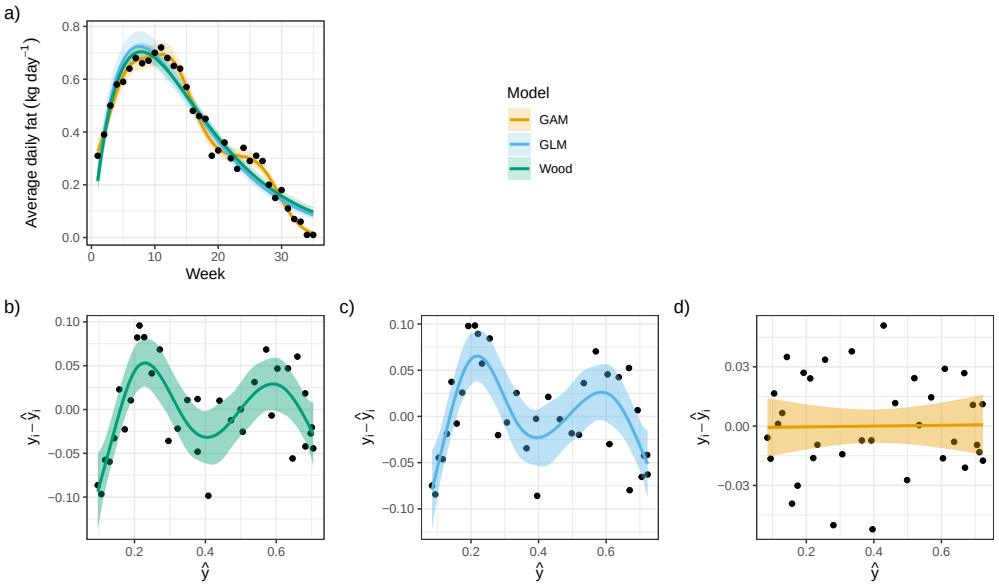


Fig. 3: Results of model fitting to the average daily fat content data from Henderson and McCulloch (1990). a) observed average daily fat content (points) and estimated lactation curves from Wood's (1967) model, a Tweedie GLM, and a Tweedie GAM (lines) with associated 95% confidence (Wood's model) or 95% credible intervals (GLM and GAM). Response residuals for Wood's model (b), Tweedie GLM (c), and Tweedie GAM (d), plus scatter plot smoothers (lines) and 95% credible intervals (shaded ribbons).

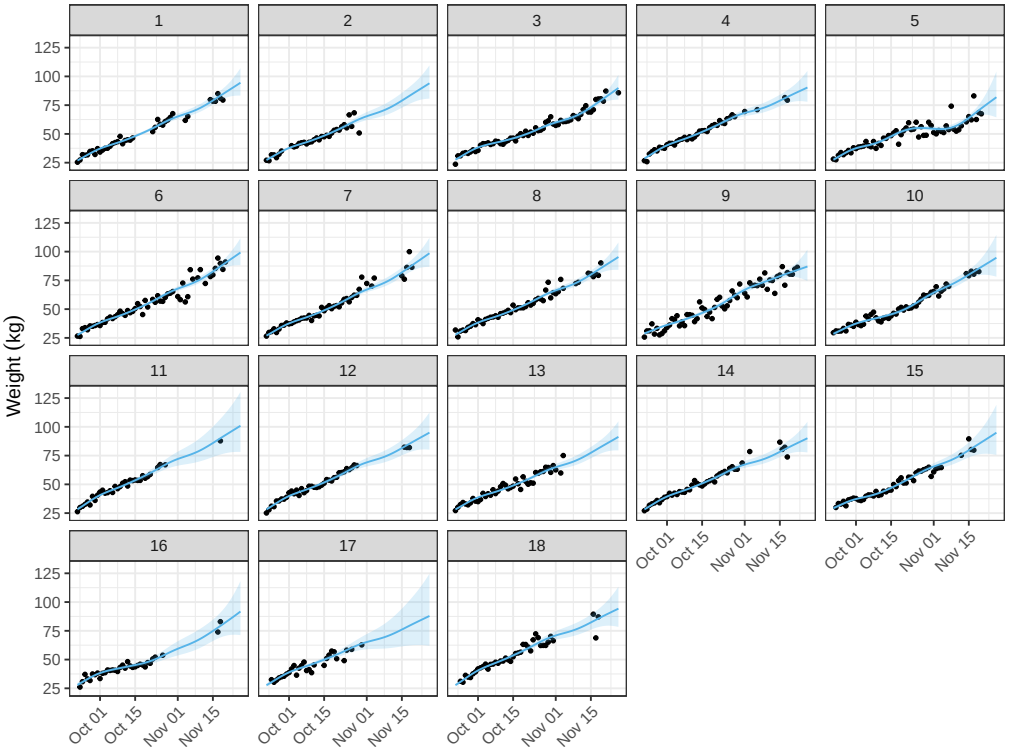


Fig. 4: Depth camera-based weight estimates from 18 commerical pigs. The data (black points) are the average of multiple measurements taken of each animal per day, 1 panel per pig. The panel labels indicate to the pig shown. The estimated growth curve for each pig obtained using generalized additive model P2 is shown by the blue line in each panel. The blue shaded ribbon is the 95% bayesian credible interval around the estimate curve.

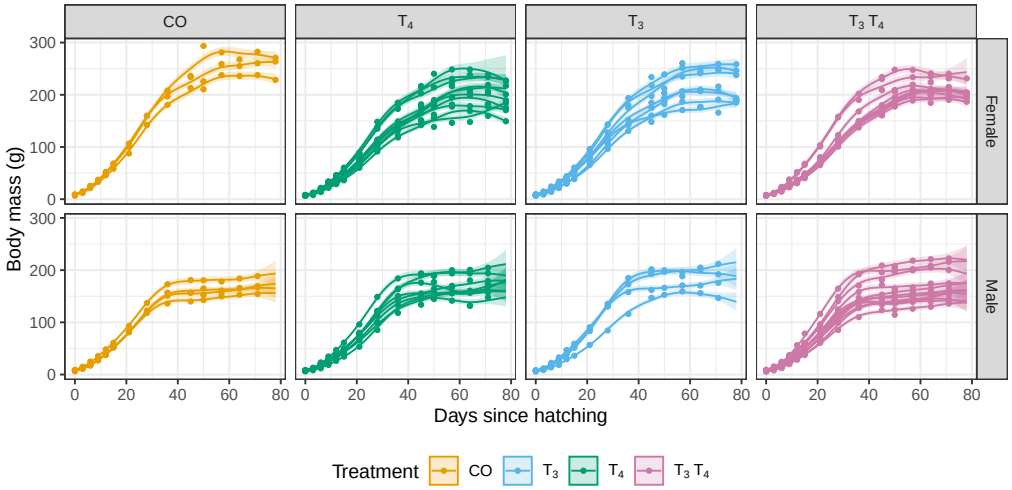


Fig. 5: Measured body mass (g) for 57 Japanese quail from an experiment on the effects of exposure to the maternal hormone $T_4$, its active metabolite $T_3$, both ($T_3T_4$) or a saline solution control (CO). The data (points) are shown along side the estimated growth curve for each quail obtained using generalized additive model Q3, which are shown by the coloured lines in each panel. The coloured shaded ribbon is the 95% bayesian credible interval around the estimate curve. The data are faceted by treatment and the sex of bird.

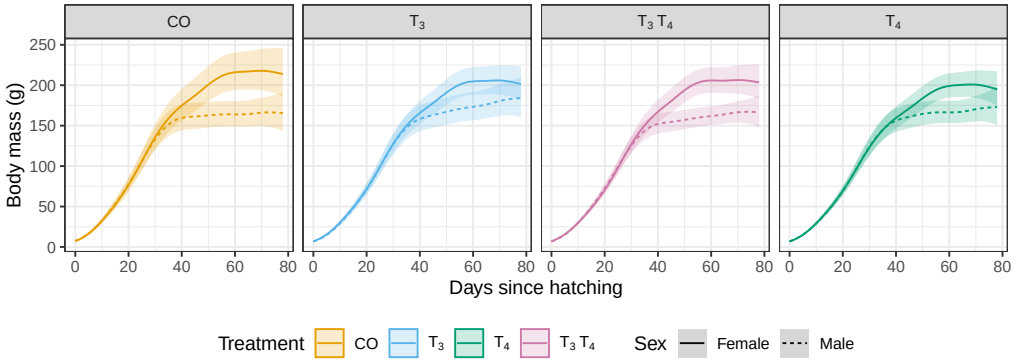


Fig. 6: Conditional value plots showing the growth rate for an average quail in each of the treatment groups by sex. The panels show the estimated curves for a particular treatment group; saline solution controls (CO), maternal hormone metabolite $T_3$, the maternal hormone $T_4$, and a combination of both $T_3T_4$. The estimated curve for male birds is shown by the solid line, and females the dashed line in each panel. The shaded band around each curve is a 95% bayesian credible interval.

## Figure captions

Fig. 1: Illustration of how penalised splines work. A spline basis expansion (a) and associated penalty matrix $\mathbf{S}$ (c) are formed for a covariate $x$. Model fitting involves finding estimates for the coefficients of the basis functions that make the fitted spline (thick, blue curve) go as close to the data (black points) as possible, without over fitting (b). In (a) and (b) the basis functions are shown as thin coloured lines and are from a B spline basis. The sum-to-zero identifiability constraint needed so that an intercept can be included in the model has been absorbed into the basis shown. The dashed horizontal line in (b) is the estimated value of the intercept. The penalty matrix (c) encodes how wiggly each basis function is in terms of its second derivative.

Fig. 2: Illustration of how the wiggliness penalty controls the resulting fit of a penalised spline. The weighted basis functions are shown as thin coloured lines. In each panel a penalised spline is shown by the solid black line, which has been fitted to the data points shown. The wiggliness value of the spline, the integrated squared derivative of the fitted spline over $x$ is given in the upper right of each panel. The spline in (a) is over fitted to the data, resulting in a very wiggly function with a large wiggliness value. The spline in (c) is over smoothed, resulting in a a simple fitted function with low wiggliness, but which does not fit the data well. The spline in (b) represents a balance between fit to the data and complexity of fitted function. The smoothing parameter for the spline, $\lambda$, is used as a tuning parameter in the model, which ultimately controls this balance between fit and complexity.

Fig. 3: Results of model fitting to the average daily fat content data from Henderson and McCulloch (1990). a) observed average daily fat content (points) and estimated lactation curves from Wood's (1967) model, a Tweedie GLM, and a Tweedie GAM (lines) with associated 95% confidence (Wood's model) or 95% credible intervals (GLM and GAM). Response residuals for Wood's model (b), Tweedie GLM (c), and Tweedie GAM (d), plus scatter plot smoothers (lines) and 95% credible intervals (shaded ribbons).

Fig. 4: Depth camera-based weight estimates from 18 commerical pigs. The data (black points) are the average of multiple measurements taken of each animal per day, 1 panel per pig. The panel labels indicate to the pig shown. The estimated growth curve for each pig obtained using generalized additive model P2 is shown by the blue line in each panel. The blue shaded ribbon is the 95% bayesian credible interval around the estimate curve.

Fig. 5: Measured body mass (g) for 57 Japanese quail from an experiment on the effects of exposure to the maternal hormone $T_4$, its active metabolite $T_3$, both ($T_3T_4$) or a saline solution control (CO). The data (points) are shown along side the estimated growth curve for each quail obtained using generalized additive model Q3, which are shown by the coloured lines in each panel. The coloured shaded ribbon is the 95% bayesian credible interval around the estimate curve. The data are faceted by treatment and the sex of bird.

Fig. 6: Conditional value plots showing the growth rate for an average quail in each of the treatment groups by sex. The panels show the estimated curves for a particular treatment group; saline solution controls (CO), maternal hormone metabolite $T_3$, the maternal hormone $T_4$, and a combination of both $T_3T_4$. The estimated curve for male birds is shown by the solid line, and females the dashed line in each panel. The shaded band around each curve is a 95% bayesian credible interval.