Journal of Animal Ecology 2007
76, 761–770

Identifying when weather influences life-history traits of grazing herbivores

MICHELLE SIMS, DAVID A. ELSTON, ANN LARKHAM*, DANIEL H. NUSSEY† and STEVE D. ALBON*

*Biomathematics and Statistics Scotland, The Macaulay Institute, Craigiebuckler, Aberdeen, AB15 8QH, UK; *The Macaulay Institute, Craigiebuckler, Aberdeen AB15 8QH, UK; and †Large Animal Research Group, Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK*

Summary

1. There is increasing evidence that density-independent weather effects influence life-history traits and hence the dynamics of populations of animals. Here, we present a novel statistical approach to estimate when such influences are strongest. The method is demonstrated by analyses investigating the timing of the influence of weather on the birth weight of sheep and deer.

2. The statistical technique allowed for the pattern of temporal correlation in the weather data enabling the effects of weather in many fine-scale time intervals to be investigated simultaneously. Thus, while previous studies have typically considered weather averaged across a single broad time interval during pregnancy, our approach enabled examination simultaneously of the relationships with weekly and fortnightly averages throughout the whole of pregnancy.

3. We detected a positive effect of temperature on the birth weight of deer, which is strongest in late pregnancy (mid-March to mid-April), and a negative effect of rainfall on the birthweight of sheep, which is strongest during mid-pregnancy (late January to early February). The possible mechanisms underlying these weather–birth weight relationships are discussed.

4. This study enhances our insight into the pattern of the timing of influence of weather on early development. The method is of much more general application and could provide valuable insights in other areas of ecology in which sequences of intercorrelated explanatory variables have been collected in space or in time.

Key-words: birth weight, mixed model, multicollinearity, random coefficient, smoothing.

Journal of Animal Ecology (2007) **76**, 761–770
doi: 10.1111/j.1365-2656.2007.01251.x

Introduction

There is increasing evidence that life-history traits of mammals are influenced by environmental factors experienced early in life both pre- and post-natally (e.g. see Lindström 1999; Beckerman *et al*. 2002; Lummaa & Clutton-Brock 2002).

Pre-natally, this process has been called foetal programming (Lucus 1991); it can influence birth weight, which in turn can affect adult reproductive success (Albon, Clutton-Brock & Guinness 1987; Kruuk *et al*. 1999). Environmental factors that influence pre-natal development are therefore of importance when considering both evolutionary processes and population dynamics, because they are likely to influence both an individual's lifetime fitness and the age-specific vital rates of individuals from a given birth cohort (Albon, Clutton-Brock & Langvatn 1992). Lumma (2003) states that foetal programming occurs when a stimulus during a critical period of early development subsequently affects body structure, physiology or metabolism. This highlights the importance of the timing of a stimulus. It is this crucial temporal element that this paper aims to identify.

The situation we address is when a density-independent environmental factor, such as a weather variable, has been measured in a sequence of time intervals; for example, a week or month in pregnancy. We wish to fit a linear model in which the response variable is the

Correspondence: Michelle Sims, Duke University Marine Laboratory, Nicholas School of Environment and Earth Sciences, Beaufort, NC 28516, USA.
E-mail: m.sims@duke.edu

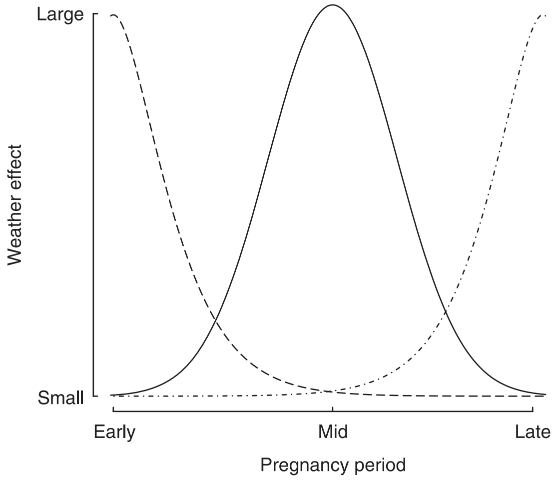

Fig. 1. Hypothetical relationships between weather and a life-history trait throughout the pregnancy period. The relationship with weather may be strongest during early (– –), mid (—) or late (– ·) pregnancy.

life-history trait and weather records at each time interval are the explanatory variables. As different stages of pregnancy are associated with different patterns of foetal development (Rhind, Rae & Brooks 2003), it is more likely that the effects of weather on the life-history trait are comparable during similar stages of pregnancy than at other times during pregnancy. Therefore, we can reasonably assume the relationship with weather changes smoothly across the pregnancy period. For example, the relationship between a life-history trait and weather may be strongest during mid-pregnancy and decline towards the beginning or end of pregnancy (Fig. 1). Standard techniques, such as an ordinary linear regression, could be used to investigate the timing of critical periods and we would expect the estimated regression coefficients to show a similar smooth pattern.

The estimated regression coefficients are interpretable when the explanatory variables are uncorrelated or nearly so. However, given the temporal structure of the data (seasonality) we expect weather conditions to be similar in neighbouring time intervals and the strength of this correlation to decline with increasing separation between intervals. Multicollinearity is encountered when there are a large number of correlated explanatory variables in a regression analysis. If this problem is ignored, uncertainty in the regression coefficient estimates will be inflated relative to if each covariate had been fitted alone and detection of time periods when weather is important is less likely. Multiple regression is therefore inappropriate and alternative procedures should be sought. One solution is to reduce the number of explanatory variables in the model perhaps by summarizing weather across a few broad temporal intervals. For example, some studies have found correlations between local weather conditions and life-history traits (see Albon *et al.* 1987; Gaillard, Delorme & Jullien 1993; Festa-Bianchet, Jorgenson & Reale 2000). Typically, these authors considered just one period of pregnancy, using intervals of 1 or 2 months, often within the last trimester when the foetus is growing most rapidly. However, recent analysis of the long-term study of red deer on the Isle of Rum, using a longer run of years, found birth weight was better correlated with temperature in a 3-month interval from February to April (Coulson *et al.* 2003) than an April to May interval used previously (Albon, Guinness & Clutton-Brock 1983; Albon *et al.* 1987; Kruuk *et al.* 1999). In practice, there has been little effort to develop methods to evaluate the influence of weather using finer-scale temporal intervals. Yet such methods would highlight the periods when the relationship between foetal development and weather is strongest and hence suggest the possible mechanisms underlying observed cohort effects. However, this requires examination of many intervals and hence having a large number of explanatory variables in the regression model.

Regression methods have been developed that will both (1) lessen the effects of multicollinearity while keeping a large number of explanatory variables in the model, and (2) use the information about the structure of regression coefficients by smoothing the estimated coefficients towards a solution expected given their known interrelationship (Eilers 1991). The procedures are based on ridge regression (Hoerl & Kennard 1970), a standard method for dealing with multicollinearity. An unknown parameter controls the amount of smoothing and 'optimal' values are chosen using cross-validation. Here, we adopt their general concepts and fit them into a linear mixed model framework. We show that using our method the value of the smoothing parameter is automatically selected within the mixed model itself, thus regularizing the procedure without requiring the use of classical selection methods such as cross-validation.

We will demonstrate the usefulness of our method in two examples investigating the timing of the influence of weather conditions prevailing during gestation on the birth weight of red deer *Cervus elaphus* (L.) and North Country Cheviot sheep *Ovis aries* (L.) in Scotland. As described above, birth weight in red deer appears to be associated with daily air temperature sometime between February and May. In contrast, rainfall appears to be the important variable, both earlier in pregnancy (8–14 weeks) and later in pregnancy (15–21 weeks), influencing the birth weight of sheep (Larkham, personal observations). Thus, previous studies identified weather effects within broad temporal intervals. The application of our method will enable examination of multiple fine-scale temporal intervals throughout the whole of the pregnancy period thus expanding our understanding of when the relationship with weather is strongest.

We explore the performance of our approach against two commonplace statistical methods for fitting regression models with many explanatory variables. This involves comparing how useful the methods are at obtaining good estimates of the regression coefficients in the presence of multicollinearity. However, a comparison is only possible when we know the true values; therefore, simulated data are used, rather than field data, as the true regressions are then known. We begin by introducing these different statistical methods for model fitting. The aim here is to illustrate the development of our approach in order to fully appreciate the methodology described in this paper. We show from the simulation study that our method, which we shall call difference penalty regression (DPR), provides a robust technique for estimating the effects of intercorrelated explanatory variables. In particular, we are able to get a better understanding of the role weather plays in the early development of mammals.

Although we develop the method in the context of traits affecting population demography, the applicability of the methods is very general. Another application in which the covariates would correspond to time intervals is phenology, with attention focusing on the effect of weather (often temperature) on the timing of an event. Alternatively, the covariates may correspond to spatial intervals. Examples in this case might be in studies of the effects of surrounding habitats on breeding success at a sample of nest sites, or the effect of surrounding water pollution on the growth of a marine sedentary organism. The paper concludes with a discussion of the observed weather–birth weight relationship and a general evaluation of the developed method.

Methods

BACKGROUND TO METHOD OF ANALYSIS

Regression using least squares

We begin with a linear regression situation relating the values of a response variable y from n independent observations to k continuous explanatory variables, w. That is, for the i^{th} observation

$$y_i = \mu + \beta_1 w_{i,1} + \beta_2 w_{i,2} + \ldots + \beta_k w_{i,k} + \varepsilon_i \qquad \text{eqn 1}$$

where is μ the intercept, β_j, $j = 1, \ldots, k$ are k regression coefficients and $\varepsilon_i \sim N(0,\sigma^2)$. Different methods, giving rise to different estimators, exist for estimating the βs. Their usefulness in obtaining accurate estimates can be explored by looking at summary measures of their sampling distributions.

A widely used criterion for comparing estimators is to look at their mean square error, which is defined as

$$mse(\hat{\beta}_j) = b(\hat{\beta}_j)^2 + Var(\hat{\beta}_j) \qquad \text{eqn 2}$$

where $\hat{\beta}_j$ is an estimator of an unknown regression coefficient β_j, $b(\hat{\beta}_j) = E(\hat{\beta}_j) - \beta_j$ is the bias of the estimator and $Var(\hat{\beta}_j)$ is the variance of the estimator. Both the mean, $E(\hat{\beta}_j)$, and the variance, $Var(\hat{\beta}_j)$, of the sampling distribution are involved in the assessment of an estimator's performance. A good estimator would be one with a small mean square error, essentially one that has little bias and a small variance.

Least squares regression (LSR) is the most common estimator of the coefficients in a linear regression model as it has the lowest variance of all unbiased ($E(\hat{\beta}_j) = \beta_j$) estimators. Estimates of the regression coefficients in eqn 1 are obtained by minimizing the sum of squared residuals

$$\sum^{n} (y_i - (\mu + \beta_1 w_{i,1} + \beta_2 w_{i,2} + \ldots + \beta_k w_{i,k}))^2. \qquad \text{eqn 3}$$

However, in certain situations, such as in the case of having intercorrelated explanatory variables, there may be better estimators whereby the mean square error criterion may be optimized by relaxing the constraint of unbiasedness. In the presence of collinearity, least squares estimates are unstable and tend to be large in absolute value. Consequently, estimates may be very different from their true values and similar estimates are unlikely to be obtained if the data collecting process were repeated.

Figure 2(a) illustrates these effects with data generated from eqn 1 with $n = 100$, $k = 17$, $\sigma = 4$; correlation between neighbouring explanatory variables, that is between $w_{i,j}$ and $w_{i,j+1}$, was set at 0·90 and $\beta_1, \ldots, \beta_k$ had a unimodal structure, peaking in the middle. Two features are clear: the estimated coefficients may be very different from the true values and successive estimates can err in opposite directions. To accommodate for such erratic behaviour, the standard errors of the estimates and confidence intervals are large. The unimodal structure of the true coefficients is therefore obscured in the presence of multicollinearity. Thus, the ability of a regression analysis to study the effects of a set of explanatory variables will be less effective.

Penalized regression using least squares

To control for the instability and large values associated with LSR estimates Hoerl & Kennard (1970) proposed the use of ridge regression (RR), a form of penalized least squares, in the presence of multicollinearity. The ridge regression estimator minimizes the sum of squared residuals plus a penalty term

$$\sum_{i=1}^{n} (y_i - (\mu + \beta_1 w_{i,1} + \beta_2 w_{i,2} + \ldots + \beta_k w_{i,k}))^2 + \lambda \sum_{j=1}^{k} \beta_j^2. \qquad \text{eqn 4}$$

RR discourages large values of the coefficients, one of the consequences of multicollinearity, by having the second term in eqn 4, $\lambda\sum_{j=1}^{k} \beta_j^2$, which penalizes large estimates. The extent of this shrinkage is controlled by the parameter λ. Methods for selecting values for λ include a ridge trace and cross-validation. Larger values for λ tend to shrink each $\hat{\beta}_j$ further towards zero, while when $\lambda = 0$ eqn 4 reduces to the LSR estimator (eqn 3).

Journal of Animal Ecology, **76**, 761–770

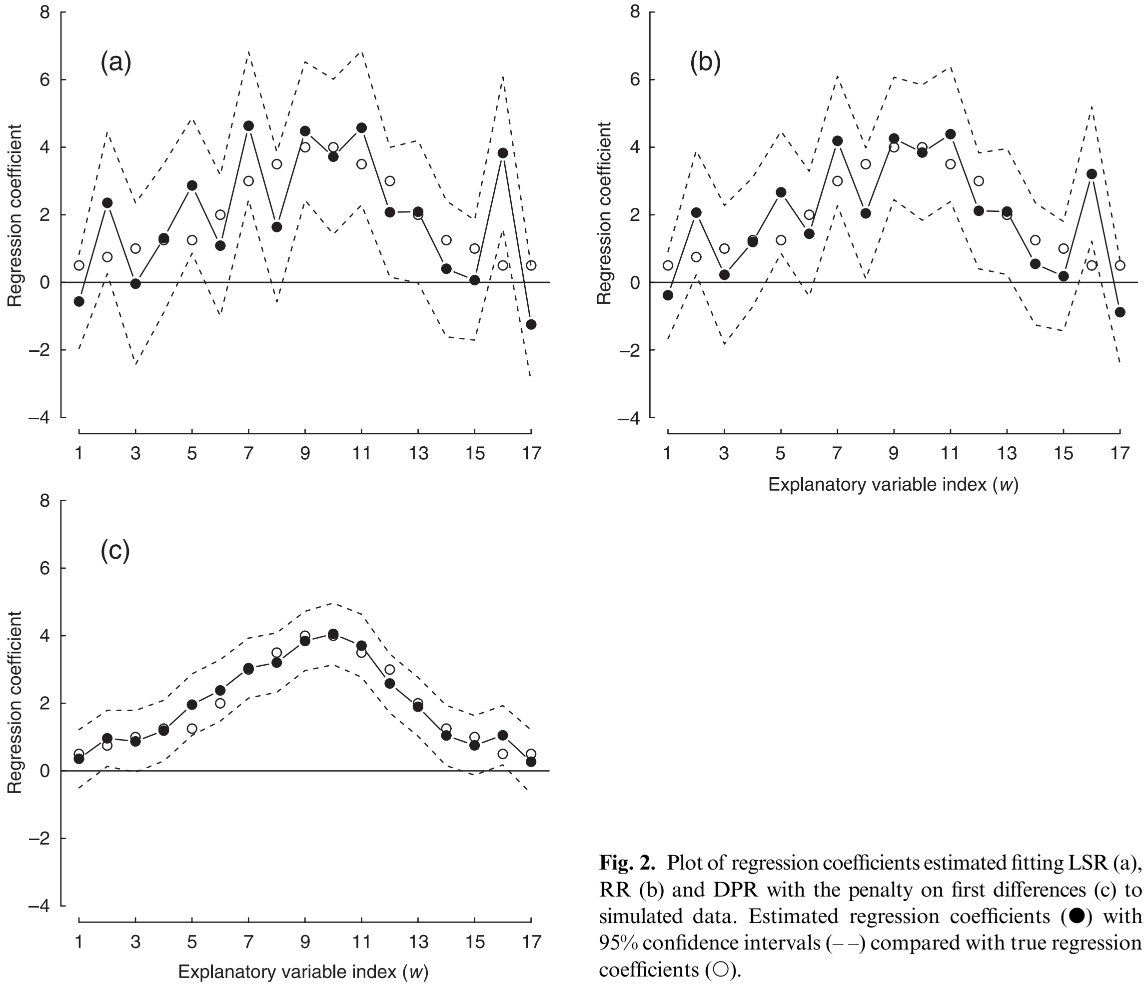

Fig. 2. Plot of regression coefficients estimated fitting LSR (a), RR (b) and DPR with the penalty on first differences (c) to simulated data. Estimated regression coefficients (●) with 95% confidence intervals (– –) compared with true regression coefficients (○).

RR can reduce the mean square error of the estimator when multicollinearity is present. More stable regression coefficient estimates may be produced with smaller standard errors than those estimated using LSR, but at the cost of generating biased estimates (a bias towards zero). However, with a suitable choice of λ this bias/variance trade-off can result in a smaller mean square error. Comparable results can be obtained using a Bayesian approach by supposing that one's prior knowledge of the coefficients is that they are drawn from a common distribution. Lindley & Smith (1972) assumed that the coefficients had the following prior

$$\beta_j \sim N(0, \sigma_\beta^2) \qquad \text{eqn 5}$$

and showed this to be equivalent to a ridge regression with $\lambda = \sigma^2/\sigma_\beta^2$. The Bayesian method has the advantage, in contrast to the rather subjective methods of choosing λ, of automatically selecting λ within the regression analysis.

RR assumes no particular interrelationship between the $\hat{\beta}_j$s, whereas those collected on a temporal scale would be expected to show a distinct pattern of correlation. In reference to our example data sets, the effect of weather in neighbouring time intervals during pregnancy is likely to be similar. For example, suppose the explanatory variables in eqn 1 represent weather conditions for a sequence of k time intervals and their associated regression coefficients represent the relationship between weather and the response variable. One would not expect the irregular changes in the regression coefficients estimated using LSR (Fig. 2a), but would rather believe the relationship changes smoothly between time intervals.

Instead of shrinking all $\hat{\beta}_j$s towards zero, Eilers (1991) proposed using a penalty term in which each $\hat{\beta}_j$ is shrunk towards adjacent estimates. This has the effect of smoothing the estimated regression coefficients. The penalty could be on first differences $\beta_j - \beta_{j-1}$ (later called differences of order 1). Hence, the penalty term in eqn 4 becomes $\lambda \sum_{j=2}^{k} (\Delta\beta_j)^2$ where $\Delta\beta_j = \beta_j - \beta_{j-1}$ for $j = 2, \ldots, k$ and the $\hat{\beta}_j$s are forced towards the values of their neighbours. Further smoothing, using this difference penalty regression (DPR), can be obtained by taking second differences in the penalty (differences of order 2); that is $\lambda \sum_{j=3}^{k} (\Delta^2\beta_j)^2$
where

$$\Delta^2\beta_j = (\beta_j - \beta_{j-1}) - (\beta_{j-1} - \beta_{j-2}) = \beta_j - 2\beta_{j-1} + \beta_{j-2}$$

for $j = 3, \ldots, k$.

The usefulness of difference penalties is illustrated by the recent developments in nonparametric curve fitting. P-spline smoothers (Eilers & Marx 1996; Marx & Eilers 1998) use difference penalties in a very general setting: the computational difficulties associated with knot selection are avoided in our examples because the covariates represent equally spaced time intervals. One area for improvement in all smoothing methods is optimizing the size of the smoothing parameter. Values are generally chosen using classical selectors such as cross-validation. However, this does not always give a satisfactory result [see comments in Eiler & Marx (1996) on the use of cross-validation to select the smoothing parameter in a P-spline to optimize smoothing].

An alternative procedure, which will self-regulate the amount of shrinkage, is to consider eqn 1 as a linear mixed model and treat the effect of weather as a random effect. The $\beta_j s$ are then assumed to be drawn from a normal distribution with zero mean and unknown variance. Random effect estimates are obtained by weighting information from the data with this prior assumption. Consequently, the estimates are shrunk towards zero, like in a ridge regression, but now the amount of shrinkage is controlled within the analysis itself. The idea of smoothing using linear mixed models has been demonstrated with the fitting of cubic smoothing splines together with fixed and random effects (Verbyla *et al*. 1999; see also Wood 2006 for a general presentation of smoothing splines in a mixed model context), and we carry the potential of this approach through to smoothing series of regression coefficients.

Our method is comparable with RR and the Bayesian approach in that it is also a biased estimator. Similar to the Bayesian approach, it assumes the $\beta_j s$ follow a common distribution and therefore has the advantage of letting the data estimate the shrinkage parameter. Hence, the amount of smoothing is dependent on the data for a particular order of differencing. Next we describe linear mixed models in more detail and show how difference penalties can be easily incorporated as random effects into the linear mixed model.

Penalized regression using linear mixed models

Eqn 1 is equivalent to a fixed effects model and may be re-written in matrix form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon} \qquad \text{eqn 6}$$

where $\mathbf{y}$ is a vector of length n, $\mathbf{X}$ is a $n \times (k + 1)$ matrix of intercorrelated explanatory variables with elements in row i equal to $[1, w_{i,1}, \ldots, w_{i,k}]$, $\boldsymbol{\theta} = (\mu, \beta_1, \ldots, \beta_k)'$ is a vector of fixed effect parameters corresponding to the unknown intercept and regression coefficients and $\boldsymbol{\varepsilon}$ is a $n \times 1$ vector of errors assumed to be independent and normally distributed with mean zero and variance $\sigma^2\mathbf{I}$. The RR estimator and its equivalent Bayesian approach estimate $\boldsymbol{\theta}$ as

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1}\,\mathbf{X}'\mathbf{y} \qquad \text{eqn 7}$$

Alternatively we can fit a linear mixed model with the regression on weather as a random effect. Thus

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon} \qquad \text{eqn 8}$$

where $\mathbf{Z}$ is a $n \times k$ matrix of the intercorrelated explanatory variables with elements $w_{i,j}$, $\mathbf{b} = (b_1, \ldots, b_k)'$ is a vector of random effect parameters, $\mathbf{X}$ is a $n \times p$ matrix of fixed effects values, $\boldsymbol{\alpha}$ is a $p \times 1$ vector of fixed effects parameters, $\mathbf{b} \sim N(\mathbf{0}, \sigma_\beta^2\,\mathbf{I})$ and $\boldsymbol{\varepsilon} \sim N(0,\ \sigma^2\,\mathbf{I})$. When the fixed effect part of the model consists of just the intercept, $\mathbf{X}$ becomes a $n \times 1$ matrix of 1s and $\boldsymbol{\alpha}$ is the scalar μ. Additional random effects can be fitted but here we consider only the correlated explanatory variables. The best linear unbiased predictor of $\mathbf{b}$ is defined as

$$\hat{\mathbf{b}} = (\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{I})^{-1}\,\mathbf{Z}'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\alpha}}) \qquad \text{eqn 9}$$

where $\lambda = \sigma^2/\sigma_\beta^2$ Note, λ in eqn 9 is equivalent to the Bayesian choice in eqn 7. Estimates of the relationships with weather, $\beta_1, \ldots \beta_k$ (eqn 1), are given by the elements of $\hat{\mathbf{b}}$.

DIFFERENCE PENALTY REGRESSION

Suppose we want to treat differences of the regression coefficients as random effects. If $\mathbf{D}$ is a difference matrix of order d then let $\mathbf{b}^* = \mathbf{D}\mathbf{b}$ represent the vector of random effect parameters. Different difference penalties may be considered: by including rows of zeros below the active part, we can treat $\mathbf{D}$ as a $k \times k$ square matrix and avoid complexity in the expression of its generalized inverse, denoted $\mathbf{D}^-$. Assuming the intercorrelated explanatory variables are recorded for intervals of equal length, a first difference matrix, which penalizes first differences, has values $D_{j,j} = -1$, $D_{j,j+1} = 1$ for $j = 1, \ldots, k - 1$ and zero elsewhere. Thus

$$b_{j+1} - b_j \sim N(0,\ \sigma_\beta^2) \qquad \text{eqn 10}$$

for $j = 1, \ldots, k - 1$. Alternatively a second difference matrix has values $D_{j,j} = 1$, $D_{j,j+1} = -2$, $D_{j,j+2} = 1$ for $j = 1, \ldots, k - 2$ and zero elsewhere. Since $\mathbf{Z}\mathbf{b} = \mathbf{Z}(\mathbf{D}^-\mathbf{D}\mathbf{b} + \mathbf{b} - \mathbf{D}^-\mathbf{D}\mathbf{b})$, we may rewrite eqn 8 as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}^*\mathbf{b}^* + \boldsymbol{\varepsilon} \qquad \text{eqn 11}$$

where $\mathbf{Z}^* = \mathbf{Z}\mathbf{D}^-$ is a transformation of the correlated explanatory variables, $\mathbf{b}^* \sim N(\mathbf{0}, \sigma_\beta^2\mathbf{I})$, and $\mathbf{Z}(\mathbf{b} - \mathbf{D}^-\mathbf{D}\mathbf{b})$ gets put into the fixed effect term $\mathbf{X}\boldsymbol{\alpha}$. Further details are given below. (Note that, for simplicity, the exposition contains only those terms required for regression on the weather covariate.) The model in eqn 11 characterizes a DPR.

The possible amount of shrinkage on the random effect parameters varies between none at all, effectively assuming independence between the explanatory

Journal of Animal Ecology, **76**, 761–770

variables, and total shrinkage so that $\mathbf{b}^* = \mathbf{0}$. Total shrinkage when using a first differences penalty would force all b_j to have the same value. A fixed effect representing this value should be included in the model so that if there was total shrinkage we would obtain the expected result. In eqn 11, $\mathbf{X}$ has columns $\mathbf{x}_p$, $p = 1, 2$, with the i^{th} elements of $\mathbf{x}_1$, $\mathbf{x}_2$ equal to 1, $\sum_{j=1}^{k} w_{i,j}$ respectively and $\boldsymbol{\alpha} = (\mu, \alpha_1)'$, where μ is the intercept and α_1 can be interpreted as an average regression coefficient about which the smoothing values vary. We then find that $\mathbf{X}\boldsymbol{\alpha} = \mu\mathbf{1} + \mathbf{Z}(\mathbf{b} - \mathbf{D}^{-}\mathbf{D}\mathbf{b})$.

If the penalty were to be on second differences then total shrinkage would force the regression coefficients to follow a linear trend along the sequence of explanatory variables. A first order polynomial of the coefficients should then be included as another fixed effect. $\mathbf{X}$ would then be a $n \times 3$ matrix with the third column equal to $\sum_{j=1}^{k} jw_{i,j}$ and $\boldsymbol{\alpha} = (\mu, \alpha_1, \alpha_2)'$ thereby allowing the underlying average regression coefficient, about which the smoothed coefficients vary, to change linearly with time interval. Similarly, for higher orders of differencing additional fixed effect terms need to be added representing higher order polynomials. Hence, for the d^{th} order of differencing there will be $d + 1$ fixed effect terms in the model. See Appendix S1 in the Supplementary material for details of the derivation of the estimates and standard errors of the weather regression coefficients $\beta_1, \ldots, \beta_k$.

The linear mixed models were fitted using the method of residual maximum likelihood (REML) (Patterson & Thompson 1971; Searle, Casella & McCulloch 1992). All subsequent examples of RR and DPR were estimated using the REML procedure in Genstat 7·2. Using the simulated data described above, a comparison of the three methods, LSR, RR and DPR is displayed in Fig. 2. The regression coefficients are much smoother using the latter approach and the unimodal structure of the true coefficients is now clear. The simulated data and the Genstat code used to fit DPR, with a penalty on first differences, to the simulated data are given in Appendices S2 and S3 of the Supplementary material. It is also possible to fit the DPR model using the lme function in R and the example code is given in Appendix S4 of the Supplementary material.

FIELD DATA

We illustrate the application of DPR with data on birthweights of two species of mammals: North Country Cheviot sheep and red deer.

Birth weights (kg) of twin lambs from North Country Cheviot sheep were collected as part of the Hill Sheep Systems Research programme (Sibbald & Maxwell 1992). We consider twin lambs born over a 10-year period between 1977 and 1986 at Sourhope Research Station (55°6′N, 2°42′W) in south-east Scotland. As there were instances where only one lamb per twin had been recorded, for computational simplicity we considered only the weight of one lamb per twin. Daily rainfall data (mm) were obtained from the Kelso Meteorological Office weather station that is situated approximately 18 km south from the study area.

Pregnancy generally lasts 21 weeks from late November to late April and so we included 21 explanatory variables in the linear mixed model representing average daily rainfall for each week during pregnancy. Weather records were individual specific, documenting rainfall at each of the 21 weeks prior to an individual's birth date. As gestation length varies little, we regard β_j as representing the relationship with rainfall during the j^{th} week. Hence, β_1 represents the effect of rainfall during the first week of pregnancy; β_{21}, the last. This contrasts with other studies that use calendar weeks and give all individuals born in the same year identical weather records irrespective of birth date. We considered penalties on first and second differences of the rainfall regression coefficients at neighbouring time intervals. Also included in the model were a continuous fixed effect for birth date, a categorical fixed effect for mother's age at the time of birth of her lamb, a continuous fixed effect for mother's pre-mating body weight, a continuous fixed effect for lamb's year of birth, a categorical fixed effect for sex of the lamb along with random effects for mother and lamb's year of birth (Larkham, personal observations).

The birth weight data for red deer were collected from a wild population living in the North Block study area on the Isle of Rum in Scotland (57°0′N, 6°20′W) between 1971 and 1998 (Clutton-Brock, Guinness & Albon 1982). Maximum daily air temperatures were obtained from a Meteorological Office weather station at Kinloch on the Isle of Rum. Pregnancy length is approximately 34 weeks with average mating in late October and subsequent calving in early June. Splitting the pregnancy period into 2-week intervals, we had 17 explanatory variables specifying mean daily temperature for each fortnight throughout pregnancy. As with the previous example, weather records were individual-specific. Some years had missing weather data. For this analysis we included only those pregnancies with complete weather records. We put a penalty on first and second differences of the temperature effects at neighbouring time intervals. Also included in the model were a continuous fixed effect for birth date, categorical fixed effects for sex of the calf and mother's reproductive status, mother's age as a linear and quadratic fixed effect along with random effects for mother identity and calf's year of birth. For further description of the data and the variables used here see Coulson *et al.* (2003).

The same model was also fitted but with mean temperature across a broad time window included as a fixed effect instead of the 17 weather covariates. Two time intervals were considered, April–May and February–April, both used in previous studies on the population, and we informally compared the estimated pattern of the weather effects during pregnancy using the three models. Comparison of the weather effects in the different models is not straightforward because

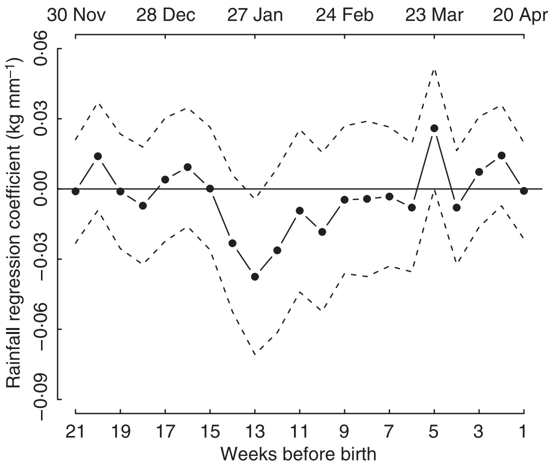

Fig. 3. Illustration of the influence of rainfall at 21 successive weeks during pregnancy on the birth weight of North Country Cheviot sheep using DPR with a penalty on first differences. Approximate dates at the mid-point of the intervals are given on the top axis. Estimated regression coefficients (●) displayed with 95% confidence intervals (– –).

the covariates were measured at different scales. A correction for this was made by dividing the estimated regression coefficient for the April–May and February–April temperature covariates by the number of fortnights within these time intervals. To investigate the extent to which the three models explained the variation in birth weight due to temperature, we compared the magnitude of their corresponding variance components.

SIMULATED DATA

Results from simulated data are presented to compare the performance of DPR with RR using linear mixed models and LSR. Data were created from the model in eqn 1 with $n = 100$, $k = 17$, $\sigma = 4$, $\mu = 100$ and $\beta_1, \dots, \beta_{17}$ have a unimodal structure (see Appendix S5 in Supplementary material for values). Assuming $w_{i,1}, \dots, w_{i,17}$ were a sequence of evenly spaced continuous variables, values for $w_{i,1}, \dots, w_{i,17}$ were generated by drawing samples of size 100 from a multivariate normal distribution

$$N_k\,(\mathbf{0}, \boldsymbol{\Sigma})$$

where $\mathbf{0}$ is a vector of zeros of length k and $\boldsymbol{\Sigma}$ is a $k \times k$ first-order autoregressive covariance matrix such that

$$\Sigma = \sigma_c^2 \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{k-1} \\ \rho & 1 & \rho & \cdots & \rho^{k-2} \\ \rho^2 & & 1 & \cdots & \vdots \\ \vdots & \vdots & & \ddots & \rho \\ \rho^{k-1} & \rho^{k-2} & \cdots & \rho & 1 \end{pmatrix}$$

where $\sigma_c^2 = 2$ and $\rho = 0{\cdot}80$. Hence the correlation between successive explanatory variables was 0·80 and correlation declined exponentially with increasing separation between variables.

A total of 10 000 data sets were generated and estimates of the regression coefficients $\hat{\beta}_1, \dots, \hat{\beta}_{17}$ were obtained using LSR, RR and DPR (with a penalty on first differences) for each data set. The mean and variance of the coefficient estimates were calculated as well as the average of the estimated variances of each coefficient $\overline{\text{var}(\beta)}$. As the true values for the regression coefficients were known, the bias and mean square error of the estimates could be acquired. At each simulation, 95% confidence intervals were also estimated as $\hat{\beta}_j \pm 1{\cdot}96 s.e.(\hat{\beta}_j)$ and coverage rates were obtained; that is, the proportion of confidence intervals containing the true regression coefficient.

Results

Difference penalty regression (DPR) with a penalty on first differences revealed a significant negative effect of rainfall during mid-pregnancy on the birth weight of lambs (Fig. 3). Females showed a negative response of $-0{\cdot}03$ kg mm^{-1} at approximately 13 weeks before birth, which is on average between late January and early February. The negative effect is most distinct when the penalty is on second differences (Fig. S1 in Supplementary material). The estimated effects of rainfall were not significantly different from zero during early and late pregnancy (Fig. 3 shows that the confidence intervals of the estimates include zero). Despite attempts to smooth the regression coefficients, we still see an irregular estimate at 5 weeks before birth compared with its neighbouring values.

Fitting a penalty on the first differences of the regression coefficients reveals a positive relationship between temperature and the birth weight of red deer (Fig. 4). The individual relationships were not significantly different from zero until about 14 weeks before birth, then rise from about 0·03 kg °C^{-1} at 14 weeks before birth to 0·045 kg °C^{-1} at 10 weeks before birth but then declined again, with the regression coefficients in the last 6 weeks of gestation again being not significantly different from zero. Thus, on average it was temperature during March and April, particularly the last 2 weeks of March and the first 2 weeks of April, when the relationship was strongest. The estimated coefficients for the two broad time intervals (February–April and April–May) were similar to each other and to the estimate for the peak (early April) identified by DPR. An even smoother result is obtained with a second differences penalty on the regression coefficients (Fig. S2 in Supplementary material).

The addition of a temperature effect in the linear mixed model for birthweight of red deer decreased the magnitude of the variance component for calf's year of birth (Table 1). In the models with an April–May or February–April temperature covariate, the size of the variance component was reduced by 41% and 45%, respectively. The DPR model revealed the greatest reduction of 52%, i.e. temperature was explaining more of the between-year variation in birthweight. There was little difference in the estimated variance for the mother and residual effect between the models.

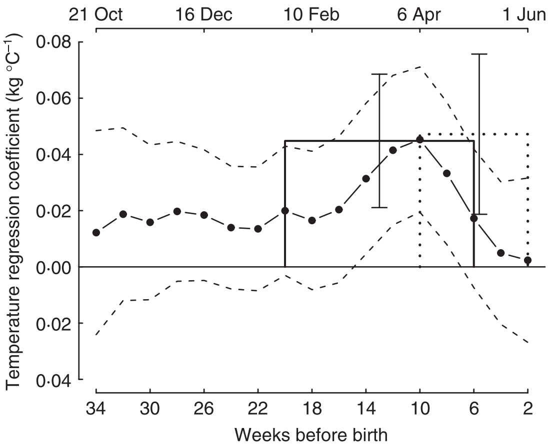

Fig. 4. Illustration of the influence of temperature at 17 successive 2-weekly intervals during pregnancy on the birth weight of red deer using DPR with a penalty on first differences. Approximate dates at the mid-point of the intervals are given on the top axis. Estimated regression coefficients (●) displayed with 95% confidence intervals (– –). Relationships with temperature summarized across broad time intervals also shown. Solid line: model with average February–April temperature as a covariate (± 1·96 SE). Dotted line: model with average April–May temperature as a covariate (± 1·96 SE).

Table 1. Variance components of linear mixed models fitted to the birth weights of red deer. Models were fitted with no temperature effect (none), with a fixed effect included for average April–May temperature (April–May) or average February–April temperature (Feb–April), and with temperature smoothed using DPR with a penalty on first differences (smoothing)

	Model for temperature effect			
Random effect	none	April–May	Feb–April	smoothing
Mother	0·637	0·630	0·633	0·623
Calf's year of birth	0·082	0·048	0·045	0·039
Residual	0·856	0·859	0·856	0·853

Table 2. Simulation results comparing least squares regression (LSR), ridge regression (RR) and difference penalty regression (DPR)

	Bias			Variance			Mean square error		
Parameter	LSR	RR	DPR	LSR	RR	DPR	LSR	RR	DPR
β_1	0·003	–0·002	0·044	0·204	0·178	0·075	0·204	0·201	0·077
β_5	0·004	0·021	0·162	0·366	0·306	0·063	0·366	0·126	0·089
β_{10}	0·007	–0·004	–0·053	0·390	0·316	0·057	0·390	0·200	0·060

	$\overline{\text{var}(\beta)}$			Coverage		
Parameter	LSR	RR	DPR	LSR	RR	DPR
β_1	0·209	0·192	0·106	0·950	0·956	0·978
β_5	0·365	0·329	0·118	0·948	0·956	0·980
β_{10}	0·395	0·350	0·116	0·950	0·958	0·994

See Appendix S5 for the complete table. Summary statistics computed from 10 000 simulations. Coverage refers to coverage of 95% confidence intervals. MSE is the mean square error. Results for difference penalty regression used a first differences penalty.

The simulation results show that the DPR performs better at obtaining estimates of regression coefficients in the presence of intercorrelated explanatory variables (Table 2; see Appendix S5 in Supplementary material for complete results). The LSR estimator appears to be unbiased but the variance is very large. In contrast, RR and DPR show an increase in bias but a substantial reduction in variance. Moreover, the mean square error is much reduced using DPR. Using DPR, we would expect $\overline{\text{var}(\beta)}$ to be an approximation of $\overline{\text{var}(\beta)}$ as is true for the LSR estimates. However, $\overline{\text{var}(\beta)}$ is consistently larger for all parameter estimates using either RR or DPR, the gap being more noticeable for the latter, arising because the variance components are estimated. This consequently increases the coverage of the 95% confidence intervals.

Discussion

In this paper we have demonstrated that a form of penalized multiple regression, with a penalty on the differences of neighbouring regression coefficients, identifies the pattern of the timing of influence of weather experienced during pregnancy on the birth weight of two species of mammals. Our approach is an improvement on previous analyses as it allows for the effect of weather at multiple, narrow time intervals to be estimated simultaneously. Hence, it provides a flexible method for identifying periods when a stimulus may have profound effects on early development and may help illuminate the potential mechanisms underlying weather–biological trait relationships. As birth weight is a major factor influencing lifetime fitness (Clutton-Brock *et al*. 1996; Kruuk *et al*. 1999) and because whole cohorts experience similar weather, the consequences of weather conditions experienced during pregnancy may have lagged effects on the dynamics of populations (Albon *et al*. 1992).

The method uses the additional information regarding the structure of the regression coefficients to reduce the negative consequences of multicollinearity. By fitting the penalty as a random effect in a linear mixed model, the amount of smoothing is regulated within the model itself. Shrinkage of the regression coefficients towards neighbouring coefficients increases when estimates from the data are less reliable. This may occur when the relative variability between the random effect and the residual error, σ^2_β/σ^2, decreases, as the size of the random effect may be partly due to the large residual error. Likewise shrinkage increases when the amount of information on each random effect decreases.

We assumed the explanatory variables are an ordered sequence at evenly spaced intervals. However, we can relax these assumptions to include variables that form a looped sequence, such as calendar month (Elston & Proe 1995) or wind direction (Eilers 1991), or are positioned at uneven intervals by modifying the differencing matrix **D**. Application of DPR depends on a suitable choice for the order of differencing. In some cases we might have prior knowledge about the pattern of the coefficients that will determine what level of differencing to penalize.

The simulation study shows that the DPR estimator is preferable to the LSR or RR when the explanatory variables form an ordered sequence; there was a noticeable decline in mean square error using DPR. However, the width of the confidence intervals calculated using DPR is overestimated resulting in higher coverage probabilities than expected. As the error is small we feel this should not hinder its applicability but rather recommend caution when interpreting results.

The main application of this paper was to identify the role weather plays during pregnancy in determining birth weight of mammals. In order to control for the effects of seasonality on birth weight, for example that individuals born later experience warmer weather and hence may be born heavier, we included birth date as a covariate in both models. Weather regression coefficients therefore represent the effects of departures from the seasonal average. The results from both examples have potentially important consequences because the variation in birth weight is a major determinant of lifetime reproductive success in large herbivores (Kruuk *et al*. 1999; Steinheim *et al*. 2002).

Our analyses confirm that temperature during late pregnancy was having a positive effect on birth weight in red deer, presumably because warm spring temperatures advanced the onset of grass growth and increased the mother's plane of nutrition (Albon *et al*. 1992). Previous studies have used April–May mean temperatures (Albon *et al*. 1983, 1987; Kruuk *et al*. 1999) or February–April mean temperatures (Coulson *et al*. 2003) in analyses of birthweight in the same population. Our results suggest these estimates encapsulated much of the time period when temperature most strongly influenced birthweight. However, we found from the smoothed regression coefficients that the time window when weather most strongly influenced birthweight was between mid-March and mid-April. The main difference between these modelling approaches was the estimated pattern of the timing of influence of temperature during pregnancy. The smoothed set of regression coefficients revealed using our approach is what we would expect given the biological background of the problem rather than effectively a step function that was produced when temperature was summarized across a single broad time window of multiple months.

Rainfall was most strongly influencing birth weight of North Country Cheviot sheep during mid-pregnancy, between the end of January and early February. A possible explanation for the negative effect might be that harsh environmental conditions, such as high rainfall, increase maternal thermoregulatory costs and may result in re-partitioning of nutrients from the foetus to the mother. This may reduce placental growth, which is at its maximum rate during mid-pregnancy (Ehrhardt & Bell 1995; Wallace *et al*. 2000). Reduced placental size can in turn reduce foetal growth rate (Mellor 1987), potentially resulting in reduced birth weight.

In addition to developing a better understanding of the timing of the impact of prevailing weather on life-history traits, penalized regression should increase the likelihood of detecting other factors contributing to the variation in fitness. For example, by explaining more of the variation due to density-independent weather factors, DPR may permit a better estimate of both the additive effects of density dependence and the often more obscure interactions between density-dependent and stochastic, density-independent factors (Milner, Elston & Albon 1999; Coulson *et al*. 2001; Hallett *et al*. 2004).

The linear mixed model approach provides an alternative method for fitting penalized regression, requires no additional smoothing parameter-selection method and alternative forms of penalty terms can be implemented by a simple transformation of the intercorrelated explanatory variables. In this paper we use our method to produce estimates of the critical periods during pregnancy when weather is influencing life-history traits of mammals. More generally, the procedure presented here provides a valuable technique for ecologists from a broad range of fields who wish to study the effects of sequences of intercorrelated explanatory variables.

Acknowledgements

The red deer data were collected from Rum, and we acknowledge the contribution of all involved in the collection process, including SNH for permission to work there, their staff for local support, Fiona Guinness, Ali Donald, Sean Morris, and many other field workers on the Kilmory deer project. We would also like to thank Tim Clutton-Brock for permission to use the red deer data, Simon Wood for advice on running the analysis in R, as well as Tim Coulson, Chris Glasbey, Xavier Lambin and Josephine Pemberton for helpful comments on earlier drafts of this paper. The work was supported by funds from the Scottish Executive Environment and Rural Affairs Department.

References

Albon, S.D., Clutton-Brock, T.H. & Guinness, F.E. (1987) Early development and population-dynamics in red deer. 2. Density-independent effects and cohort variation. *Journal of Animal Ecology*, **56**, 69–81.

Albon, S.D., Clutton-Brock, T.H. & Langvatn, R. (1992) Cohort variation in reproduction and survival: implications for population demography. *The Biology of Deer* (ed. R.D. Brown), pp. 15–21. Springer Verlag, Berlin.

Albon, S.D., Guinness, F.E. & Clutton-Brock, T.H. (1983) The influence of climatic variation on the birth weights of red deer (*Cervus-elaphus*). *Journal of Zoology*, **200**, 295–298.

Beckerman, A., Benton, T.G., Ranta, E., Kaitala, V. & Lundberg, P. (2002) Population dynamic consequences of delayed life-history effects. *Trends in Ecology and Evolution*, **17**, 263–269.

Clutton-Brock, T.H., Guinness, F.E. & Albon, S.D. (1982) *Red Deer: Behaviour and Ecology of Two Sexes*. University Chicago Press, Chicago, IL.

Clutton-Brock, T.H., Stevenson, I.R., Marrow, P., MacColl, A.D., Houston, A.I. & McNamara, J.M. (1996) Population fluctuations, reproductive costs and life-history tactics in female Soay sheep. *Journal of Animal Ecology*, **65**, 675–689.

Coulson, T., Catchpole, E.A., Albon, S.D., Morgan, B.J.T., Pemberton, J.M., Clutton-Brock, T.H., Crawley, M.J. & Grenfell, B.T. (2001) Age, sex, density, winter weather, and population crashes in Soay sheep. *Science*, **292**, 1528–1531.

Coulson, T., Kruuk, L.E.B., Tavecchia, G., Pemberton, J.M. & Clutton-Brock, T.H. (2003) Estimating selection on neonatal traits in red deer using elasticity path analysis. *Evolution*, **57**, 2879–2892.

Ehrhardt, R.A. & Bell, A.W. (1995) Growth and metabolism of the ovine placenta during mid-gestation. *Placenta*, **16**, 727–741.

Eilers, P.H.C. (1991) Penalised regression in action: estimating pollution roses from daily averages. *Environmetrics*, **2**, 25–47.

Eilers, P.H.C. & Marx, B.D. (1996) Flexible smoothing with B-splines and penalties. *Statistical Science*, **11**, 89–102.

Elston, D.A. & Proe, M.F. (1995) Smoothing regression coefficients in an overspecified regression model with interrelated explanatory variables. *Journal of the Royal Statistical Society Series C. (Applied Statistics)*, **44**, 395–406.

Festa-Bianchet, M., Jorgenson, J.T. & Reale, D. (2000) Early development, adult mass, and reproductive success in bighorn sheep. *Behavioral Ecology*, **11**, 633–639.

Gaillard, J.M., Delorme, D. & Jullien, J.M. (1993) Effects of cohort, sex, and birth date on body development of roe deer (*Capreolus capreolus*) fawns. *Oecologia*, **94**, 57–61.

Hallett, T.B., Coulson, T., Pilkington, J.G., Clutton-Brock, T.H., Pemberton, J.M. & Grenfell, B.T. (2004) Why large-scale climate indices seem to predict ecological processes better than local weather. *Nature*, **430**, 71–75.

Hoerl, A.E. & Kennard, R.W. (1970) Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, **12**, 55–67.

Kruuk, L.E.B., Clutton-Brock, T.H., Rose, K.E. & Guinness, F.E. (1999) Early determinants of lifetime reproductive success differ between the sexes in red deer. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **266**, 1655–1661.

Lindley, D.V. & Smith, A.F.M. (1972) Bayes estimates for the linear model. *Journal of the Royal Statistical Society. Series B*, **34**, 1–41.

Lindström, J. (1999) Early development and fitness in birds and mammals. *Trends in Ecology and Evolution*, **14**, 343–348.

Lucus, A. (1991) Programming by early nutrition in man. *The Childhood Environment and Adult Disease* (eds G.R. Bock & J. Whelan), pp. 38–55. John Wiley and Sons, Chichester.

Lummaa, V. (2003) Early developmental conditions and reproductive success in humans: downstream effects of prenatal famine, birthweight, and timing of birth. *American Journal of Human Biology*, **15**, 370–379.

Lummaa, V. & Clutton-Brock, T.H. (2002) Early development, survival and reproduction in humans. *Trends in Ecology and Evolution*, **17**, 141–147.

Marx, B.D. & Eilers, P.H.C. (1998) Direct generalized additive modeling with penalised likelihood. *Computational Statistics and Data Analysis*, **28**, 193–209.

Mellor, D.J. (1987) Nutritional effects on the fetus and mammary-gland during pregnancy. *Proceedings of the Nutrition Society*, **46**, 249–257.

Milner, J.M., Elston, D.A. & Albon, S.D. (1999) Estimating the contributions of population density and climatic fluctuations to interannual variation in survival of Soay sheep. *Journal of Animal Ecology*, **68**, 1235–1247.

Patterson, H.D. & Thompson, R. (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**, 545–554.

Rhind, S.M., Rae, M.T. & Brooks, A.N. (2003) Environmental influences on the fetus and neonate – timing, mechanisms of action and effects on subsequent adult function. *Domestic Animal Endocrinology*, **25**, 3–11.

Searle, S.S., Casella, G. & McCulloch, C.E. (1992) *Variance Components*. Wiley, New York.

Sibbald, A.R. & Maxwell, T.J. (1992) Contrasting approaches to systems research for sheep production in the hills and uplands of the United Kingdom. *Global Appraisal of Livestock Farming Systems and Study of their Organisational Levels: Concepts, Methodology and Results* (ed. A. Gibon & G. Matheron), pp. 79–93. Office for Official Publications of the European Communities, Luxembourg.

Steinheim, G., Mysterud, A., Holand, O., Bakken, M. & Adnoy, T. (2002) The effect of initial weight of the ewe on later reproductive effort in domestic sheep (*Ovis aries*). *Journal of Zoology*, **258**, 515–520.

Verbyla, A.P., Cullis, B.R., Kenward, M.G. & Welham, S.J. (1999) The analysis of designed experiments and longitudinal data by smoothing splines. *Journal of the Royal Statistical Society Series C. (Applied Statistics)*, **48**, 269–311.

Wallace, J.M., Bourke, D.A., Aitken, R.P., Palmer, R.M., Da Silva, P. & Cruickshank, M.A. (2000) Relationship between nutritionally-mediated placental growth restriction and fetal growth, body composition and endocrine status during late gestation in adolescent sheep. *Placenta*, **21**, 100–108.

Wood, S. (2006) *Generalized Additive Models: an Introduction with R*. Chapman & Hall/CRC. Boca Raton, Florida.

Received 21 July 2006; accepted 27 March 2007

Supplementary material

The following supplementary material is available for this article online.

Appendix S1. Derivation of the estimated regression coefficients and their standard errors from a difference penalty regression (DPR).

Appendix S2. Example Genstat code to fit a DPR model with a penalty on first differences.

Appendix S3. Simulated data used to fit DPR in Fig. 2c.

Appendix S4. Example R code to fit a DPR model with a penalty on first differences.

Appendix S5. Simulation study results comparing least squares regression (LSR), ridge regression (RR) and difference penalty regression (DPR).

Fig. S1. The relationship between rainfall experienced during pregnancy and the birth weight of North Country Cheviot sheep.

Fig. S2. The relationship between temperature experienced during pregnancy and the birth weight of red deer.

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/full/10.1111/j.1365-2656.2007.01251.x

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.