23 Investigating the effects of rice farming on aquatic birds with mixed modelling

Elphick, C.S., Zuur, A.F., Ieno, E.N. and Smith, G.M.

23.1 Introduction

Ecologists are frequently interested in describing differences among the ecological communities that occur in habitats with different characteristics. In an ideal world, experimental methods would standardise situations such that each habitat variable could be altered separately in order to investigate their individual effects. This approach works well in simple ecosystems that can be replicated at small spatial scales. Unfortunately, the world is not always simple and many situations cannot be experimentally manipulated. Investigating specific applied questions, in particular, often can be done only at the spatial scales at which the applied phenomena occur and within the logistical constraints imposed by the system under study. In such cases, one is often left with the choice between collecting "messy" data that are difficult to analyse or avoiding the research questions entirely. In this chapter, we investigate just such a case, in which applied ecological questions were of interest, but experimental influence over the system was not possible.

Agricultural ecosystems dominate large areas of the Earth, and they are used by a wide variety of species. Consequently, conservation biologists have become increasingly interested in understanding how farmland can contribute to biodiversity protection (McNeely and Scherr 2002; Donald 2004). Rice agriculture, in particular, has great potential to provide habitat because it is a dominant crop worldwide, and because the flooded conditions under which rice is typically grown, simulate —if only approximately— wetland habitats, which have been widely drained in agricultural areas. In the Central Valley of California, USA, rice is grown in an area of extremely intensive agriculture that historically was a vast complex of seasonally flooded wetland and grassland habitat. In the past, these wetlands supported millions of migratory waterbirds during winter, and the area remains one of North America's most important areas for wintering waterbirds (Heitmeyer et al. 1989).

Traditionally, most California rice fields have been flooded only in the summer months when rice is being grown. Recent legislative changes, designed to improve air quality by phasing out the practice of burning residual straw left over after harvest, have resulted in many farmers flooding their fields during winter in order to speed up the decomposition of rice straw before the subsequent growing season. This activity more closely simulates the historic flooding pattern and has been viewed as a potential boon for wetland birds.

In the original study, the main goals were to determine whether flooding fields result in greater use by aquatic birds, whether different methods of manipulating the straw in conjunction with flooding influences how much fields are used, and whether the depth to which fields are flooded is important. Various straw manipulation methods are used by farmers in the belief that they enhance decomposition rates, and usually involve cutting up the straw to increase the surface area, increasing the contact between the straw and the soil, or both. The effects of water depth are of interest because water is expensive and there are economic benefits to minimizing water use. Limiting water consumption is also important to society because there are many competing demands for a limited water supply. Reduced water levels may have agronomic benefits too, because anoxic soil conditions, which would slow straw decomposition, are less likely with shallow flooding.

The data used in this chapter were collected during 1993–1995, and come from winter surveys of a large number of fields in which waterbirds were counted and identified to species. For the purposes of this study "waterbirds" are defined as Anseriformes (swans, geese and ducks), Podicipediformes (grebes), Ciconiiformes (herons, ibises, and allies), Gruiformes (rails, cranes and allies) and Charadriiformes (shorebirds, gulls, and allies). The data have been analysed previously, both to look for effects on individual species (Elphick and Oring 1998) and to examine composite descriptors of the waterbird community (Elphick and Oring 2003).

Although the study was planned with experimental design principles in mind, the reality of the situation limited the rigour with which these ideals could be followed. Different straw management treatments could not be assigned randomly because their application depended on the separate decisions made by individual farmers. Replication was possible but was similarly constrained by the actions of different growers. Moreover, because all fieldwork was conducted on privately owned land, data collection was subject to the constraints placed by the landowners. Eventually, farms were selected such that each management technique was represented by as many fields as possible and that fields with different treatments were interspersed among each other. Layered on top of these constraints, the nature of the data created various problems for common statistical analyses. In short, from an ecologist's perspective the dataset was an analysts' nightmare! Because of all these issues, the original analyses relied primarily on non-parametric statistical tests and were not well suited to examining different explanatory variables simultaneously or for testing interactions among explanatory variables. In this case study chapter we will apply mixed modelling techniques to deal with the complicated structure of the data.

In the initial analysis, flooding fields clearly had a strong effect on waterbird use, but the effects of the way in which flooding occurred were more ambiguous. Our main questions here, then, concern whether the method of straw manipulation and the water depth in flooded fields influenced waterbird use of the fields. Additionally, we test whether there was geographic and temporal variation in bird numbers. To fully understand all the steps carried out in this case study chapter, it might be helpful to review the theory on mixed models (Chapter 8) afterwards.

23.2 The data

The basic experimental design for this study involved collecting data that were nested at three spatial scales. The main units of study were fields, and it was at this scale that straw management treatments were applied. Some (but not all) fields were subdivided into units referred to as 'checks'. Checks are simply sub-sections of a large field that are separated from each other by narrow earthen levees in order to help farmers control water depth. In this study, data were collected separately for individual checks because water depth typically varied a great deal among the checks within a field, but very little within checks. Collecting data for each check separately, therefore, makes it easier to test for depth effects by ensuring uniform depth within each sampled unit. Fields were also spread over a large portion of the rice-growing region, but were clustered into three geographic 'blocks'. Sampling over a large area was considered important to ensure that the results were broadly applicable. But clustering sampled fields were also necessary in order to facilitate data collection. In our analysis, we wanted to test for differences among these blocks both to determine whether there were geographic differences in the use of fields by birds, and to determine whether the main results applied throughout the rice-growing region. The overall design, then, has checks nested within fields, which are nested within geographic blocks. In all there were 26 fields in block 1, 10 fields in block 2 and 6 fields in block 3 (note that depth data were lacking for a few fields included in the original study, and so the sample sizes here differ slightly from the totals given in the previous papers). The number of checks in a field varied from only 1 in the smallest fields to as many as 16 in the largest. A schematic overview of the data is presented in Figure 23.1.

Birds were counted during surveys that were conducted at approximately 10-day intervals (see Elphick and Oring 1998 for details). A total of 25 surveys were made, 12 between Nov 1993 and Mar 1994 and another 13 between Nov 1994 and Mar 1995. In principle we could make the schematic overview in Figure 23.1 considerably more complex by adding "Survey" as another level in the structural hierarchy. With a few exceptions, each field was visited during every survey. Blocks 1, 2 and 3 had a total of 446, 750 and 836 observations, respectively. We omitted three observations that had extreme water depths (> 60 cm). Possible explanatory variables are water depth (which varies at the check level), straw treatment (which varies at the field level and is labelled as Sptreat in this chapter; see Table 23.1), Block, Field, Survey, and Year.

A complicating factor is that both fields and checks varied in size. The response variables are either numbers or densities of 45 different bird species. If we wanted to work with numbers, then the size of the area would have to be used as an offset or weighting variable; for simplicity in this analysis we used densities.

From an analyst's perspective a major problem with the dataset is that 95% of the observations are equal to zero, which might cause all kinds of statistical problems, including violation of homogeneity (Chapter 5) or overdispersion (Chapter 6), if GLM models were to be applied. The specific aim of this case study chapter is to demonstrate how univariate methods (especially mixed modelling) can be used. Therefore, we pooled the data for all waterbird species to give a single variable (called Aqbirds) and tested models that explain patterns of overall waterbird abundance. Hence, we will not consider the application of multivariate methods like ANOSIM and NMDS (Chapters 10 and 15), which might also help explain the relationships between waterbird use and the various explanatory variables.

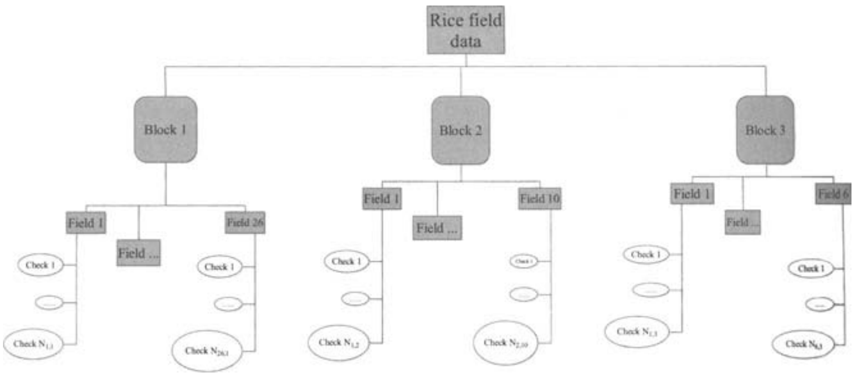

Figure 23.1. Schematic overview of the rice field data. In principle we could make the schematic overview considerably more complex by adding 'Year' and 'Survey as additional levels at the bottom of the structural hierarchy. The surveys were conducted from November to March in each winter.

Table 23.1. Levels of the treatment variable SPTREAT and the corresponding numerical codes used in this chapter.

Treatment	Numerical Code In This Chapter	Description
fld	1	Field flooded with no straw manipulation
rl+fld	2	Field rolled to mix straw and soil, and then flooded
fld+rl	3	Field flooded, and then rolled to mix straw and soil
chp+fld	4	Straw chopped up, and field then flooded
inc+fld	5	Straw incorporated into soil, and field then flooded
rmv+fld	6	Straw removed entirely, and field then flooded

23.3 Getting familiar with the data: Exploration

When deciding how to analyse the rice field bird data, we first had to make a series of decisions about how to structure our models. The first problem was that the data have a hierarchical structure, with checks nested within fields, and fields within blocks. A possible model is

Aqbirds = constant + Depth + factor(Block) + factor(Field) +....

where factor() represents a nominal variable. In the initial dataset, 'Block' was labelled as 1 to 3 and 'Field' was labelled as 1 to 26 for the fields within block 1, 1 to 10 for the fields within block 2 and 1 to 6 for the fields from block 3. This labelling was a problem because field 1 in block 1 is not the same as field 1 in blocks 2 or 3; yet the function factor() would erroneously aggregate all observations with the same field value. Consequently, we re-labelled the fields so that each has a unique label with levels 1 to 42. Using this nominal explanatory variable in the model above means that we must estimate 41 (42 levels minus the baseline) separate intercepts for the field variable alone, which is a lot of parameters to estimate, especially as we are not actually interested in the effects of individual fields. We cannot ignore these effects totally because some of the variation in the data might be due to between-field variation, and accounting for this variation will help us to test for other patterns. To reduce the number of parameters that must be estimated, therefore, we treat Field as a random component rather than as a fixed effect in the model. This approach basically means that we assume that the fields in our study are a random sample from a large number of possible fields, and that we can model all of the variation among fields as a random process that can be described in simple mathematical terms. To be more precise, instead of paying the price of 41 degrees of freedom in order to estimate separate intercepts for every field, we assume that all fields have intercepts that are normally distributed with mean 0 and a certain variance. With this formulation, the only unknown component that must be estimated is this variance, which 'costs' us only one degree of freedom.

We can make the variance structure a bit more complicated and account for the nested design (see Quinn and Keough 2002 for examples) with something like:

Aqbirds = constant + Depth + Sptreat + Block/Field/Year/Survey + noise

The notation Block/Field is a computer notation for nested random effects, and in this study we can extend this approach to incorporate four nested levels. In theory, we could use random components for Field, Block, Survey and Year. We are, however, interested in the specific geographic differences among the blocks, and in any seasonal pattern exhibited by Survey. Consequently, these variables are better left as fixed effects. It would be suitable to model Year as a random component, but there were only two years studied. Generally it is best to have at least four or five levels if one is to treat a variable as a random component instead of fixed because this results in more stable variance estimation and the degree of freedom savings are minimal otherwise. So, we keep the models simple and use something of the form:

Aqbirds = constant + Depth + Sptreat + Year + Survey + Block + Field + noise (23.1) where only Field and noise are random components, and the rest are considered to be fixed. We call this model a mixed effects model (Chapter 8). Deciding whether to view variables as random or fixed was relatively easy, mainly driven by the underlying questions and the nature of the variables.

The next problem was to determine how to test for temporal patterns. Although each year's data were collected in the same period, the timing of each round of surveys differed somewhat between years due to unexpected restrictions on access to certain sites. For example, although dates for the first survey in each winter coincide, by the fourth survey the timing was quite different (survey 4 occurred from 4–7 January in the first winter and from 19–21 December in the second). This mismatch means that it is misleading to treat a given survey i as temporally equivalent in the two years. Consequently, we decided to analyse each year's data separately. Hence, we removed Year from our previous model and planned to apply the following model separately to data from each year:

$$\text{Aqbirds} = \text{constant} + \text{Depth} + \text{Sptreat} + \text{Survey} + \text{Block} + \text{Field} + \text{noise} \quad (23.12)$$

This model implies a linear relationship between bird densities and water depth, which would not necessarily be expected from a biological perspective. So, we next need to determine whether the effect of depth is indeed linear, and the best way to do this is through data exploration. We should also verify that there are no outliers in these data. Figure 23.2 shows a Cleveland dotplot (Chapter 4) of the response variable Aqbirds for both years combined. Although there are no single extremely large observations, the data would benefit from a transformation as approximately 25 observations (from the second winter) are at least twice as large as the remaining values. QQ-plots (not presented here) show clearly that a $\log_{10}$ transformation is better than a square root or no transformation (see Chapter 4 for a discussion on transformations and how to choose among them). A constant value of 1 was added to each bird density value to avoid problems caused by taking the log of zero. After pooling data for all species the percentage of zero observations in the response variable was much reduced, but still accounted for 35% of the observations.

Figure 23.3 shows a coplot of $\log_{10}(Y+1)$ transformed waterbird densities versus Depth conditional on the variables Sptreat and Year. To aid visual interpretation, a LOESS smoother with a span of 0.7 (Chapter 7) has been added. Note that there appears to be evidence for a non-linear relationship between depth and waterbird densities.

A boxplot of the log transformed waterbird data conditional on survey and year is given in Figure 23.4. The 12 boxplots on the left-hand side are from the winter of 1993/94, the 13 on the right from 1994/95. Note that there is a clear seasonal pattern in the first winter, but that this pattern is less clear in the second winter. This difference may relate to very different weather patterns in the two years, with extensive rainfall and natural flooding in the second year that caused many fields to stay wet longer than normal (Elphick and Oring 1998). If the data from both years were to be analysed together, the patterns in the boxplots indicate that interaction terms between Year and other variables may be required.

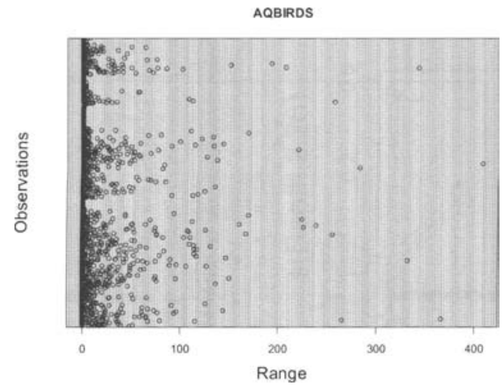

Figure 23.2. Cleveland dotplot of waterbird densities. The horizontal axis represents the values of the observations, and the vertical axis corresponds to the order of the data. The observation at the bottom is the first row in the spreadsheet, and the observation at the top the last.

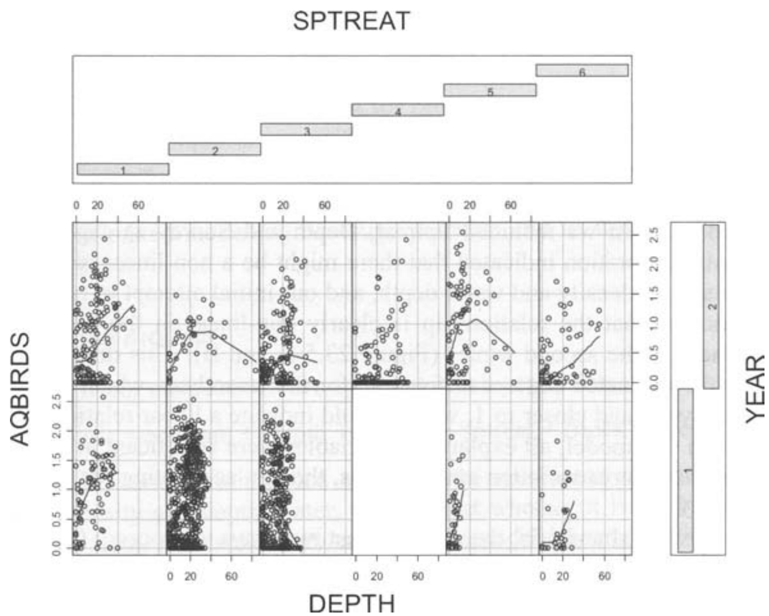

Figure 23.3. Coplot of waterbird densities versus depth conditional on Sptreat and Year. The lower six panels correspond to treatments 1 to 6 (see Table 23.1) in year 1; the upper six panels to the same treatments in year 2. Smoothing curves were added to aid visual interpretation. Treatment 4 was not measured in year 1.

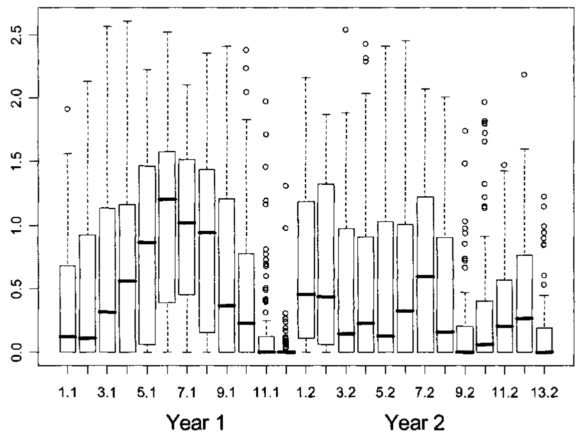

Figure 23.4. Boxplots of log transformed waterbird data conditional on survey and year. The first 12 boxplots from the left are for year 1, the others for year 2. The labelling '*i.j*' refers to year '*j*' and Survey *i*.

23.4 Building a mixed model

Now that we have a sense of what the data look like, we can start building an initial model. First we included Sptreat, Depth and Survey as explanatory variables. Data exploration indicated that there might be a non-linear relationship between waterbird density and water depth, and our initial attempt to use an additive model showed that the relationship is clearly non-linear in the first winter and nearly linear in the second winter (Figure 23.5-A and B). This difference is indicated by the estimated degrees of freedom for the smoothers, which are smaller in the second year, and closer to 1, which would indicate a linear relationship (Chapter 7). With this model, all explanatory variables were significant at the 5% level. This analysis ignores variation among fields, though, accounting for it only as part of the error term.

From this initial analysis, there are at least two ways to proceed: (i) Apply additive mixed modelling to allow for a non-linear depth effect and the between-field variation using a random component or (ii) apply linear mixed modelling using a polynomial function of depth and between-field variation using a random component. So, the problem is how to model the non-linear depth effect. Because (i) additive mixed modelling has already been used in two other case study chapters, and (ii) software for linear mixed modelling is better developed than for additive mixed modelling, we use the former here.

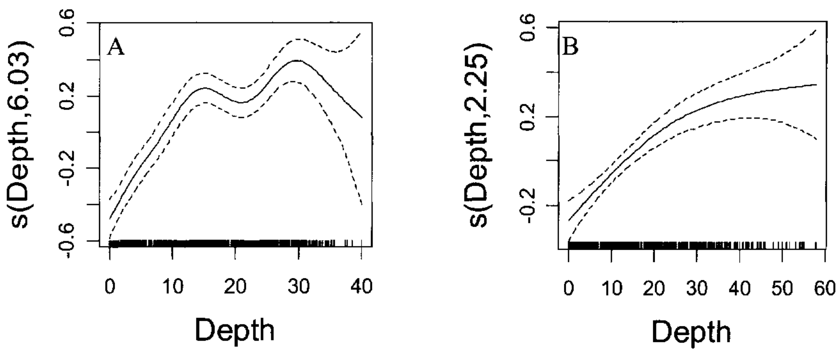

Figure 23.5. Smoothing curve obtained by an additive model using the year 1 data (panel A) and the year 2 data (panel B), showing the effect of water depth on waterbird densities. Dotted lines are 95% confidence bands. Vertical lines at the bottom indicate the values of depth at the corresponding site. The vertical axes represent the contribution of Depth to the fitted bird values (Chapter 7). Cross-validation was used to estimate the degrees of freedom for each smoother.

The limited availability of software for (generalised) additive mixed modelling is because the technique is reasonably new (Ruppert et al. 2003, Wood 2004, 2006). Application of these methods at present requires expert knowledge and programming skills. As shown in other case study chapters in this book, (generalised) additive mixed modelling has huge potential and its use in ecology can be expected to grow.

Linear mixed models

The first linear mixed model we applied to each year's dataset was as follows:

M_0: Aqbirds = constant + Depth + Depth^2 + Sptreat + Block + Field + noise

The Depth^2 term was included to model the non-linear relationship between waterbird density and depth. Later, we will test whether it is really necessary to include this term. All explanatory variables in model M_0 are nominal except for Depth and Depth^2 and only Field was treated as a random component. Unfortunately, this model crashed, producing the message: 'NA in data'. The reason for this error message was that some straw treatments only occurred in one block causing a perfect collinearity between Block and Sptreat. To solve this problem we would either have to eliminate certain straw treatments or drop one of the confounded terms from the analysis. As we are primarily interested in the straw effect, we chose to omit Block. The new model becomes

M_1: Aqbirds = constant + Depth + Depth^2 + Sptreat + Field + noise

The next step was to assess whether a seasonal effect should be included in the model. To do this we plotted standardised residuals from model M_1 versus Survey (Figure 23.6). The resulting boxplot shows a clear seasonal pattern, suggesting that adding a time component is important:

M_2: Aqbirds = constant + Depth + Depth^2 + Sptreat + Survey + Field + noise

The residuals from this model did not show a seasonal pattern when medians were compared, but the amount of variation still varied among surveys, which violates the homogeneity assumption (Figure 23.7). Because heterogeneity of variance is a serious problem, the *p*-values produced by this model should not be used. Mixed modelling provides a solution, especially if we use separate variances to different portions of the data.

To clarify what we have done so far, the terms Depth, Depth^2, Sptreat and Survey were all used as fixed components in model M_2. Field was used as a random component and is an important part of the model because it reduces the portion of the variation that cannot be explained at all. The remaining unexplained variance was treated as another random component named 'noise', although this notation is a bit sloppy. Elsewhere in the book the residuals that describe this 'noise' have been referred to as ε. The homogeneity assumption applies to these residuals, which are assumed to be normally distributed with expectation 0 and variance σ^2. With mixed modelling, we can account for the heterogeneity by assuming a different value for the residual variance associated with each survey (labelled σ_1^2 to σ_{12}^2 in year 1), rather than just assuming that all residuals are distributed with a mean of 0 and a single variance of σ^2. This gives the following model:

M_3: Aqbirds$_j$ = constant + Depth + Depth^2 + Sptreat + Survey + Field + ε_j

where $\varepsilon_j \sim N(0, \sigma_j^2)$ and Field $\sim N(0, \sigma_{Field}^2)$. Incidentally, instead of using separate variances σ_j^2 for each survey, we could try to account for heterogeneity by using different variances for each straw treatment. In the next section, we will discuss model M_3 for the year 1 data and once we have found the optimal model in terms of random terms we will try to identify unnecessary fixed terms.

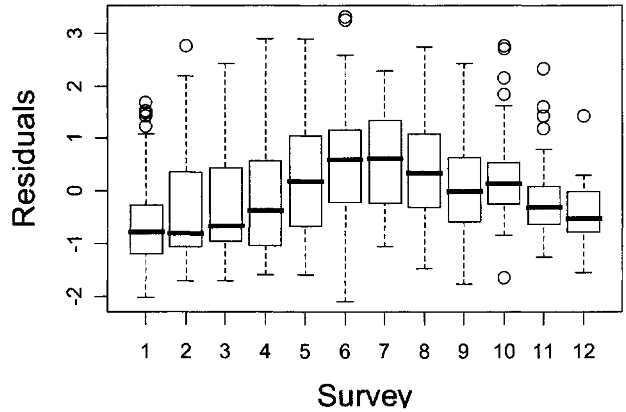

Figure 23.6. Standardised residuals of model M_1 plotted versus Survey for year 1. A strong seasonal trend occurs in both the median values and in the amount of variation within each survey.

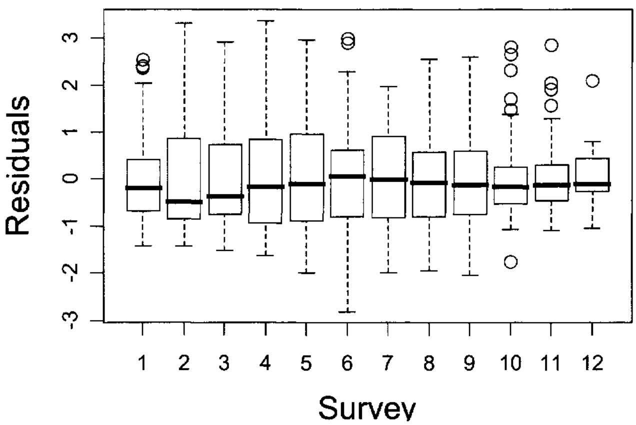

Figure 23.7. Standardised residuals of model M_2 plotted versus Survey for year 1. There is no clear seasonal variation in the median residuals values per survey anymore, although the variance heterogeneity remains, with especially smaller residuals in the later surveys.

23.5 The optimal model in terms of random components

In the previous section we developed a model that contained the random component Field and, in model M_3 also introduced 12 different variance components, one for each survey. Although the modifications introduced in model M_3 solve the variance heterogeneity problem, we are paying a high price in terms of degrees of freedom, because we now have to estimate an extra 11 parameters. Using the AIC or the likelihood ratio test, we can compare models M_2 and M_3 and assess whether this price results in significant model improvement:

Model	df	AIC	logLik	L.Ratio	p value
M_2	20	2146.58	−1053.29		
M_3	31	2006.26	−972.13	162.32	<0.001

In this comparison the smaller AIC for M_3 indicates that the more complex model is the better one and the likelihood ratio test allows us to go further and say that using the 12 variances is a significant ($p < 0.001$) improvement. We can also determine whether we need such a construction for Sptreat as well. In this case, we compared model M_2 with a model M_4 that containing six separate variances, one for each straw treatment. In this case, however, the likelihood ratio test gave $p = 0.094$ indicating that M_2 is the better of these two models. If M_4 had proven better than M_2, then we would have had to choose between this model and M_3. Because these models are not nested, a likelihood ratio test would have been inappropriate and we would have had to choose by comparing AIC values.

So far, our best model is M_3, but this model assumes independent residuals ε. Because the data form a time series, this assumption may be invalid, so we tested the effect of adding a simple auto-correlation structure in which each survey is treated as a point in time. We used an ARMA(1,0) structure for the noise, see also Chapters 16, 27 and 36. In such a model we allow for a relationship over time in the error term: $\varepsilon_t = \theta\varepsilon_{t-1} + \eta_t$ where η_t is independently and normally distributed noise. In the first year, we have up to 12 sequential observations for each check in a particular field. The model will impose the same auto-correlation structure for all checks. Hence, it will give one value for θ describing the overall auto-correlation in the error term for all checks. For longer time series, more complicated auto-correlation structures might be used. Model M_5 is the same as M_3, except that the error term is now allowed to be ARMA(1,0). When we tried to run this model we initially got error messages, caused by some checks with very short time series (only four or five surveys). This problem arose because there were a few checks in which depth measurements could not be made on every survey because depth stakes that had been placed in fields at the start of the field season were removed or chewed through at some sites. Vandalism by muskrats and theft were not accounted for in the study design. Waterbird densities were still obtained for these checks on every survey, but the data could not be used in our analysis because of the missing depth data. Luckily, this problem affected few checks, but the reader who wants to carry out the analysis should remove checks 5 and 6 of field 302, and checks 2–5 of field 305 for the year 1 data. Given these data discrepancies, we refitted M_3 with the reduced dataset in order to compare it to M_5:

Model	df	AIC	logLik	L.Ratio	p value
M_3	31	1925.33	−931.66		
M_5	32	1866.60	−901.30	60.2	<0.001

The likelihood ratio test shows that allowing for auto-correlation between the residuals gives a significant improvement. The estimated value of θ was 0.274. We will discuss the significance levels of the fixed terms later, but it is interesting to mention one of them here and emphasise the implications of auto-correlation. In model M_3, the straw treatment effect produced a p-value of 0.007, but after auto-correlation was accounted for in model M_5, the p-value for this variable was 0.045. In other words, if we had ignored auto-correlation and used model M_3 we would have concluded, wrongly, that the effect of straw management is highly significant when in fact it only just approaches the $p = 0.05$ significance threshold.

Dropping fixed terms from the linear mixed model

Having found the optimal model for the year 1 data in terms of random components, we now investigate whether any of the fixed components can be dropped from the model or if interaction terms should be added. Our starting model is M_5, and we consider each of the following alternatives:

- Drop only Survey (M_6).
- Drop only Sptreat (M_7).
- Add an interaction between Depth and Sptreat (M_8).
- Drop the quadratic Depth term to test whether the depth relationship is linear or non-linear (M_9).
- Drop both the linear and quadratic Depth terms to test whether depth is related to waterbird density at all (M_{10}).

This leads to the following models:

M_6: Aqbirds = constant + Depth + Depth^2 + Sptreat +
Field + Variance for each survey + auto-correlated noise

M_7: Aqbirds = constant + Depth + Depth^2 + Survey +
Field + Variance for each survey + auto-correlated noise

M_8: Aqbirds = constant + Depth + Depth^2 + Sptreat + Survey +
Depth* Sptreat +
Field + Variance for each survey + auto-correlated noise

M_9: Aqbirds = constant + Depth + Sptreat + Survey +
Field + Variance for each survey + auto-correlated noise

M_{10}: Aqbirds = constant + Sptreat + Survey +
Field + Variance for each survey + auto-correlated noise

To compare mixed models with different fixed components, the estimation routine should use maximum likelihood estimation instead of restricted maximum likelihood estimation (REML). See Chapter 8 for details; note that AIC values obtained with these methods should not be compared with each other. The results are given in Table 23.2. First, M_5 and M_6 are compared to test the null hypothesis that the regression parameters for Survey equal zero. The likelihood ratio test shows that this hypothesis can be rejected ($p < 0.001$) and that dropping Survey results in a significantly worse model. The next comparison also shows that reducing the model causes a significant worsening, and that straw treatment affects waterbird densities. This result, however, is only weakly significant ($p = 0.045$), indicating that the Sptreat effect is much smaller than the Survey effect.

The last three models all examine the effects of water depth. Adding an interaction between Depth and Sptreat slightly improves the model (i.e., it produces a smaller AIC value). But this difference gives a p-value of 0.055, suggesting that it is of borderline significance. The quadratic Depth term also produces only a marginal difference, although there is significant improvement when this term is included in the model. In contrast, removing depth entirely, by dropping both the Depth and Depth^2 terms, results in a highly significant worsening of the model.

So, which model should we choose? Clearly Survey and Depth should be in the chosen model, as dropping either leads to a substantially worse model. Including Sptreat, Depth^2, or the Sptreat × Depth interaction all cause marginal improvements in the model, but none of these terms is as important as the other two. Whether they should all be included, then, depends on how rigidly you want to stick to the 0.05 significance criterion. Formally, choosing the final model should involve an iterative process in which the entire process outlined here is repeated after each variable is dropped, in a backward selection style.

Table 23.2. Comparing models with different fixed effects. The AIC for M_5 is 1778.92. Either the AIC or (better) the likelihood ratio test can be used to compare the nested models.

Test models	Term deleted	AIC	L-Ratio	p-value
M_5 versus M_6	Survey	1956.830	199.904	<0.001
M_5 versus M_7	Sptreat	1780.620	9.694	0.045
M_5 versus M_8	Add interaction Sptreat and depth	1777.645	9.280	0.055
M_5 versus M_9	Depth^2	1781.325	4.399	0.035
M_5 versus M_{10}	Depth and Depth^2	1793.191	18.266	<0.001

23.6 Validating the optimal linear mixed model

Assuming that we accept model M_8, the next thing we have to do is to inspect the residuals and check whether normality and homogeneity assumptions hold. Figure 23.8-A shows the fitted values versus standardised residuals. Note that the residuals show a pattern of bands that is characteristic to all linear regression, GLM, mixed models and GAM models when there are lots of observations with the same values; see Draper and Smith (1998) for an explanation. Given that our dataset has so many zeros, this pattern is not surprising and we should not worry about straight lines of residuals. The residuals are standardised, hence, values larger than 2 or –2 are suspect (Chapter 5). There is no clear violation of the homogeneity assumption, and only a few observations are larger than 2 given the size of the dataset. Figure 23.8-B shows a histogram of these residuals and indicates that they are slightly skewed.

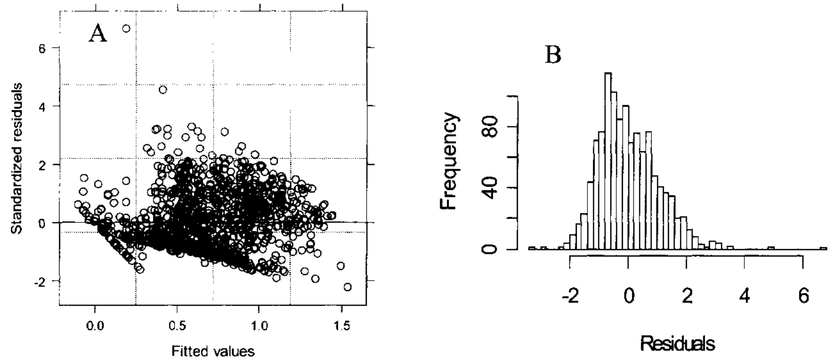

Figure 23.8. A: Fitted values versus standardised residuals obtained by model M_8. B: Histogram of residuals obtained by model M_8.

23.7 More numerical output for the optimal model

The last thing we have to do is to discuss the numerical output for model M_8 (obtained by REML estimation).

```
Random effects:
Formula: ~1 | Field
            (Intercept)     Residual
StdDev:      0.099          0.479

Correlation Structure: ARMA(1,0)
Formula: ~Time | Field/Check
Parameter estimate(s):
Phi1
0.278
```

This information tells us that variation among fields is quite small. The variance between fields and the residual variance are 0.0099^2 and 0.479^2, respectively. Hence, the variation among fields is considerably less than the residual variance. Because the Field effect is small, we also compared various mixed models with corresponding GLS models (a GLS is a mixed model without the random intercept). Comparing AIC values suggested that the mixed models were better, indicating that the random Field component is needed even though the effect is small. The auto-regressive parameter has a value of 0.274. The following lines give the different monthly variances.

```
Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | Survey
 Parameter estimates:
     1      2      3      4      5      6      7      8
 1.000  1.390  1.343  1.407  1.320  1.439  1.209  1.277
     9     10     11     12
 1.289  1.073  0.717  0.350
```

In the analysis, survey 1 is arbitrarily set as the baseline survey, and the variance parameters for the other surveys are scaled relative to this baseline. Examining the pattern in these variance estimates shows that the first survey and the last three surveys have a relatively lower spread than the others. In other words, variation is lower at the start and end of the winter season when bird numbers are low, and variation increases mid-winter when there are large numbers of wintering waterbirds present in the region (see Figure 23.4).

Let us now focus on the estimated parameters and standard errors (obtained by REML estimation) of the fixed components.

	Value	Std. Error	Df	*t*-value	*p*-value
(Intercept)	0.139	0.126	1118	1.097	0.273
Depth	0.037	0.010	1118	3.733	<0.001
I(Depth^2)	-0.001	0.000	1118	-2.483	0.013
Sptreatfldrl	-0.129	0.094	31	-1.363	0.183
Sptreatincfld	-0.160	0.147	31	-1.088	0.285
Sptreatrlfld	-0.074	0.081	31	-0.905	0.372
Sptreatrmvfld	-0.149	0.141	31	-1.053	0.301
Survey2	0.137	0.078	1118	1.768	0.077
Survey3	0.186	0.076	1118	2.450	0.015
Survey4	0.278	0.084	1118	3.317	0.001
Survey5	0.480	0.081	1118	5.930	<0.001
Survey6	0.697	0.087	1118	8.015	<0.001
Survey7	0.656	0.079	1118	8.287	<0.001
Survey8	0.622	0.083	1118	7.517	<0.001
Survey9	0.417	0.087	1118	4.785	<0.001
Survey10	0.427	0.104	1118	4.107	<0.001
Survey11	0.117	0.098	1118	1.199	0.231
Survey12	-0.032	0.093	1118	-0.341	0.733
Depth:Sptreatfldrl	-0.011	0.006	1118	-1.797	0.073
Depth:Sptreatincfld	0.026	0.020	1118	1.312	0.190
Depth:Sptreatrlfld	-0.006	0.005	1118	-1.003	0.316
Depth:Sptreatrmvfld	-0.021	0.011	1118	-1.953	0.051

Just as in linear regression, one should interpret the *p*-values with great care, and the *p*-values in Table 23.2 are better for assessing the importance of each fixed component. It is, however, useful to look at the estimated parameters and their signs. In addition, this output provides some insight into the nature of the differences between levels for the nominal variables. According to this output, the linear depth term is positive and highly significant, indicating that waterbird densities increase with water depth, and the quadratic term is negative and weakly significant, suggesting that this pattern may weaken, or even reverse, in deeper conditions (see also Figure 23.3). The parameter estimates for Sptreat arbitrarily use the first treatment (Sptreatfld) as a baseline, and the negative values for the other treatments indicate that densities were highest for the baseline treatment. The standard errors and p-values for these parameter estimates, however, indicate that differences among treatments are relatively slight, which accords with the earlier conclusion that this effect is only marginally significant. (Note that the need to exclude a few checks due to missing data, resulted in one straw treatment being dropped from the final analysis.) Parameter estimates for Survey are also scaled relative to the first level for the variable (survey 1). In this case, most parameters are positive, reflecting the higher densities of birds in subsequent surveys. More careful inspection shows that Survey parameter estimates steadily increase until survey 6, and then decline. This change matches the buildup of birds during early winter, as migrants arrive from their northern breeding grounds, followed by a gradual decline in late winter as birds begin to move north again (Figure 23.4). Lastly, the parameters that describe the Sptreat × Depth interaction indicate that the effect of depth in fields where straw has been removed (Sptreatrmvfld) is somewhat different from that in the other straw treatments.

Year 2 data

Results for the year 2 data were very similar to those presented for year 1. This analysis needed a separate variance for each survey, a random field component and an auto-correlation structure ($\theta = 0.183$). There was a significant survey effect ($p < 0.001$), a significant straw treatment effect ($p = 0.010$), a weak interaction between straw treatment and depth ($p = 0.056$), and a strong non-linear depth effect ($p < 0.001$ for the linear term and $p = 0.001$ for the quadratic term). We leave it as an exercise for the user to validate the model.

23.8 Discussion

In this chapter, linear mixed modelling techniques were used to analyse the waterbird data. For the year 1 data, detailed model selection indicated that there were highly significant survey and depth effects. The pattern of spread in the residuals varied among surveys, and a random field component was needed in the model. A likelihood ratio test also indicated that we should include an auto-correlation term. Without accounting for auto-correlation, the model indicated a highly significant straw management effect, but once the auto-correlation was added, p-values for straw management and its interaction with depth indicated only weak significance.

This analysis improved on the earlier attempt to analyse these data (see Elphick and Oring 1998, 2003) in several ways. In the previous analyses, the problems created by the hierarchical structure of the data collection, the variance heterogeneity, and the large number of zeros were dealt with by simply testing each explanatory variable separately using non-parametric tests. This approach was not very satisfactory because it meant that the different variables were not tested simultaneously, which made it difficult to parse out their relative importance. In addition, the earlier analysis side-stepped the temporal pattern in the data, by using mean values calculated across each winter for each check. By using data from each survey, separately, the new analysis presented here provides a picture of how waterbird densities change over time, while simultaneously describing and accounting for the depth and straw treatment effects addressed in the earlier analysis. The new analysis could not, however, solve all of the recognized limitations of the earlier attempt. For example, we had to drop the Block variable here because it was confounded with certain straw treatments. Consequently, we were unable to improve on the earlier treatment of geographic effects.

Overall the re-analysis of these data did not radically change our understanding of the dataset. The new analysis, however, does provide stronger support for some of the inferences made in the earlier papers and a deeper appreciation of how the different explanatory variables interact.

In our analysis, we used waterbird densities per check as our response variable. Another approach would have been to use counts as the response variable, but to do that we would need to take into account the variation in check size. One option would be to use the size of the check as a weighting factor, for example checks with small areas could be down-weighted. Alternatively we could use Area as an explanatory variable in the model. Thirdly, we could use a model in which the size of the check (labelled as 'Area') is used as a so-called offset variable. This means that it has no regression parameter. It can be used in a GLM model with a Poisson distribution, as follows:

$$\text{bird number} = e^{\log(Area)+Depth+Depth^2+Sptreat+\ldots}$$

The term log(Area) is fitted as an offset variable, and all other explanatory variables are treated as usual. To do this, however, we would also need to account for the nested structure of the data, which requires generalised linear mixed modelling techniques that are outside the scope of this book.

Acknowledgement

Data collection for this chapter was made possible by the generous help of many California rice growers, and financial support from the Nature Conservancy, Bureau of Reclamation, Central Valley Habitat Joint Venture, Ducks Unlimited, Inc., and the Institute for Wetland and Waterfowl Research.