36 Time series analysis of Hawaiian waterbirds

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

36.1 Introduction

Surveys to monitor changes in population size over time are of interest for a variety of research questions and management goals. For example, population biologists require survey data collected over time to test hypotheses concerning the patterns and mechanisms of population regulation or to evaluate the effects on population size of interactions caused by competition and predation. Resource managers use changes in population size to (i) evaluate the effectiveness of management actions that are designed to increase or decrease numbers, (ii) monitor changes in indicator species, and (iii) quantify the effects of environmental change. Monitoring population size over time is particularly important to species conservation, where population decline is one key to identifying species that are at risk of extinction.

Our goal in this case study chapter was to analyse long-term survey data for three endangered waterbirds that are found only in the Hawaiian Islands: Hawaiian stilt (*Himantopus mexicanus knudeseni*), Hawaiian coot (*Fulica alai*), and Hawaiian moorhen (*Gallinula chloropus sandvicensis*). The survey data come from biannual waterbird counts that are conducted in both winter and late summer. Surveys were initiated for waterfowl in the 1940s on most of the major Hawaiian Islands, and they were modified in the 1960s to better monitor the endangered waterbirds as well (Engilis and Pratt 1993). In the 1970s, surveys were expanded to encompass all of the main Hawaiian Islands. Surveys are coordinated by the Hawaii Division of Forestry and Wildlife, and the goal is to complete each statewide survey in a single day to reduce the risk of double counting individuals.

We analysed winter survey data to reduce the contribution of recently hatched birds to survey numbers, which can increase count variability. There are two previously published analyses of population trends of these species. Reed and Oring (1993) analysed winter and late-summer data for Hawaiian stilts and concluded that there was a statewide increase in population size during the 1970s and 1980s. Engilis and Pratt (1993) analyzed data from 1977–1987 for all three of the species addressed in this chapter, and found that annual rainfall patterns had a strong effect on counts. They concluded that, for this time period, moorhen and coot numbers varied with no clear trend, whereas stilt numbers rose slightly. Trend analyses in these papers were based on visual inspection (Engilis and Pratt 1993) or used linear and non-linear regressions (Reed and Oring 1993). Since these techniques ignore time structure, specific time series analysis tools are applied in this case study chapter. We limited our analyses to the islands of Oahu, Maui and Kauai (Figure 36.1) because these islands have the longest survey history and collectively contain the vast majority of the population for each species.

Our primary questions were to determine whether there are detectable population size changes over time, whether trends are similar among time series, and to evaluate the effect of rainfall described for a shorter time span by Engilis and Pratt (1993).

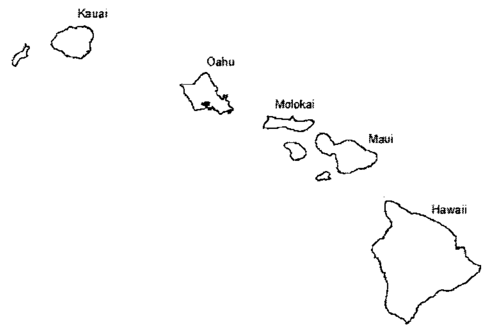

Figure 36.1. Sampling area, showing the main Hawaiian Islands.

36.2 Endangered Hawaiian waterbirds

Hawaiian stilt, coot, and moorhen breed and live year-round in low-elevation wetlands, and they are endangered because of habitat loss and invasion by exotic species that either prey upon them or change the nature of their habitat (Griffin et al. 1989). These waterbirds differ significantly in detectability, which contributes to concerns about how accurately the biannual waterbird surveys enumerate individuals, and whether the number of birds counted is an accurate index of population size (Chang 1990; Engilis and Pratt 1993). Data are considered most accurate for Hawaiian stilts, which spend most of their time in the open where they can be easily seen, and least accurate for moorhen, which live in dense emergent vegetation and have cryptic behaviour (Chang 1990).

Hawaiian stilts and coots are found on all of the major islands (Hawaii, Kauai, Maui, Molokai, Oahu), as well as on Niihau, which shares birds seasonally with Kauai (Engilis and Pratt 1993; Reed et al. 1998). The Hawaiian moorhen's distribution once covered five islands, but now is restricted to Oahu and Kauai. For all three species, population size was estimated only sporadically before the 1950s. Reports, however, tend to describe populations that were fairly large in the 1800s, with declines noticeable by the end of the nineteenth century and continuing until the 1940s (Banko 1987a,b). Since the 1960s, population sizes have increased (Engilis and Pratt 1993; Reed and Oring 1993), with current statewide estimates of approximately 1500 stilts, 3000–5000 coots, and 300–700 moorhen. Because Hawaiian moorhen are secretive, their survey results are considered to be an underestimate of the true number (Chang 1990). The time series used in our analyses are given in Table 36.1. Rainfall data came from the National Climate Data Center (http://cdo.ncdc.noaa.gov/ancsum/ACS) for the Kahului airport on Maui.

Table 36.1. Species, islands, and names used in this case study. Stilts and coots move seasonally between Kauai and Niihau, and so for this time series, we only used data from years in which both islands were surveyed for these two species. Moorhen do not occur on Maui, so there is no time series for that combination.

	Species	Island	Name in this chapter
1	Hawaiian Stilt	Oahu	Stilt_Oahu
2	Hawaiian Stilt	Maui	Stilt_Maui
3	Hawaiian Stilt	Kauai	Stilt_Kauai_Niihau
4	Hawaiian Coot	Oahu	Coot_Oahu
5	Hawaiian Coot	Maui	Coot_Maui
6	Hawaiian Coot	Kauai	Coot_Kauai_Niihau
7	Hawaiian Moorhen	Oahu	Moorhen_Oahu
8	Hawaiian Moorhen	Kauai	Moorhen_Kauai

36.3 Data exploration

The first question we asked was whether there are any outliers or extreme observations because of their potential undue influence on the estimated trends. Cleveland dotplots and boxplots (Chapter 4) indicated that various time series had observations that were distinctly larger than in other years within that series. We compared Cleveland dotplots and boxplots of square root and $\log_{10}$ transformed data. Both transformations resulted in considerably fewer influential observations, and the choice of which one to use was merely subjective. Consequently, we used the square root transformed data because it changes the real data less than does the stronger $\log_{10}$ transformation. The second important step in a time series analysis is to make a plot of each variable versus time; lattice graphs are a useful tool for this. In making such a graph, the question one should ask is whether the absolute differences between the time series are important or whether an increase in one abundant species is equally important as a proportionally similar increase in a less abundant species. In the second case, normalised data (Chapter 4) should be used. For this analysis, we made all time series equally important and Figure 36.2 shows a lattice plot of all eight normalised time series. The graph highlights the fact that not all series start at the same time, some series fluctuate more than others, and there are various missing values (these are the gaps in the lines). Note that some time series techniques cannot cope well with missing values (e.g., MAFA). The time series for stilts and coots on Kauai_Niihau have many missing values because Niihau surveys were initiated late in the survey period.

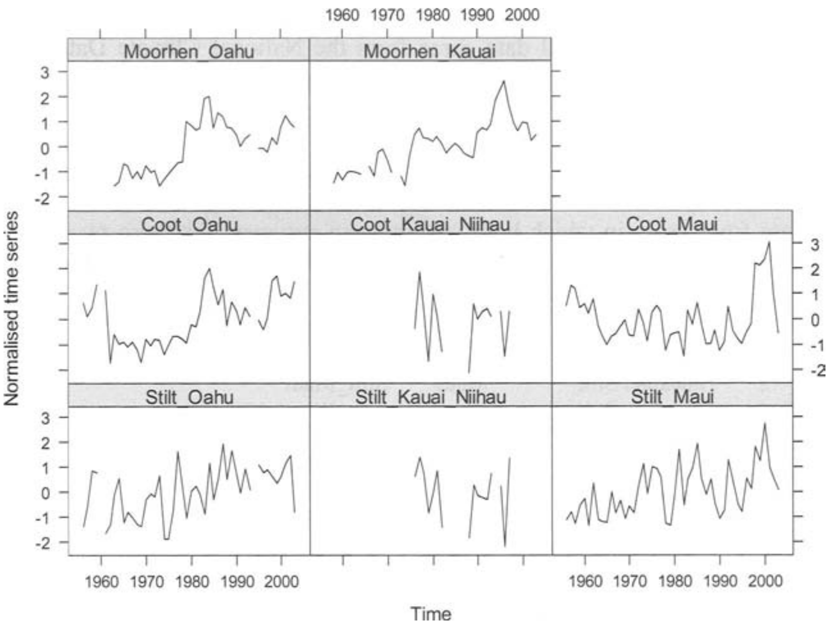

Figure 36.2. Lattice plot of all time series. The horizontal axis shows the time (in years) and the vertical axis the normalised bird abundance.

To obtain insight into the strength of the relationships over time for individual time series, auto-correlation functions were calculated. All series except those for coot and stilt on Kauai_Niihau show an auto-correlation pattern that might indicate the presence of a trend. Recall from Chapter 16 that an auto-correlation function measures the relationship between a time series Y_t and its own past Y_{t-k}. The auto-correlation of a time series with a strong trend will have a high auto-correlation for the first few lags k, compare also Figure 36.3 and Figure 36.2.

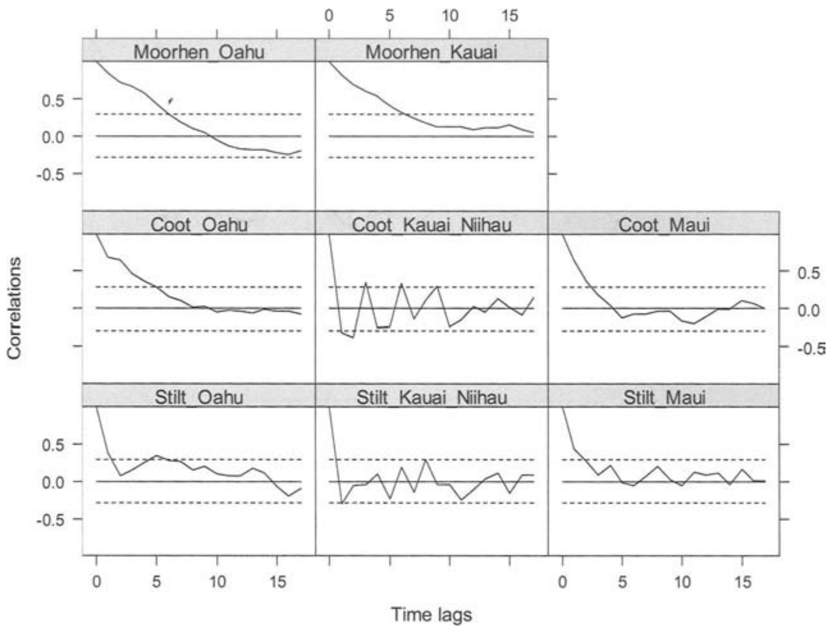

Figure 36.3. Auto-correlation functions for all time series.

36.4 Three ways to estimate trends

LOESS smoothing

The first approach we applied to visualise the possible presence of common trends is LOESS smoothing (Chapter 17). In this method, each (normalised) time series is smoothed (using time as the explanatory variable) and all smoothing curves are plotted in one graph (Figure 36.4). The thick line is the average, which can be calculated because the original series were normalised (Chapters 4 and 17). Note that missing values were omitted from the analysis. If there are only a few missing values, they can be replaced by an appropriate estimate, for example using interpolation or an average value. However, in this case two series have a considerable number of missing values and replacing them by an estimate would be inappropriate. The shape of the smoothing curves does indicate a general increase after about 20 years (i.e., starting in the mid-1970s). However, there is considerable variation among the curves.

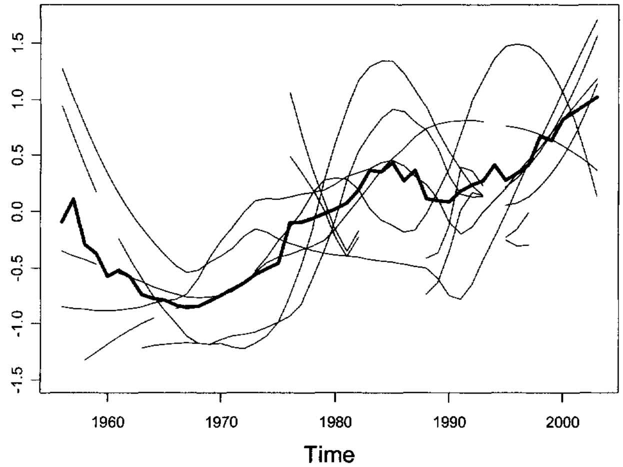

Figure 36.4. Curves obtained by LOESS smoothing. Breaks in lines indicate years in which a time series had missing values. The horizontal axis represents time in years, and the vertical axis shows the values of the smoothers. The time series were normalised (Chapter 4); hence, the vertical axis does not have units. The thick line is the average of the smoothers in each year.

Min/max auto-correlation factor analysis (MAFA)

A more formal way to estimate common trends is min/max auto-correlation factor analysis (Solow 1994). The method was explained in Chapter 17 and in various case study chapters (Chapters 33 and 34). Just like principal component analysis, this method gives axes. However, these axes are derived in such a way that the first axis has the highest auto-correlation with a time lag of one year, which is typically associated with a smooth trend. MAFA extracts the main trends in the time series, and the first MAFA axis can be seen as the most important underlying pattern, or index function in the multivariate time series dataset. This technique cannot cope well with missing values, and because both time series from Kauai_Niihau contained lots of missing values, we decided to omit these time series from the MAFA analyses. MAFA axes are presented in Figure 36.5.

The first MAFA axis looks like a step function: reasonably stable up to the mid-1970s, followed by a rapid increase to the early 1980s, and then stable again. The second axis has high values around 1960, a peak in the mid-1980s and a drop in the mid-1990s. Just as in principal component analysis (PCA), in MAFA we can infer which of the original variables is related to the first axis, to the second axis, and to both, by using each variable's loadings on each axis (Chapter 17). Here, we can address the same question: Which of the original six time series are related to these MAFA axes? One way to answer this question is to calculate Pearson correlations between the original time series and the axes (Table 36.2). A high positive value indicates that the bird time series follows the same pattern as the MAFA axis. A high negative value means that it has the opposite shape.

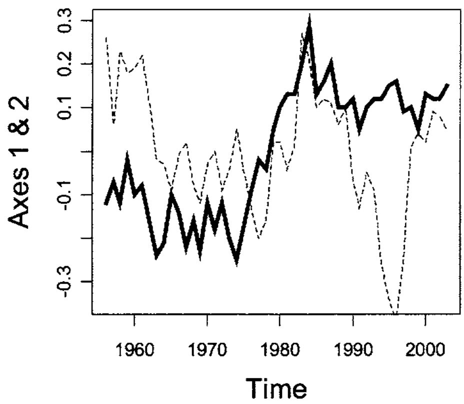

Figure 36.5. Results from the MAFA. The thick line is the first MAFA axis, and the dotted line is the second axis.

Table 36.2. Correlations between the first two MAFA axes and the original time series used for the MAFA.

Time Series	MAFA Axis 1	MAFA Axis 2
Stilt_Oahu	0.56	−0.12
Stilt_Maui	0.41	0.02
Coot_Oahu	0.71	0.47
Coot_Maui	0.05	0.34
Moorhen_Oahu	0.89	0.41
Moorhen_Kauai	0.65	−0.69

Results indicate that the bird time series for stilts, coots and moorhens on Oahu, for moorhens on Kauai, and to a lesser extent for stilts on Maui follow the pattern of the first trend. Indeed, one can recognise this shape in the corresponding time series in Figure 36.2. The second axis is mainly positively related to coot on Oahu and negatively to moorhen on Kauai, with weaker positive relationships with moorhen on Oahu and coot on Maui. Again, one can spot patterns of the second MAFA axis in these time series, especially the big peak (or drop depending on the sign of the correlation) during the mid-1990s. The correlation between the first two MAFA axes and rainfall is not significantly different from 0 at the 5% level. This means that there is no relationship between the two main underlying patterns in the six selected time series (or: MAFA axes) and rainfall. A permutation test indicated that the auto-correlation of the first, and of the second, MAFA axis is significantly different from 0 at the 5% level.

Dynamic factor analysis (DFA)

A more formal tool to estimate common trends and the effects of explanatory variables is dynamic factor analysis (Chapters 17, 33, 34). The DFA model decomposes the eight bird time series into:

- A linear combination of common trends.
- Effects of explanatory variables.
- An intercept for each time series.
- Noise.

The difference between MAFA and DFA is that in MAFA all common information is extracted in the form of trends. In DFA, the common trends represent the information that is common in a set of time series, and that cannot be explained by the explanatory variables. Information that is 'not common' enough is routed to the noise component. As to this noise component, it can be modelled as something that is independent for each bird time series, or alternatively, the noise term of one bird time series can be allowed to interact with that of another bird time series. This brings us close to a generalised least squares approach in which an error covariance matrix can be used to model interactions between noise components (Chapter 17). The interactions of the noise associated with two time series can be seen as unexplained information that exists between the two series, but that forms a pattern not strong enough to justify the need for an extra common trend in the model. Technically, a non-diagonal error covariance matrix is used (Zuur 2003a,b, 2004). For these data, DFA models containing such an error matrix performed better (judged by the AIC) compared with a diagonal error matrix.

DFA can deal better with missing values than can MAFA because it makes use of the so-called Kalman filter and smoother (Zuur et al. 2003a), and therefore all eight time series were used. But it should be noted that DFA cannot do magic with this respect. We noticed that the algorithm became unstable if more than one trend was used. This is probably due to the non-random distribution of the missing values. Therefore, DFA models with only one common trend were used.

Explanatory variables in the DFA

Conservation protection laws for coots and stilts came into force in 1970 and for moorhen in 1967. To test whether these laws had any effect on population sizes, nominal variables can be introduced. For example, we created a new variable called C_{1970} that had values of zero up until 1969 and values of one thereafter. The same was done for C_{1967}; zeros up to 1966 and ones from 1967 onwards. The DFA model allows one to include explanatory variables, but it will produce a different regression coefficient for each time series, and therefore describes a separate effect for each of the eight time series. So, using either C_{1967} or C_{1970} will allow the variable to influence all three species, which might not make sense given the different timing of protection. The shape of the first MAFA axis, moreover, suggests that any population increase is delayed by a few years and starts after 1974. This pattern makes biological sense because any response to management is unlikely to be instantaneous, and because the implementation of new management activities is likely to lag a few years behind new legislation. Also, despite their ecological differences, these three species all occur in the same wetlands and face similar threats. Consequently, protection of management for one species is likely to also have benefits for the others. An alternative, therefore, is to test another nominal variable; let us call it C_j. It is defined as

$$C_j = \begin{cases} 0 & \text{for the years 1956 to j - 1} \\ 1 & \text{for the years j to 2003} \end{cases}$$

We then applied the DFA model (in words):

8 bird time series = intercept + 1 common trend + rainfall + C_j + noise

The questions that then arise are whether rainfall is significant, which value for j should be used, whether C_j is significant and how the trend looks for different values of j. Note that the trend is basically a smoothing curve, and that, in this application, an increasing trend would indicate an overall pattern of population recovery after past declines.

Just as in linear regression we obtain t-values for the regression coefficients (rainfall and C_j) and the AIC can be used to find the optimal model and also the optimal value of j. The AIC is a model selection tool, for which the lower the AIC value the better (Chapter 17). Table 36.3 shows the AIC values for various DFA models. The model with no explanatory variables has the highest AIC, indicating that it is the worst model. We also considered models with only rainfall (AIC = 754.08) and only the C_j variables, but these were also sub-optimal models. Models that contained both rainfall and C_j were better.

Figure 36.6 shows the estimated trends for some of these models. The upper right panel (labelled 'none') is the trend obtained by the model with rainfall but no C_j value used as explanatory variables. The trend has the shape of a step function. The models with rainfall and C_j where j is between 1966 and 1974 all have similar AIC values and their trends are similar in shape. At j = 1975, the AIC drops and the trend shape changes.

So, what is the DFA doing? To understand this, it might be useful to compare results with that of the MAFA. The MAFA produced two common trends, the first one showing stability, followed by a period of increase, and followed by stability again. The function C_j in the DFA model with one common trend and rainfall is just picking up either the transition at the lower part of the step function or that at the upper part depending on the value of j. The next question is, then, which model is the best? From a statistical point of view, it is the model with j = 1979 because it has the lowest AIC. The fact that the model with C_{1979} is better than with C_{1975} indicates that the difference between the second period of stability and the other years is more important for explaining the overall pattern in the data than the difference between the first period of stability and the rest. A further model improvement might be obtained if we use two dummy variables simultaneously in the model, C_{1975} and C_{1979} as it would effectively replace the first MAFA axis by two dummy variables. For now, however, we discuss the results obtained by the DFA model with C_{1975} because the timing of the start of the recovery is of more interest to conservation biologists wanting to understand the mechanisms that underlie the reversal of population declines.

Estimated regression coefficients for this model are presented in Table 36.4. In the DFA model, rainfall has a negative and significant effect on several time series. Why is this the case in the DFA when rainfall was not important in the MAFA? Well, let us look in detail at the time series for which it is actually important. Based on the sign, magnitude and significance levels, rainfall is important for both stilt and coot on both Maui and Kauai_Niihau. But two of these time series were not used in the MAFA and the other two time series were not strongly related to the first two MAFA axes!

Table 36.3. DFA results using all eight time series. Only DFA models with one common trend were considered.

Model	Explanatory Variable	AIC	Model	Explanatory Variable	AIC
1	None	770.74	11	Rainfall & C_{1972}	751.70
2	Rainfall	754.08	12	Rainfall & C_{1973}	749.91
3	C_{1966}	768.80	13	Rainfall & C_{1974}	748.68
4	C_{1969}	768.74	14	Rainfall & C_{1975}	730.68
5	Rainfall & C_{1966}	744.52	15	Rainfall & C_{1976}	722.83
6	Rainfall & C_{1967}	751.50	16	Rainfall & C_{1977}	720.61
7	Rainfall & C_{1968}	753.47	17	Rainfall & C_{1978}	725.17
8	Rainfall & C_{1969}	748.58	18	Rainfall & C_{1979}	717.49
9	Rainfall & C_{1970}	753.47	19	Rainfall & C_{1980}	724.29
10	Rainfall & C_{1971}	753.02			

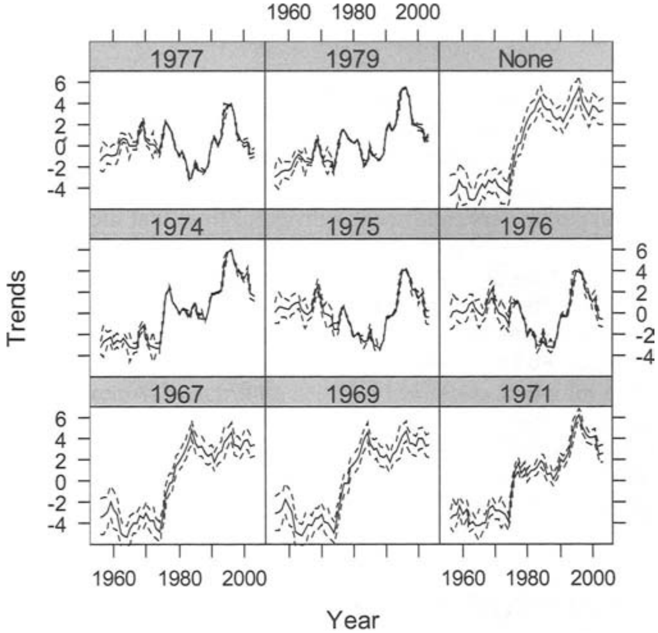

Figure 36.6. Common trend obtained by a DFA model using one common trend, rainfall, and C_j as explanatory variables. The y-axis represents the partial fit of the trend (Chapter 17). It shows the standardised population size. Dotted lines are 95% confidence bands for the trend. The numbers above the panels refer to different values for j.

The next question is: Why is C_{1975} in the model? The time series for which C_{1975} is important were also strongly related to the first MAFA axis. Hence, C_{1975} takes over part of the role of the first MAFA axis.

Besides the effects of the explanatory variables rainfall and C_{1975}, there is a certain amount of common information in the time series. This is captured by the common trend (see the panel for j = 1975 in Figure 36.6). Just as in MAFA one can ask the question: Which of the original time series show this pattern? Again, correlations between the trend and the original time series can be used. These can either be presented in a graph or tabulated (Table 36.4). In this case, only three time series had correlations that were not close to zero and, hence, contained a pattern similar to the DFA trend. These were Coot_Maui and both moorhen series. All other time series are mainly related to the explanatory variables. Notice that Moorhen_Oahu shows a negative correlation with the trend, whereas the other two time series show positive correlations. Note also that the correlations are calculated for available data only (missing values were not substituted by averages).

One can indeed recognise a bump or a dip (moorhen on Oahu) during the mid-1990s in these three time series; see also Figure 36.2.

Table 36.4. Estimated regression parameters, standard errors and t-values obtained by a DFA model with one common trend, rainfall, and C_{1975} as explanatory variables and a non-diagonal error covariance matrix. The last column contains the cross-correlations between the trend and the original time series. Bold face indicates regression parameters significantly different from 0 at the 5% level.

	Rainfall			C_{1975}			Trend
	Estimate	SE	t-value	Estimate	SE	t-value	
Stilt_Oahu	0.00	0.02	0.21	**1.00**	**0.27**	**3.76**	−0.004
Stilt_Maui	**−0.06**	**0.01**	**−4.56**	**0.95**	**0.22**	**4.29**	−0.079
Stilt_Kauai_Niihau	**−0.06**	**0.02**	**−3.21**	0.00	1.60	0.00	0.139
Coot_Oahu	−0.02	0.02	−1.50	**0.87**	**0.26**	**3.30**	−0.162
Coot_Maui	**−0.06**	**0.02**	**−4.12**	0.14	0.27	0.52	**0.300**
Coot_Kauai_Niihau	**−0.05**	**0.02**	**−2.23**	1.20	1.60	0.75	−0.002
Moorhen_Oahu	0.00	0.01	−0.01	**1.26**	**0.25**	**4.80**	**−0.422**
Moorhen_Kauai	0.01	0.01	1.33	**1.82**	**0.35**	**5.13**	**0.431**

36.5 Additive mixed modelling

The results obtained in the previous section indicate that there is at least one major pattern over time underlying all eight time series. In this section, we show how smoothing techniques can be used to gain similar information. To apply the methods in this section, the data were re-organised such that all eight time series were placed under each other in the same spreadsheet. Hence, there is now one long column of data. A new nominal variable 'ID' was created with values 1 to 8, identifying the time series. The columns with rainfall and year were copied eight times as well. As a result, the re-organised dataset contained only four columns. In order to make the additive modelling (AM) results directly comparable with the other methods that we have used, we first standardised the time series just as we did for the DFA and MAFA analyses. This step is not strictly necessarily as we will discuss later. We can start the AM analysis in different ways. If we had no prior knowledge from the DFA or MAFA, we could start the analysis by applying the AM to each individual time series:

$$\text{Birds}_t = \text{constant} + f_1(\text{Rainfall}_t) + f_2(\text{Time}_t) + \varepsilon_t$$

where $f_1()$ and $f_2()$ are smoothing functions and ε_t is independently normally distributed noise. The index t refers to year. The problem with this model is that the data form time series, and therefore, the independence assumption on ε_t may be incorrect (Chapters 16, 26). Additive *mixed* modelling (AMM) allows one to include an auto-correlation structure, and the simplest choice is an AR-1 structure:

$$\varepsilon_t = \rho\varepsilon_{t-1} + \gamma_t$$

where ε_t is now allowed to be correlated with noise from previous years and γ_t is independently normally distributed noise. Other auto-correlation structures were discussed in Chapters 16 and 26. It is also possible to apply the AMM on the combined data matrix:

$$\text{Birds}_{ti} = \text{constant} + f_{1i}(\text{Rainfall}_{ti}) + f_{2i}\,(\text{Time}_{ti}) + \varepsilon_{ti} \tag{36.1}$$

$$\text{where} \quad \varepsilon_{ti} = \rho\varepsilon_{t\text{-}1,\,i} + \gamma_{ti}$$

The index i refers to time series and runs from 1 to 8. The constant does not contain an index i because the time series were standardised and therefore have a mean of zero. Technically, this model is fitted using the 'by' command in the gamm function in the R library mgcv (Wood 2006). The 'by' command allows one to include an interaction between smoothers and nominal variables. The model in equation (36.1) assumes that the rainfall-bird relationship and the long-term trend are different for each time series. This means that a lot of precious degrees of freedom have to be estimated. Yet, another possible AMM, assuming that the rainfall effect and trend is the same for all time series, takes the form:

$$\text{Birds}_{ti} = \text{constant} + f_1(\text{Rainfall}_{ti}) + f_2(\text{Time}_{ti}) + \varepsilon_{ti} \tag{36.2}$$

$$\text{where} \quad \varepsilon_{ti} = \rho\varepsilon_{t\text{-}1,\,i} + \gamma_{ti}$$

Note that we have omitted the index i from the smoothers f_1 and f_2 in equation (36.1). The model in (36.2) is nested within the model in (36.1), and therefore we can use a likelihood ratio test to see which one is better. Results indicated that there was no evidence (p = 0.84) that the model in (36.1) is better than the AMM in equation (36.2). This does not necessarily mean that there is only one rainfall pattern, and one long-term trend, but it does mean that a model with two smoothers is better than the model with 16 smoothers.

One can also raise the question of whether the three species have the same trend, and such a model is given by

$$\text{Birds}_{ti} = \text{constant} + f_{1,species}(\text{Rainfall}_{ti}) + f_{2,species}(\text{Time}_{ti}) + \varepsilon_{ti} \tag{36.3}$$

$$\text{where} \quad \varepsilon_{ti} = \rho\varepsilon_{t\text{-}1,i} + \gamma_{ti}$$

In this model, each species is allowed to have a different rainfall effect and a different trend. Because the model in equation (36.1) is nested within (36.3) the likelihood ratio test was applied and there was no evidence to prefer the more complicated model with three trends (p = 0.17).

Obviously, one can also ask whether the three islands have the same trend, and such a model is given by

$$\text{Birds}_{ti} = \text{constant} + f_{1,island}(\text{Rainfall}_{ti}) + f_{2,island}\,(\text{Time}_{ti}) + \varepsilon_{ti} \tag{36.4}$$

$$\text{where} \quad \varepsilon_{ti} = \rho\varepsilon_{t\text{-}1,i} + \gamma_{ti}$$

Each island is now allowed to have a different rainfall effect and a different trend. Again, model (36.1) is nested within model (36.4) and the likelihood ratio test gave a p-value of 0.04. Note, however, that these p-values should be interpreted with care due to its approximate nature. We have decided to present the results of the model in equation (36.4). The numerical output is given by

	edf	*F*-statistic	*p*-value
s(Rainfall):S1	2.46	2.47	0.010
s(Rainfall):S2	1.00	13.31	<0.001
s(Rainfall):S3	3.65	4.87	<0.001
s(Year):S1	4.64	4.95	<0.001
s(Year):S2	1.00	20.59	<0.001
s(Year):S3	1.00	0.29	0.59

These are approximate significance values of smooth terms. The notation s(Rainfall):S1 refers to the smoother for rainfall at island 1, whereas (sYear):S1 gives the smoother for the temporal trend on the same island (1 = Oahu, 2 = Maui, 3 = Kauai_Niihau). Except for the trend for Kauai_Niihau, all smoothers are significant. An estimated degree of freedom (edf) of 1 means that the relationship is modelled by a straight line. Hence, the rainfall effect for Maui and the long-term trend at Maui are linear, and all other smoothers are non-linear. The auto-correlation parameter ρ was 0.39. The smoothing curves are plotted in Figure 36.7. The shape of these curves indicates that on Oahu, there was a strong trend over time with an increase from the mid-1970s until the mid-1980s. Layered on top of this pattern, there was a general decrease in bird numbers with rainfall; the more rainfall, the lower the bird numbers. On Maui, there is a decrease in bird numbers with rainfall, but an increasing trend over time. Finally, on Kauii_Niihau, there is a non-linear rainfall effect, with numbers generally decreasing with more rain, but increasing again when there is very high rainfall.

The model can be extended in various ways: (i) We can add dummy variables that measure the effect of conservation protection in the same way as we did for the DFA, and (ii) instead of standardising the time series, we can add a nominal variable ID with values 1 to 8 (allowing for different mean values per time series) and use eight different variances (allowing for different spread per time series). It is also possible to use different spreads (variances) per species or per island. Avoiding standardisation is useful if the model will be used for prediction.

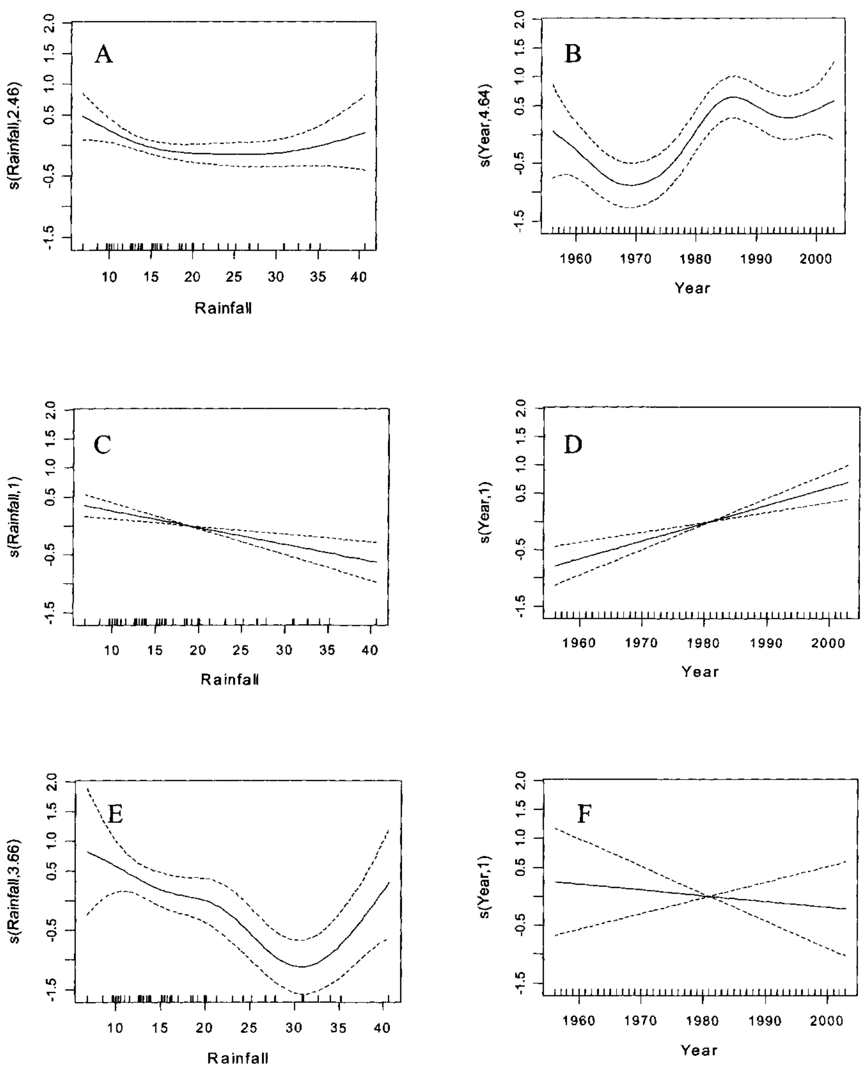

Figure 36.7. Smoothing curves obtained by additive mixed modelling for Year and Rainfall using an additive mixed model. The dotted lines represent the 95% confidence bands. The vertical axis represents the contribution of the smoother to the fitted values. Cubic splines were used. Panels A and B refer to Oahu, C and D to Maui and E and F to Kauai_Niihau.

36.6 Sudden breakpoints

Both the MAFA and DFA indicated that there was a sudden increase in bird abundance in a majority of the time series. To confirm this interpretation, we applied chronological clustering (Chapter 19). This technique calculates a distance matrix between the years. Each element in the matrix represents the association between two years (as measured over all eight variables). We decided to use the Whittaker index (Bell and Legendre 1987; Legendre and Legendre 1998). The problem with simple clustering is that it could produce groups of non-sequential years, ignoring the temporal dependence between years, which is crucial in an analysis involving time series. Instead of applying ordinary clustering on the distance matrix, chronological clustering creates groups that are constrained to include only sequential years. In this method, two clustering parameters have to be set, alpha and the connectedness. Following the recommendations in Legendre et al. (1985), we kept the connectedness fixed at 0.5 and presented results for different values of alpha. Small values of alpha mean that we end up with rather conservative groupings. This can also be seen as a bird's eye overview: Only the major breakpoints are obtained. Increasing the value of alpha gives greater resolution and more groups. Results in Figure 36.8 indicate that there is a clear breakpoint in 1975, which corresponds with our earlier analyses. Other measures of similarity gave identical breakpoints. Two time series (those for Kauai_Niihau) started in 1976 and could have influenced the location of the breakpoint. We repeated the analysis without these time series, however, and a nearly identical picture was obtained except that the main breakpoint shifted slightly to 1978.

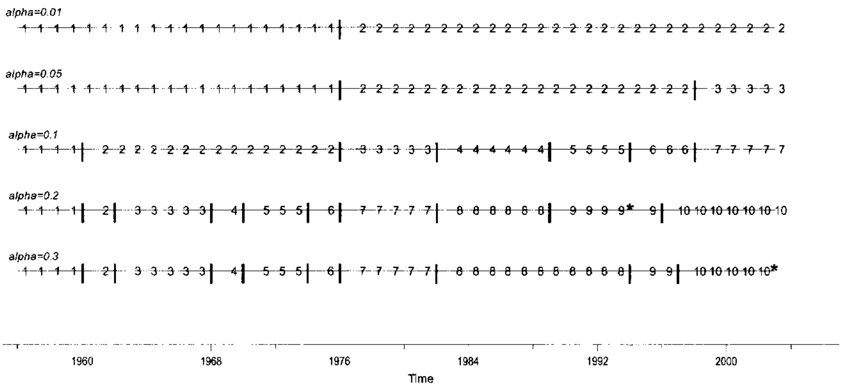

Figure 36.8. Results of chronological clustering applied on all eight bird time series. Vertical bars identify breakpoints for different alpha levels.

36.7 Discussion

The main underlying questions for this study were (i) whether the eight time series have similar trends or patterns over time, and (ii) whether there are any sudden changes. To answer these questions, various time series techniques, discussed earlier in this book, were applied. The time series dataset used here is not large, although it is relatively long for an annual monitoring program in the field of ecology. Nonetheless, it turned out that it was not an 'easy' dataset from a statistical point of view. No single technique was conclusive; however, the case study demonstrates how results from all these techniques combined can give insight into what is going on. Moreover, results from both optimal and sub-optimal models were informative.

The analyses showed that there was a major increase starting in the mid-1970s that continued up to the end of the 1970s. Because the start of two of the time series coincided, more or less, with the beginning of this period of increase, it is reasonable to wonder if the change is an artefact caused by the pattern of data collection. The basic result, however, was the same for both analyses that include these two time series and for those that do not.

We used DFA to identify when the change in population sizes took place. Using dummy variables, the year 1975 was identified as the year in which the increase started. It is unlikely that the two time series starting in 1976 are responsible for the shift as all time series were normalised in the DFA. Besides, the MAFA showed the same breakpoint and it did not use these two time series. The additive mixed modelling indicated an increase from the early 1970 up to the mid-1980s on Oahu.

Chronological clustering also shows the same breakpoint in the mid-1970s, but it does not indicate why the breakpoint occurred. MAFA and DFA gave more details on reasons for the pattern.

To conclude, the analysis suggests that there is a rainfall effect primarily associated with the stilt and coot time series from Maui and Kauai, providing additional support for the idea that fewer birds are recorded in years when there is a lot of rain. On top of this, several time series were affected by a major population increase in the 1970s, which occurred not long after endangered species legislation was introduced to protect the birds under study. Finally, three series contained an additional common pattern: coot on Maui and moorhen on Kauai showed an overall increase, whereas moorhen on Oahu decreased.

Acknowledgements

We thank the Hawaii Division of Forestry and Wildlife, and the Hawaii Natural Heritage Program for access to the biannual waterbird survey data, and the Hawaii Division of Forestry and Wildlife for their long-term coordination of these surveys.