Time series analysis of Holly Blue records in Yorkshire – uncovering a parasitoid

David R. R. Smith Department of Psychology, University of Hull
Email: davidsmith.butterflies@gmail.com

Introduction

The recorded numbers of the butterfly Holly Blue *Celastrina argiolus* have been shown to cycle up and down every 4-6 years and this has been attributed to the caterpillar population being parasitised by the larvae of the ichneumon *Listrodomus nycthemerus* (Revels, 1994, 2006). Revels provided convincing evidence of the host-parasitoid relationship between the butterfly Holly Blue and the parasitic ichneumon *L. nycthemerus* by collecting Holly Blue caterpillars and rearing them through to chrysalis. Over a period of 15 years he showed that the proportion of chrysalises that produced either Holly Blue butterflies or *L. nycthemerus* fluctuated; at the peak of the cycle almost all chrysalises hatched into a Holly Blue adult butterfly and at the trough of the cycle almost all chrysalises (99%) produced an adult *L. nycthemerus*. The cycle from peak to trough took 4-6 years to repeat with parasitic wasp numbers lagging behind butterfly numbers by one or two generations. Though Holly Blue young caterpillars are attacked by the tiny braconid wasp *Cotesia inducta* it is *L. nycthemerus* which appears to largely drive the cycle by attacking the older caterpillars (Revels, 1994, 2006). *L. nycthemerus* has no other known host than the Holly Blue caterpillar while *C. inducta* attacks a number of Lycaenidae including White-letter Hairstreak *Satyrium w-album* (Shaw *et al*., 2009).

That Yorkshire's Holly Blue populations showed booms and busts has been remarked on before (see Frost, 2005) who described a pattern repeating roughly every 5 years from 1990 to 2002. We now have data from a longer sequence of years so the potential for detecting convincing patterns in our data is stronger. More recently, in an article reviewing lepidopteral colonisers of Yorkshire an analysis of Holly Blue plotted the number of tetrads from which it was recorded between 1995 and 2017 (Smith, 2018, Fig. 14, replotted below as Fig. 1). The data in Figure 1 hints at some sort of periodic cycle that reduces in amplitude as the recording year approaches the present; it is suggestive of a damped oscillation such as happens when your car's shock absorbers minimise vertical movement when you pass over a pothole in the road. Interestingly, *L. nycthemerus* has never been recorded in Yorkshire (W.A.Ely, pers. comm.)[1].

The rationale for this article is to first present a better measure of Holly Blue population numbers – tetrad numbers are a measure of distribution that only *correlate* with abundance – so we shall use a true (sampled) population measure (number of individual butterflies recorded). Second, any periodic cycle appearing in our abundance measure of Holly Blue will be better revealed if we perform some simple analysis of the time series which will help us to strip out trends in population numbers which might obscure periodicity. Third, a mathematical treatment will allow the quantification of how strong any periodicity (signal) we observe in the data is relative to all other frequency components (the noise). The over-riding rationale for this article is that the uncovering of periodic cycles would provide indirect evidence for the long established presence of *L. nycthemerus* in Yorkshire.

============== INSERT FIGURE 1 ABOUT HERE ==============

---

[1] This was the case until during the writing of this paper a report of the first ever sighting of *L. nycthemerus* was received! Please see this issue XXXX.

Method

The Butterfly Conservation Yorkshire (BCY) database was searched for Holly Blue records for the period 1995 to 2017 from the five Watsonian vice-counties (VC61-VC65) comprising the county of Yorkshire for recording purposes. From these records the number of Holly Blue butterflies seen across all of Yorkshire for each year between 1996 and 2017 was calculated[2].The annual count is across all broods (Yorkshire Holly Blue populations have two broods with a partial third brood in some years). This provides the Holly Blue time series Y_i: a regularly sampled (=annual) series of successive data points that are temporally auto-correlated, e.g., the number of Holly Blues in a year is not independent of the previous year(s) Holly Blue numbers. The time series consists of members y_1, y_2, …, y_{22} where subscript 1 refers to the first entry (=1996), and subscript 22 is the last entry (=2017). The time series is assumed to be additive, i.e., the random fluctuations (noise) in the data are approximately constant over time.

Analysis of the Holly Blue time series took place within the *R* statistical environment (R version 3.4.4, R Core Team, 2018), using additional time series analysis packages where necessary (Ulrich, 2018). The mathematical treatment was performed with MATLAB and the Signal Processing Toolbox (Release 2007a, The Mathworks).

Results

Figure 2 is the Holly Blue time series, which shows the annual count of Holly Blue butterflies for each recording year across the whole of Yorkshire for 1996 to 2017. There is quite a convincing periodic pattern with a cycle of around 4 years which reduces in amplitude as we approach the present day. There is a suggestion that the period of the cycle lengthens towards the present; from 4 years between 1996—2004 to around 6 years between 2004—2017.

============== INSERT FIGURE 2 ABOUT HERE ==============

Before performing a mathematical analysis of the Holly Blue time series a simple analysis of the time series was conducted to justify whether it would be reasonable to perform a spectral analysis of the time series.

We can consider the Holly Blue time series as a mixture of three components: a long-term non-cyclic *trend* component which could be due to, for instance, the geographic spread of the Holly Blue as it reacts to changes in climate or perhaps simple drift in observer numbers; a *cyclic* component (confusingly often referred to as the 'seasonal' component in time-series analysis) which this article is principally interested in; and a *random* component (noise). For a reasonably brisk introduction to time-series analysis see Kendal and Ord (1990) or, for a slower-paced introduction, see the Open University book 'Time Series' (2006).

Figure 3 shows the Holly Blue time series Y_i decomposed into its trend y_i^T, cyclic y_i^C ('seasonal') and random y_i^E additive components. Thus the Holly Blue time series is taken to be defined as Y_i

---

[2] The number of Holly Blue reported in 1995 was extremely low (=23). 1995 provides the first extant records of Holly Blue in the BCY database and is doubly confounded by the small numbers of observers sending in records and the recent movement of Holly Blue northwards in latitude. For these reasons the 1995 count was discarded from the time series.

= $y_i^T + y_i^C + y_i^E$. The Holly Blue time series is shown in the top row of the figure ('observed'). The trend component (second row, Fig. 3) is calculated as a simple symmetrically-weighted moving five year mean: $(\sum_i^{i\in\{i,i\pm1,i\pm2\}} y_i)/5$, e.g., the 1998 (3rd) entry in the trend (*T*) component y_3^T of the Holly Blue time series is 906 (calculated as the mean of five years centred on 1998, $(\sum_i^{1:5} y_i)/5$ = (271 + 1560 + 1997 + 537 + 164)/5); the 1999 (4th) trend component entry is 933 calculated as $(\sum_i^{2:6} y_i)/5$, and so on. The y_i^T trend component is undefined for the first and last two years in the series (1996—1997 and 2016—2017) as the mean 'window' sums across years for which we do not have data (1994—1995 and 2018—2019). The cyclic component (third row, Fig. 3) is calculated by *assuming* a 5-year cycle (which is the best guess and between a 4-year and a 6-year cycle) and building a 'model' cycle by finding the average of the first, second, …, fifth elements of each cycle within the time series. This cyclic seasonal component y_i^C is then centred within the figure. The random component y_i^E (fourth row, Fig. 3) is the remainder when the trend and seasonal components are subtracted from the time series: $y_i^E = Y_i - y_i^T - y_i^C$.

The trend component y_i^T indicates that Holly Blue numbers have been declining since about 2004—2006. The cyclic component y_i^C is an averaged *model* of a five-year cycle in the data – the excursion around the ~0 baseline is about ±~500 – compared to a peak and trough excursion around the ~1065 baseline of ±~600 in the observed time series Y_i. The random component y_i^E has a mean around ~40 with a standard deviation of ±~450. Preliminary analysis achieved by a decomposition of the time series (see Fig. 3) indicates that hunting for periodicity in the data is warranted as the cyclic component is substantial in size relative to the trend and random components.

============= INSERT FIGURE 3 ABOUT HERE =============

The identification and quantification of the periodicity in the time series was achieved by performing a single spectrum (Fourier) analysis of the Holly Blue time series. The fundamental theorem of Fourier holds that any complex wave (or time series) is equivalent to an infinite number of sine and cosine component waves of various amplitudes, frequencies and phases added together (for a classic review see Bracewell (1965) and for a more applied review see James (1995)). Thus we can take any given time series and analyse that complex series into its constituent periodic frequencies; we are interested in whether that breakdown into constituent periodic frequencies reveals any especially strong frequency component. If there is a very strong frequency that dominates the other constituent frequencies then this would be strong evidence that there exists a cycle in our time series. Moreover, we should be able to quantify the relative strength of that dominant cycle against the other constituent frequencies.

There are two steps we need to do to prepare our data for single spectrum analysis. First we 'detrend' the series by subtracting y_i^T; the rationale for this is that slow changes in Holly Blue numbers due to the arrival and possibly decline of Holly Blue in Yorkshire, changes in observer numbers, etc., will only act to dominate and obscure other more meaningful patterns in our time series. Second, we need to subtract the global mean $\overline{(Y_i - y_i^T)}$ from the series because this represents a constant zero-frequency baseline that will overwhelm all other constituent frequencies.

Therefore the time series, $Y_i - y_i^T - \overline{(Y_i - y_i^T)}$, was passed through a single spectrum analysis. The series passed for single spectrum analysis consisted of 18 points because the first and last two entries of the original time series were lost as y_i^T is calculated as a five-year moving mean which is undefined for the first and last two entries in y_i^T. Figure 4 shows the Periodogram Power Spectral Density Estimate (PPSDE) – this figure shows how much power there is as a function of frequency (Hz) in the Holly Blue time series. We can see that there are three peaks in the PPSDE. We can discount the peak around 9Hz as this arises from an 'edge effect' directly linked to the length of the time series; we would always expect a peak at $\frac{n}{2}$ Hz where *n* is the length of the time series. We are left with two peaks at 4Hz and 6Hz. The dominant (strongest power) peak is at 4Hz with another smaller peak at 6Hz; the 6Hz peak is ~6dB weaker than the 4Hz peak. Because the y-axis is in power dB this means that the 4Hz cycle is 4 times stronger than the 6Hz cycle. Both the 4 and 6Hz peaks are considerably stronger than other component frequencies that are not adjacent to the 4 and 6Hz centre frequencies[3]; for instance, the 2Hz component in the PPSDE is ~20dB weaker than the 4 and 6Hz components meaning that the 2Hz cycle component is ~100 times as weak as the 4 and 6Hz frequency components.

============== INSERT FIGURE 4 ABOUT HERE ==============

Discussion

The analysis of the time series abundance data for Yorkshire's Holly Blue show clear evidence of a 4-year cycle and, to a lesser extent, a 6-year cycle. Both cycle frequencies stand out against other non-adjacent frequencies (such as 2Hz) by a factor of 100. This can be taken as very strong indirect evidence of a clear 4-year and 6-year cycle of boom and bust in Holly Blues numbers – the obvious candidate would be *L. nycthemerus,* which is known to drive such host-parasitoid cycles in Holly Blue populations (Revels, 1994, 2006). The breaking news of the discovery of *L. nycthemerus* in Yorkshire this year (see this issue XXXX) gives direct evidence of the presence of this parasitic wasp. The analysis of the time series abundance data for Holly Blue indicates that the parasitic wasp has long been established in Yorkshire. *L. nycthemerus* may have travelled with, or closely followed, the Holly Blue northwards into Yorkshire around the mid 1990s.

It is interesting to note that the cycle appears to have lengthened from 4 years (1996—2004) to 6 years in the last half of the time series (from about 2004), and this coincides with the beginning of a long-term decline in Holly Blue numbers (see second row, Fig. 2). Lengthening time periods in predator-prey cycles can be associated with a positive nonlinear trend (John Bowers, pers. comm.) but here we have a lengthening of cycle period with a negative quasi-linear trend. One interpretation of the lengthening period of the cycle is that as Holly Blue caterpillars become rarer it is harder for *L. nycthemerus* to find hosts to parasitize, which leads to a slower build up of wasp numbers and consequently a longer cycle period.

It should be acknowledged that recording variation applies at different time scales. An attempt to remove long-term trends (due to possible factors such as changes in the geographic spread of Holly

---

[3] Note that the 'broadness' of the peaks around 4 and 6 Hz, with significant power around the peak frequency, is an inevitable consequence of having a short time series which induces 'leakage' into adjacent frequencies. A longer time series would allow for sharper peaks because the sampling resolution of a series increases with length.

Blue or changes in recorder numbers) was made by subtracting the five-year moving average from the time series. However, there was no compensation for recording variation at smaller time scales which might affect the seasonality (cycle) and noise components of the time series. For this reason, fluctuations in population levels in butterflies should ideally be studied using transect data where recorder effort is held constant. However, Holly Blue has the strongest correlation (r^2 = 0.85) between transect and casual abundance data of all UK butterfly species (Mason *et al*., 2017), so this should not be much of an issue for this study. Given that we see a clear cyclic component in data collected from a substantial part of Yorkshire (for instance, Holly Blue was present in 887 of 3720 recorded tetrads (=23.8%) between 2004-2017, Smith, 2018) we can be reasonably confident that *L. nycthemerus* is established across large parts of Yorkshire.

Acknowledgements

I am grateful for the Butterfly Monitoring Scheme (UK BMS) and Butterfly Conservation and their volunteer recorders for the butterfly record~~er~~s from Yorkshire upon which this research is based. This article has benefitted from the helpful comments of a reviewer.

References

Bracewell, R.N. (1965). *The Fourier Transform and its Applications*. McGraw-Hill, New York.

Frost, H. M. (2005). *The Butterflies of Yorkshire*. Butterfly Conservation. Holly Blue entry written by Frost, H. M. and Parkinson, D. (p.141).

James, J.F. (1995). *A student's guide to Fourier transforms with applications in physics and engineering*. Cambridge University Press.

Kendall, M. & Ord, J.K. (1990). *Time series* (3rd ed.). Edward Arnold.

MATLAB (2007). The Mathworks, Inc, Natick, Massachusetts, USA.

Mason, S.C., Hill, J.K., Thomas, C.D., Powney, G.D., Fox, R., Brereton, T. & Oliver, T. H. (2017). Population variability in species can be deduced from opportunistic citizen science records: a case study using British butterflies. *Insect Conservation and Diversity 11*: 131-142.

Open University Course Team. (2006) Time Series. Open University Press. Attached to the OU Practical Modern Statistics course (order code M249/CB2).

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Revels, R. (1994). The rise and fall of the Holly Blue. *British Wildlife 4*: 236-9.

Revels, R. (2006). More on the rise and fall of the Holly Blue. *British Wildlife 17*: 419-24.

Shaw, M.R., Stefanescu, C. & van Nouhuys, S. (2009). Parasitoids of European butterflies. *In* Settele, J., Shreeve, T., Konvicka, M. & van Dyck, H. (Eds) *Ecology of butterflies in Europe*. Cambridge University Press, 130–156.

Smith, D.R.R. (2018). How have recent lepidopteral colonisers fared in Yorkshire? *The Naturalist 143*: 107-129.

Ulrich, J. (2018). TTR: Technical Trading Rules. R package version 0.23-4. https://CRAN.R-project.org/package=TTR

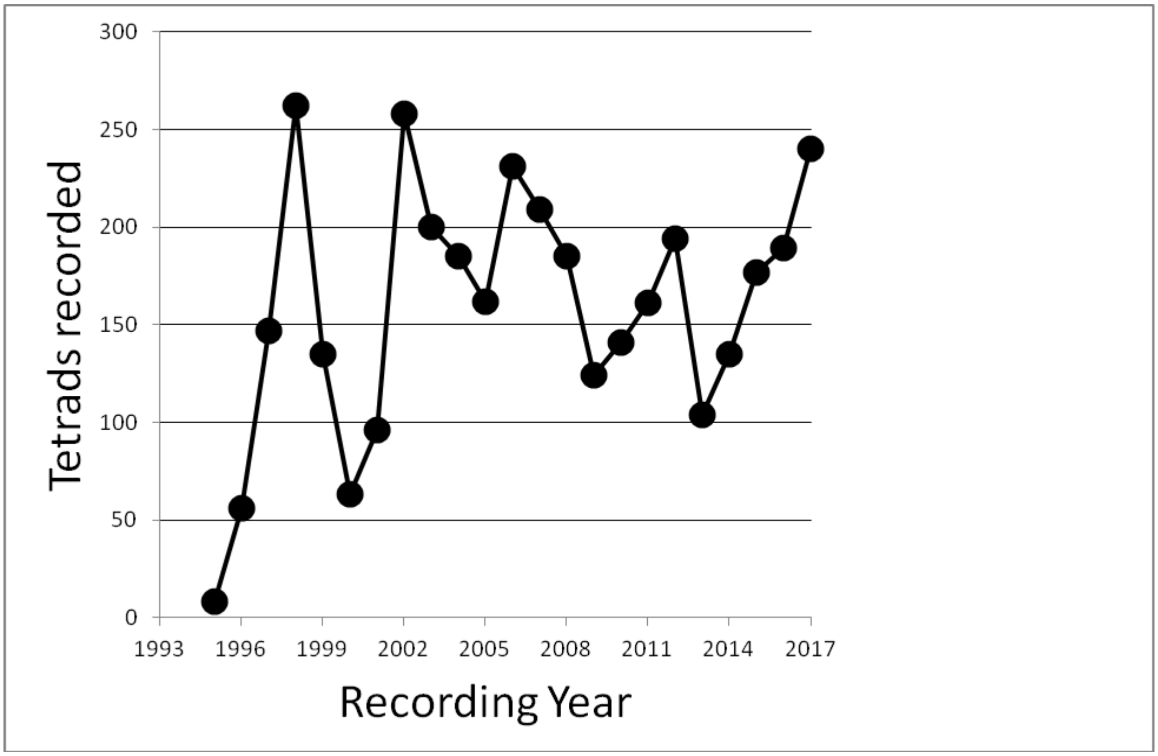


Figure 1. Data replotted from Figure 14 of Smith (2018). Number of tetrads in Yorkshire where Holly Blue was seen as a function of recording year. The data points are uncorrected raw tetrad counts per year, e.g., there has been no correction factor applied for the tendency for recorded tetrads to increase each year due to greater numbers of recorders involved in butterfly recording. These data points are equivalent to the open circle symbols in Figure 14 of Smith (2018).

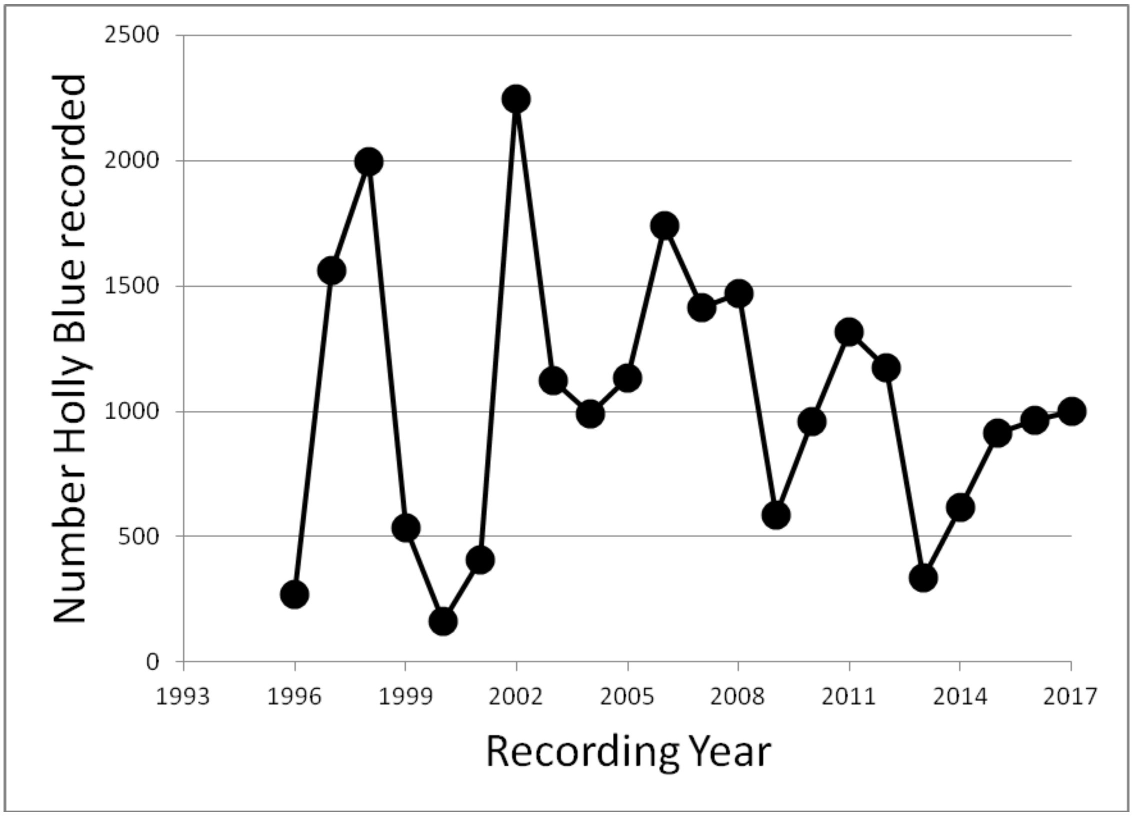
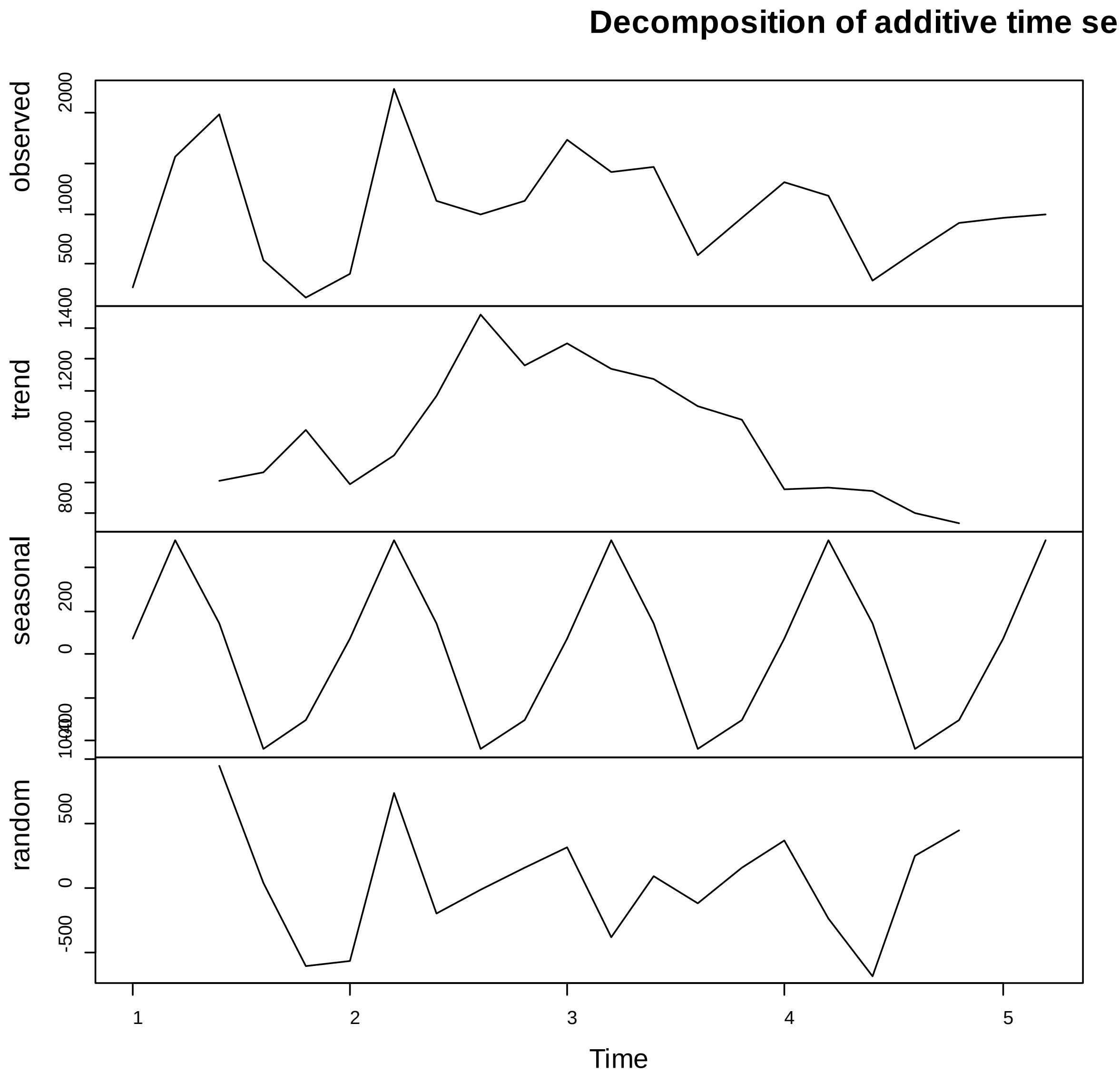


Figure 2. Annual counts of individual Holly Blue butterflies across the whole of Yorkshire as a function of recording year.



Decomposition of additive time ser
observed
2000
1000
500
1400
trend
1200
1000
800
seasonal
200
0
random
500
0
-500
1
2
3
4
5
Time

Figure 3. The decomposition of the Holly Blue time series (top row) into three components: a long-term trend (second row); a model cyclic seasonal component (third row); and a random component (fourth row). The x-axis is marked in cycles where the first cycle (‘1’) starts at 1996 and the second cyle (‘2’) starts at 2001, and so on.

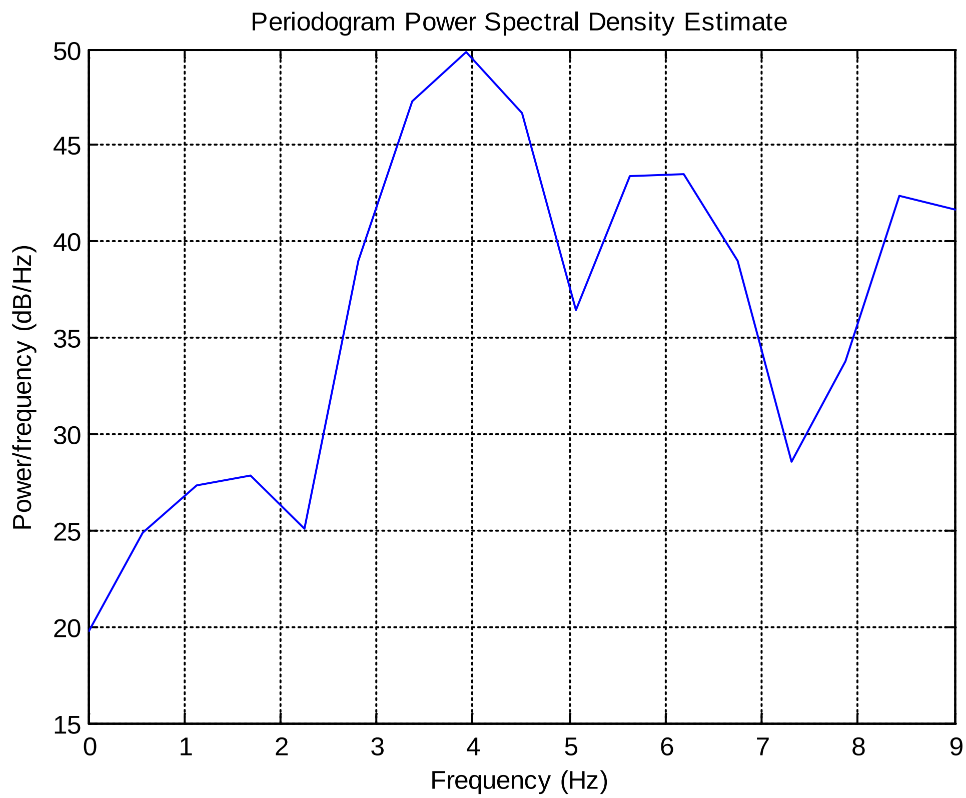


Figure 4. The Periodogram Power Spectral Density Estimate. The power of each frequency component in the single spectrum analysis of the Holly Blue time series was extracted using a nonparametric periodogram with a Hamming window. The x-axis is marked in frequencies measured in Hz, such that 4Hz is equal to a 4-year cycle, and so on.