Research Update

Preliminary Analysis of Fringilla Insect Data

Robert Wright

16th March 2022

Supervisor: Gerard Talavera

As part of my internship at the *Botanical Institute of Barcelona*, I have been performing an exploratory analysis on (what I refer to as) the *Fringilla insect dataset*. As far as I am concerned, the data and method of acquisition have been first presented in a paper by Shapoval and Buczyński [2012]. The following report is designed to give an update on the progress of analysis as well as highlight some questions (in section 3) that came up during the work. All project files can be found on github.

Contents

1. Data Exploration **2**
1.1. Missing Values 3

2. First Appearance **4**
2.1. Temperature Records 6

3. Open Questions **7**

A. Appendix **8**

References **10**

1. Data Exploration

The first step in understanding the insect data is to investigate the yearly abundance. Figure 1 reveals that the data (all traps combined) contains three dominating species: *Vanessa atalanta*, *Inachis io* and *Vanessa cardui*. While *V. atalanta* is recorded consistently in recent years, *V. cardui* and *I. io* show high variation. Most prominently, *I. io* shows pronounced outbreaks in 1985, 2002 and 2011.

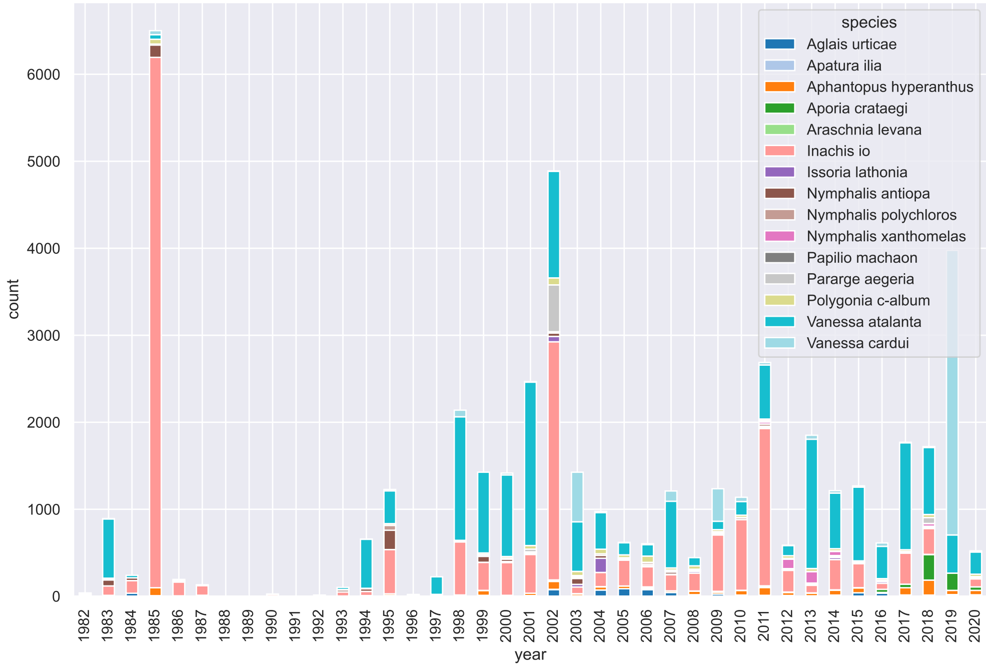

Figure 1: The total count per year split into single species.

1.1. Missing Values

The dataset has a high number of missing values in early years, with a considerable 'improvement' from 1998 onwards. Contrary to expectations, the number of missing values in temperature and insect count is not equal. Apparently there have been days in which insect count was monitored, but temperature was not recorded, and vice versa. Consequently, in order to yield more consistent results, the data analysis is constrained to the time period from 1998 to 2020.

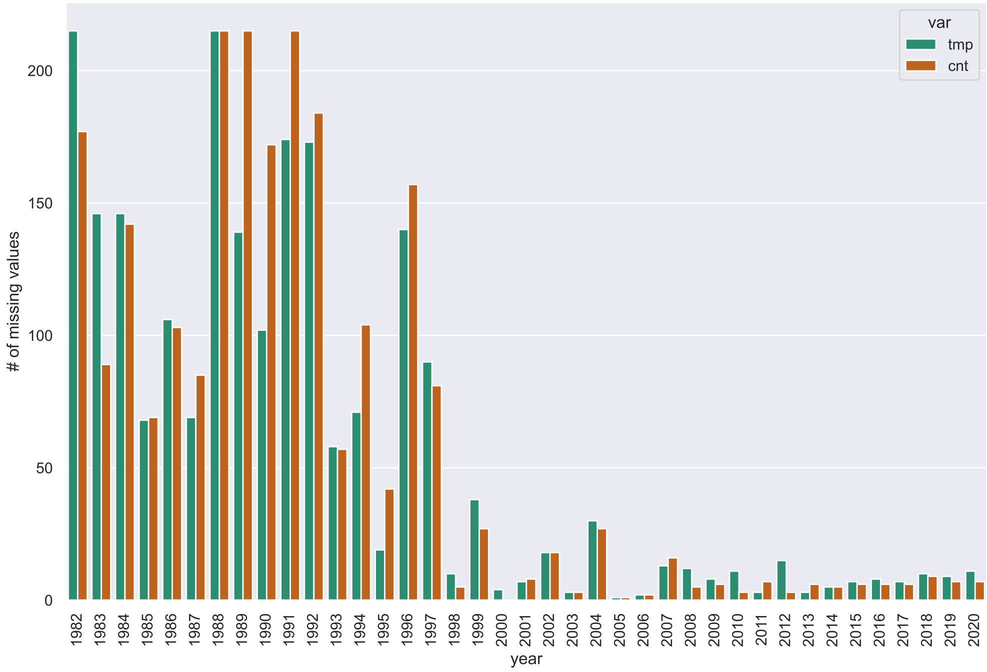

Figure 2: Number of missing values in temperature (*tmp*; green) and count (*cnt*, orange) per year.

2. First Appearance

Continuing an approach of Roy and Sparks [2000], in the following, the timing of first (and last) appearance per year is analysed. Here, the term 'appearance' refers to the earliest (and latest) insect count in the recordings of a single year. In the previously mentioned publication, the authors used data recordings from the *British Butterfly Monitoring Scheme* for the years 1976 to 1998 collected at over 100 sites. Among other things, they found a significant temporal shift to earlier mean first appearances in 13 out of 35 species analysed.

In order to identify linear trends in the Fringilla dataset, a least-squares regression is applied on first appearance with year as the explanatory variable. One of the fitted models is exemplarily plotted in fig. 3. The statistical parameters of all regressions are summarized in table 1 of appendix A. It turns out that three species show a significant temporal trend

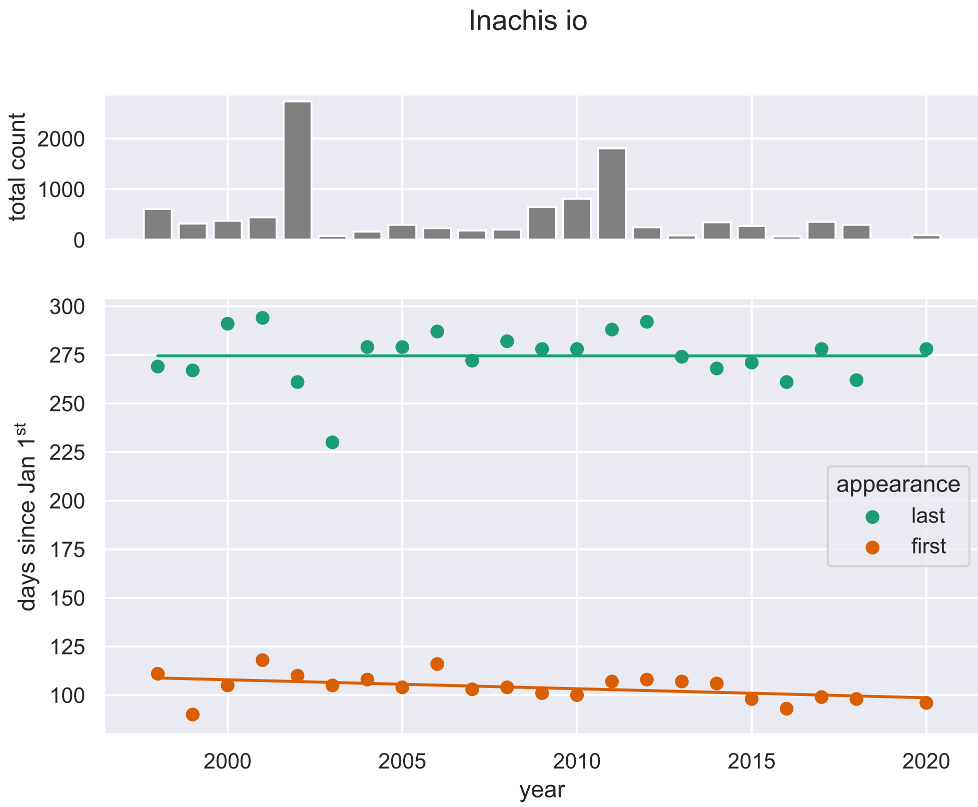

Figure 3: *Top:* Total count of *I. io* for the years 1998 to 2020, two outbreaks are visible. *Bottom:* Linear regression on first (orange) and last (green) appearance. Note that the y-axis is in days since Jan 1st of the respective year.

towards earlier appearance: *V. atalanta*, *I. io* and *P. c-album*. However, the data does not reveal a significant relationship for delayed last appearance in recent years for any species.

Proceeding, it would be of interest to relate climate variables to first appearance and investigate the underlying driving factors for the observed change in phenology. As the Fringilla data already includes temperature records, analysing the influence of temperature on first appearance immediately suggests itself.

2.1. Temperature Records

The Fringilla data includes the daily maximum temperature [Shapoval and Buczyński 2012], however, due to the large number of missing values in early years as well as during the winter months (no monitoring from November to April), comparing the records to and eventually using other sets of temperature data is advisable. A potential replacement dataset might be the *Climate Research Unit gridded Time Series* (CRU TS), covering all land domains (except Antarctica) on a 0.5° latitude by 0.5° longitude grid. This dataset is further discussed by Harris et al. [2020]. As fig. 4 illustrates, the temperature data displays contradictory behaviour: While—speaking of overall trends—the temperature recordings from Fringilla decrease, the annual mean temperature computed from CRU TS becomes warmer.

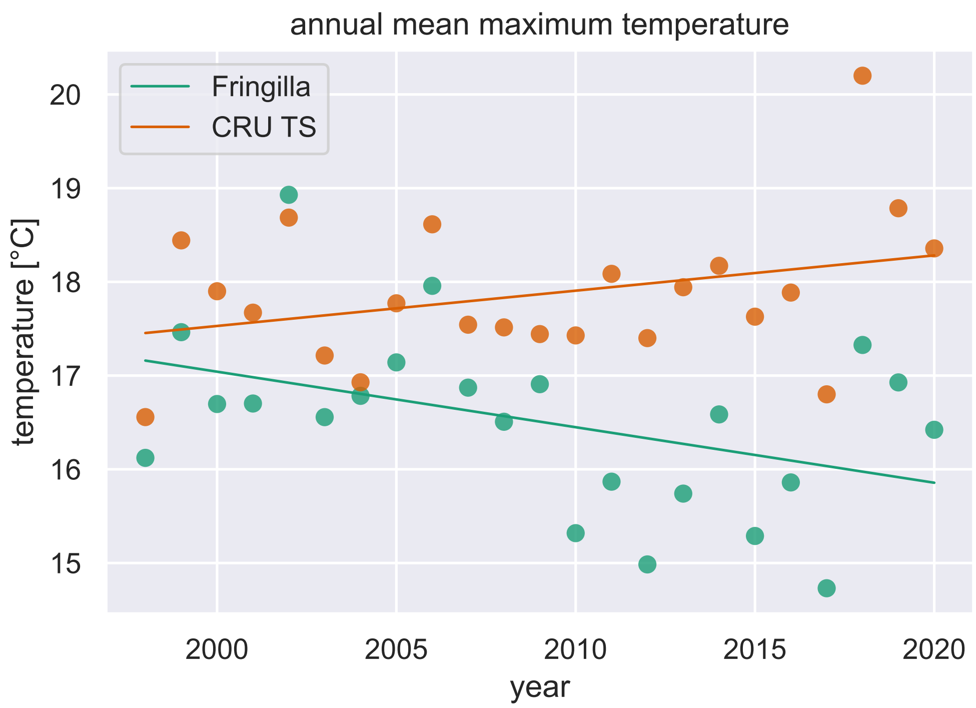

Figure 4: Comparison of the annual mean maximum temperature computed from the Fringilla dataset and CRU TS. Besides the data points, a linear fit is plotted. The annual mean is computed on the closed interval from 1st April to 31st October.

3. Open Questions

To summarize, here is a list of the questions which arose during the exploratory data analysis:

Data Exploration

- Does the combination of individual traps introduces any kind of bias to the total count? As fig. 5 in appendix A indicates, some traps were not used throughout the entire period of data acquisition.
- The outbreaks of local *I. io* are of interest to us. Were there any attempts you have heard of to find their cause(s)?
- The number of missing values decreases drastically after 1997 (see section 1.1). Did something fundamentally change in the way the traps were monitored?

First Appearance

- According to Fringilla dataset, the temperature is declining in recent years. However, this trend is not supported by CRU TS. Why? Did the method of recording temperature change? Was really the **maximum** value of daily temperature recorded as stated by Shapoval and Buczyński [2012]?
- Is CRU TS an appropriate alternative or do you use any other climate datasets?
- What are your thoughts or do you have some general remarks on the analysis of the first appearance? In case you know of any other publication, which might be of interest in this regard, please feel free to share it with us.

A. Appendix

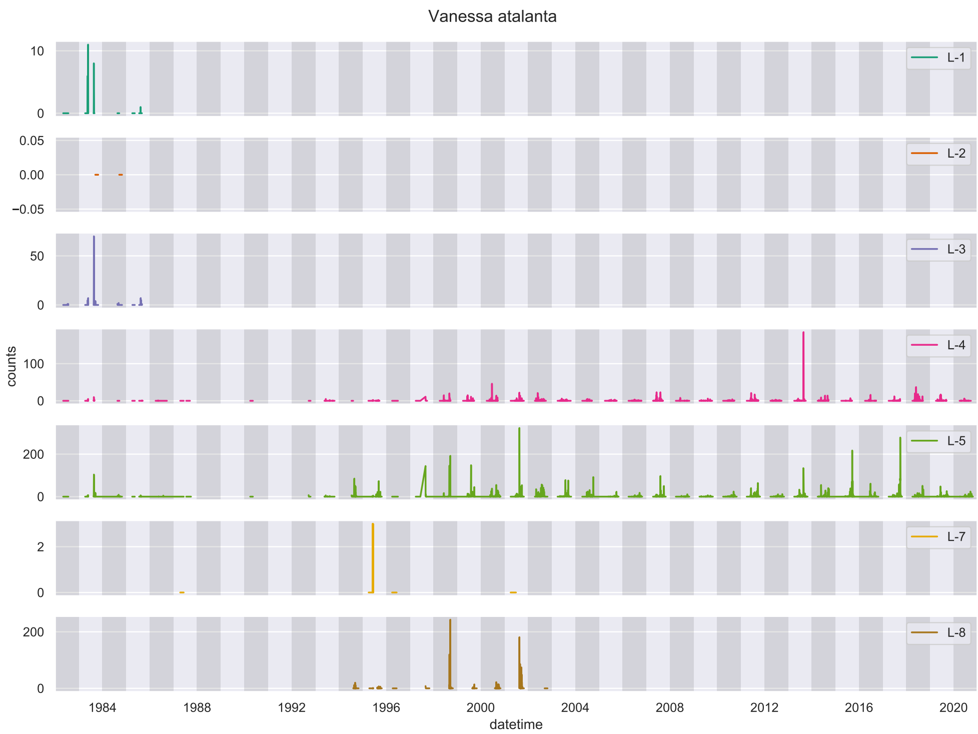

Figure 5: Daily count of insects caught in individual traps throughout the entire period of time. For illustrative purposes, *V. atalanta* with relatively high abundances was selected. The traps are labelled from *L-1* to *L-8*.

species	appearance	r^2 [%]	p-value	change (+10y)
Vanessa atalanta	first	24.0	0.018	−13.94
	last	1.3	0.602	−1.63
Vanessa cardui	first	6.1	0.28	−9.17
	last	1.3	0.618	−2.97
Inachis io	first	19.7	0.039	−4.62
	last	0.0	0.998	−0.01
Issoria lathonia	first	7.8	0.22	−18.86
	last	7.8	0.22	−7.04
Aglais urticae	first	12.9	0.092	−28.87
	last	0.4	0.762	1.68
Aporia crataegi	first	7.6	0.285	−3.5
	last	1.7	0.62	3.79
Apatura ilia	first	9.2	0.465	−5.14
	last	94.4	0.0	−41.92
Aphantopus hyperanthus	first	3.4	0.397	−2.68
	last	4.6	0.324	2.08
Araschnia levana	first	0.2	0.884	2.66
	last	1.4	0.686	4.49
Nymphalis antiopa	first	0.0	0.945	0.99
	last	2.4	0.487	−6.74
Nymphalis polychloros	first	12.3	0.11	−15.75
	last	1.3	0.612	−6.52
Nymphalis xanthomelas	first	14.0	0.208	−32.13
	last	0.2	0.896	3.09
Papilio machaon	first	13.0	0.276	−19.57
	last	15.6	0.23	9.84
Polygonia c-album	first	20.1	0.036	−13.21
	last	9.3	0.168	4.83
Pararge aegeria	first	0.4	0.81	4.11
	last	14.1	0.138	−13.05

Table 1: Statistical output of regression analysis of linear trend in first and last appearance. The coefficient of determination r^2 is the square of the Pearson correlation coefficient. The p-value is derived from a hypothesis test whose null hypothesis is that the slope is zero, using Wald Test with t-distribution of the test statistic. Significance is indicated by yellow filling. Values for change per decade are number of days.

References

Harris, I. et al. 'Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset'. In: *Scientific Data* 7.1 (Apr. 2020). DOI: 10.1038/s41597-020-0453-3.

Roy, D. B. and T. H. Sparks. 'Phenology of British butterflies and climate change'. In: *Global Change Biology* 6.4 (Apr. 2000), pp. 407–416. DOI: 10.1046/j.1365-2486.2000.00322.x.

Shapoval, A. and P. Buczyński. 'Remarkable Odonata caught in ornithological traps on the Courish Spit, Kaliningrad Oblast, Russia'. In: *Libellula* 31 (Jan. 2012), pp. 97–109.