

## COMPLEX DEMODULATION: A NOVEL TIME SERIES METHOD FOR ANALYSING SEASONAL INFECTIOUS DISEASES

A. B. HOGAN<sup>✉1</sup>, K. GLASS<sup>1</sup> and R. S. ANDERSSSEN<sup>2</sup>

(Received 18 May, 2016; accepted 26 October, 2016; first published online 21 April 2017)

### Abstract

Understanding how seasonal patterns change from year to year is important for the management of infectious disease epidemics. Here, we present a mathematical formalization of the application of complex demodulation, which has previously only been applied in an exploratory manner in the context of infectious diseases. This method extracts the changing amplitude and phase from seasonal data, allowing comparisons between the size and timing of yearly epidemics. We first validate the method using synthetic data that displays the key features of epidemic data. In particular, we analyse both annual and biennial synthetic data, and explore the effect of delayed epidemics on the extracted amplitude and phase. We then demonstrate the usefulness of complex demodulation using national notification data for influenza in Australia. This method clearly highlights the higher number of notifications and the early peak of the influenza pandemic in 2009. We also identify that epidemics that peaked later than usual generally followed larger epidemics and involved fewer overall notifications. Our analysis establishes a role for complex demodulation in the study of seasonal epidemiological events.

2010 *Mathematics subject classification*: primary 65T99; secondary 62-07.

*Keywords and phrases*: seasonality, periodicity, infectious disease, epidemic, influenza, time series analysis.

### 1. Introduction

Many infectious diseases display annual seasonal patterns, and it is of interest to quantify how these seasonal patterns vary over time. Here, we use complex demodulation to recover the changing amplitude and phase from seasonal infectious disease data, relative to the underlying dominant frequency. Understanding changes in amplitude and phase helps to characterize and describe differences in seasonal patterns.

<sup>1</sup>Research School of Population Health, The Australian National University, Canberra, Australia; e-mail: [alexandra.hogan@anu.edu.au](mailto:alexandra.hogan@anu.edu.au), [kathryn.glass@anu.edu.au](mailto:kathryn.glass@anu.edu.au).

<sup>2</sup>CSIRO Data61, Canberra, Australia; e-mail: [bob.anderssen@data61.csiro.au](mailto:bob.anderssen@data61.csiro.au).

© Australian Mathematical Society 2017, Serial-fee code 1446-1811/2017 \$16.00

Motivation for this approach is the success with which complex demodulation has led to an improved understanding of seasonal processes in a variety of health and other situations [9, 12, 14, 17, 20]. In the analysis of the seasonality of suicide time series data [14], applying complex demodulation avoided arbitrary splitting of the data, and led to the conclusion that the strength of suicide seasonality was associated with the absolute number of suicides. Kingan et al. [11] applied complex demodulation to geomagnetic data from a number of observatories in order to improve the analysis of geomagnetic storms. Through the use of complex demodulation, it was established for the first time that the phase of the daily solar geomagnetic variation changed during a geomagnetic storm.

In an earlier work, we explored the use of complex demodulation to analyse time series data for respiratory syncytial virus (RSV) and bronchiolitis in Western Australia [10], and found a very high correlation between the timing of RSV and bronchiolitis epidemics, but weaker correlation between the magnitude of these epidemics. The aim of the present research is to formalize the application of complex demodulation to seasonal epidemic data and to show more broadly how complex demodulation can be used to enhance current understanding of the dynamics of infectious diseases. We first apply the method to synthetic data sets that display key features of epidemic time series data, and use this to understand how the underlying structure in the time series data can be depicted by the amplitude and phase of the demodulated data. We then apply the method to national influenza notification data for Australia, in order to investigate changes in epidemic size and timing.

## 2. Complex demodulation methodology

The purpose of complex demodulation is to extract and compare information from cyclical time series data. Specifically, it is used to extract the slowly varying amplitude and phase within the time series, relative to the period of the cyclical event.

**2.1. Mathematical details** In this section, we introduce the mathematical details of complex demodulation, inspired by the excellent introduction to complex demodulation within the framework of Fourier analysis given by Bloomfield [2]. Some of these mathematical details were previously presented in summary form in the supplementary material of Hogan et al. [10].

Let  $x_t$  represent a time series of data points with dominant frequency  $f_0$  that may be identified using Fourier analysis. Given that the time series exhibits this strong frequency, it follows that it has a periodic-like signal  $x(t)$  with added noise  $z_t$ , which can be modelled as

$$x_t = x(t) + z_t, \quad x(t) = R_t \cos(2\pi(f_0 t + \phi_t)), \quad (2.1)$$

where  $R_t$  is the instantaneous amplitude, and  $\phi_t$  is the instantaneous phase.

**2.1.1 Model justification.** The motivation for choosing the model in this form, as explained by Hao et al. [8] and Thomas [18], relates to the analytic modelling of

signals in communication theory. If the signal  $x(t)$  is a narrow-band signal about a central frequency  $f_0$ , which is larger than the bandwidth of its spectrum  $X(f)$ , then the inverse Fourier transform  $x_+(t)$  of the positive part  $X_+(f)$  of  $X(f)$  becomes the analytic signal associated with  $x(t)$ . If  $x_+(t)$  is written in polar coordinates as  $A(t) \exp(ig(t))$ , it follows that the corresponding  $x_t$  takes the form

$$x_t = A(t) \cos(g(t)), \quad (2.2)$$

where  $A(t)$  and  $g(t)$  are, respectively, the envelope and phase function of  $x(t)$ . To generate expressions for  $A(t)$  and  $g(t)$ , we use the above assumption that  $x(t)$  is a narrow-band signal, and work with the spectrum

$$W(f) = X_+(f + f_0),$$

which has shifted  $X_+(f)$  by the frequency  $f_0$ .

On writing the inverse Fourier transform of  $W(t)$  in polar coordinates as  $|w(t)| \exp(ih(t))$ , it algebraically follows that the magnitude  $|w(t)|$  corresponds to the envelope function  $A(t)$ , while the phase function  $g(t)$  takes the form

$$g(t) = f_0 t + h(t). \quad (2.3)$$

Combining equations (2.2) and (2.3), and including a noise component  $z_t$ , thereby justifies the structure of equation (2.1).

*2.1.2 Implementation of complex demodulation.* Equation (2.1) can be rewritten in terms of exponential functions, giving

$$x_t = \frac{1}{2} R_t (e^{2\pi i(f_0 t + \phi_t)} + e^{-2\pi i(f_0 t + \phi_t)}) + z_t. \quad (2.4)$$

There are three key steps that comprise the complex demodulation procedure. The first step is to demodulate the time series, the second is to apply a filter, and the third is to extract the amplitude and phase. Therefore, the time series is first demodulated by multiplying equation (2.4) by  $e^{-2\pi i f_0 t}$ , where  $e^{-2\pi i f_0 t}$  is referred to as the *demodulator*. After rearranging, the new expression  $y_t = x_t e^{-2\pi i f_0 t}$  takes the form

$$y_t = \frac{1}{2} R_t e^{2\pi i \phi_t} + \frac{1}{2} R_t e^{-2\pi i(2f_0 t + \phi_t)} + z_t e^{-2\pi i f_0 t}. \quad (2.5)$$

The objective is then to retain only the first term in equation (2.5), so as to allow the extraction of  $R_t$  and  $\phi_t$ . Applying a filter of length at least  $1/f_0$  will achieve this by removing the second and third terms. For the present situation, a simple moving average filter with window of length  $1/f_0$  is selected, but other possible filters were discussed by Bloomfield [2]. As the second term in the expression oscillates at frequency  $-2f_0$ , it will be eliminated. We assume that the final term, the noise component, is not smooth, and therefore will also be eliminated by an appropriate filter [2]. The application of a filter meeting the above requirements may be expressed as

$$Y_t = \Gamma[y_t] = \frac{1}{2} R_t e^{2\pi i \phi_t},$$

where  $Y_t$  is the demodulated, filtered data. The final step is to extract the instantaneous amplitude  $R_t$  and the instantaneous phase  $\phi_t$  from  $Y_t$ . Then the amplitude and phase are, respectively,

$$R_t = 2|Y_t|,$$

$$\phi_t = \text{Im} \frac{\log(Y_t)}{2\pi}.$$

The `angle` and `unwrap` functions in MATLAB may be used to compute the phase.

**2.2. Application to synthetic data** The motivation for using synthetic data is that the changing pattern in the time series is known explicitly, and can thereby be used to identify how this pattern is reflected in the structure of the complex demodulation amplitude and phase. Several synthetic data sets were created using the function

$$x(t) = \{1 + a \cos(2\theta t)\}\{\cos^2(2\theta t + b \cos(2\theta t))\}, \quad (2.6)$$

which, with various choices for  $a$ ,  $b$  and  $\theta$ , displays key features that are often observed in epidemic time series data. The parameter  $a$  can be varied to allow either for peaks of the same size in each period ( $a = 0$ ), or for alternating peak sizes ( $0 < a \leq 1$ ). The parameter  $b$  can be varied to allow for either equally spaced peaks ( $b = 0$ ), or a pattern where the peak is delayed every second period ( $0 < b \leq 1$ ). In the subsequent analysis, the terms “annual” and “biennial” are used, respectively, to describe peaks of the same and alternating sizes. The term “regular” is used to describe data where peaks occur at the same point in each period, and the term “delayed” is used to describe data where the peak is delayed every second period.

The results of the complex demodulation process applied to the synthetic data sets are shown in Figures 1 and 2. For each of these data sets, the period is  $1/f_0 = 0.25$ . In Figures 1 and 2, the data sets are plotted in the top panels (together with a moving average of the data with a window length of 0.25), the amplitudes are shown in the third panel and the phases are depicted in the fourth panel.

Figure 1 shows annual (solid red line) and biennial (dashed black line) synthetic data sets, where the spacing between peaks is the same for each period. For the annual data set, as expected, the filtered amplitude and phase are both uniform. For the biennial data set, the amplitude captures the alternating peak size, whereas, for the phase, small deviations are present. These deviations arise due to the presence of other frequencies in the data (in this case, a frequency of  $f_0 = 2$ , giving rise to the biennial pattern) that are not removed by the selected annual moving average filter (panel (ii)). If a demodulator with a frequency of  $f_0 = 2$  and a corresponding filter are applied to the biennial synthetic data set, the filtered amplitude and phase are instead uniform. We also ran the complex demodulation routine with additional synthetic data, where the relative differences in the heights of the small and large peaks are increased, and found that the deviations are more pronounced (data not shown).

Figure 2 shows two biennial synthetic data sets: panel (i) has a delay before every second (smaller) peak, while panel (ii) has equally spaced peaks. The filtered

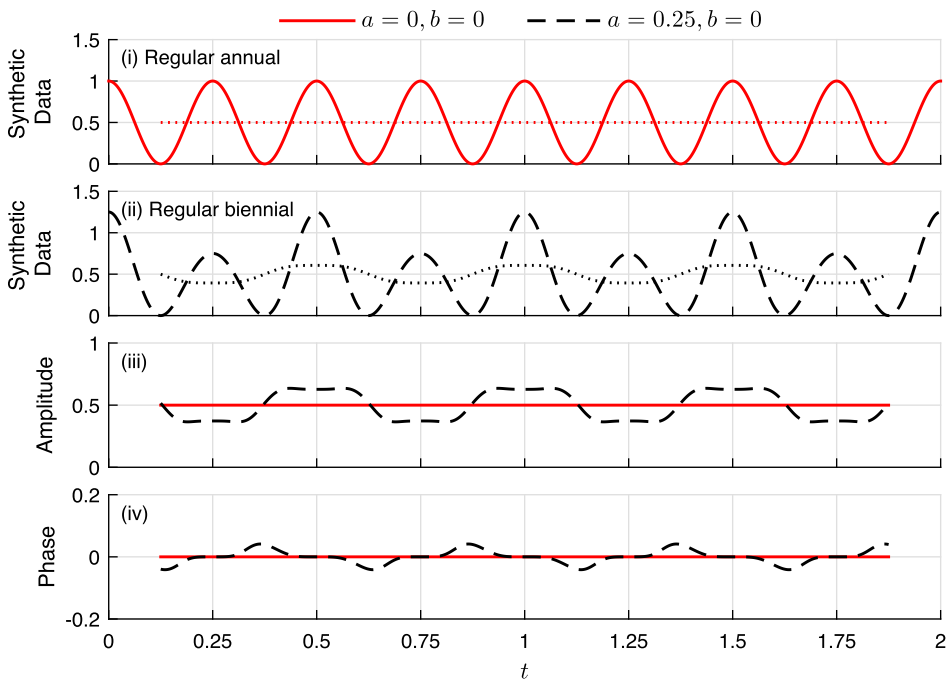


FIGURE 1. Complex demodulation applied to two synthetic data sets: a regular annual pattern (solid red line) and a regular biennial pattern (dashed black line), with  $\theta = 2\pi$ . Panels (i) and (ii) show the annual and biennial simulated data, along with a centred moving average (the dotted line in each panel). Panels (iii) and (iv) show the demodulated and filtered amplitude and phase, respectively. The synthetic data sets were created using equation (2.6) with parameters listed in the legend above (colour available online).

amplitudes (panel (iii)) are closely aligned, and the phase (panel (iv)) captures the steady epidemic timing for the regular biennial data set (dashed black line) and the alternating between-peak intervals for the delayed biennial data set (solid red line).

Comparing the features of the regular biennial pattern and the corresponding features of the delayed biennial pattern in Figure 2, we can identify how the changing structure of  $x(t)$  is reflected in the amplitude and phase. First, the centred moving average in panel (i) rises and falls more rapidly than its counterpart in panel (ii). It rises (falls) because the next peak is arriving earlier (later) than that of the demodulator. Secondly, in panel (iv), the phase for the delayed biennial data shows a stronger difference when compared to that for its regular biennial counterpart, relative to the demodulator. This is expected, as it is the phase that captures the nature of the alternating peak spacing pattern. In particular, it increases sharply in a short inter-peak interval, and drops gradually in a long inter-peak interval. These relationships are of interest, because they can be used to show how patterns in the amplitude and phase can be used as indicators of structure in real data, such as when timing differences between peaks are not immediately apparent. They also illustrate how a centred

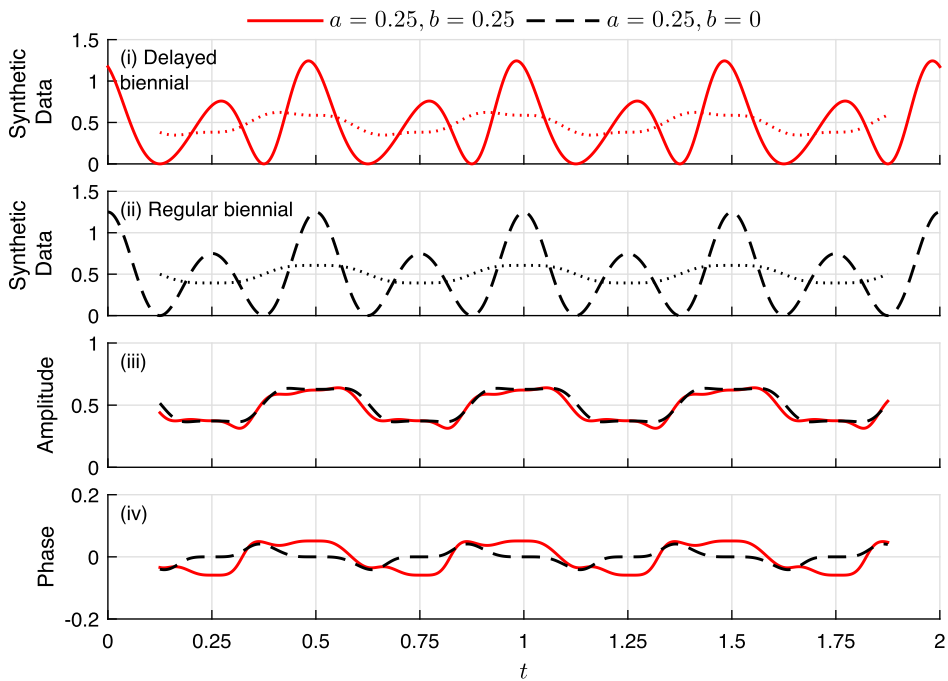


FIGURE 2. Complex demodulation applied to two synthetic data sets: a regular biennial pattern (dashed black line) and a delayed biennial pattern (solid red line), with  $\theta = 2\pi$ . Panels (i) and (ii) show the data, with a centred moving average (the dotted line in each panel). Panels (iii) and (iv) show the demodulated and filtered amplitude and phase, respectively. The synthetic data sets were created using equation (2.6) with parameters listed in the legend above (colour available online).

moving average contains information about both the timing and size of the peaks, whereas the amplitude and phase depict these components separately.

**2.3. Application to influenza data** Influenza presents a substantial global health burden, with yearly epidemics in the winter months in many temperate countries, and less defined seasonality in tropical regions [19]. There are several theories concerning the drivers of influenza seasonality, including viral evolution, host immune response and social and environmental factors, but none of these factors conclusively explain influenza dynamics. Influenza seasonality seems likely to be caused by interaction of many different factors [13]. While identifying the drivers of influenza seasonality has proved challenging, complex demodulation allows us to visualize changes in this seasonality over time.

In this paper, we examine influenza data as a representative example of seasonal infectious disease time series data, to investigate how the characteristics of influenza epidemics have shifted over the last ten years in Australia. Complex demodulation highlights these differences quantitatively, allowing identification of features that are not easily identified from visual inspection of the data. For this analysis, we

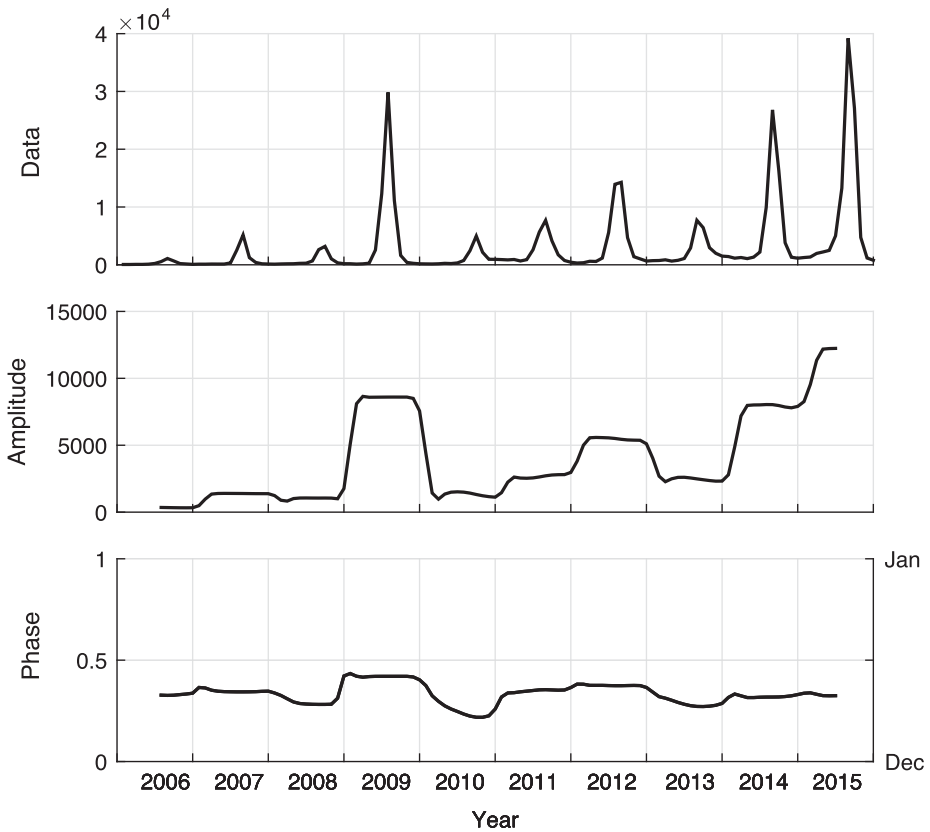


FIGURE 3. National Australian influenza notification data, and the filtered complex demodulation amplitude and phase. The data were obtained from the Australian Government Department of Health [1].

used publicly available data for laboratory-confirmed influenza notifications from the Australian National Notifiable Diseases Surveillance System (NNDSS) [1]. We extracted the number of monthly notifications from 2006 to 2015 inclusive.

The complex demodulation results for the national influenza data are shown in Figure 3. The upper panel presents the monthly data for January 2006 to December 2015. The middle panel depicts the filtered amplitude of the demodulated data and the lower panel represents the filtered phase. The phase is presented in the interval  $[0, 1]$ , which corresponds to a monthly range of December to January.

The results show several interesting features. In particular, the method extracts both the size and timing components of the pandemic in 2009. While the magnitude of the pandemic is clearly apparent in the time series data and reflected in the amplitude plot, the phase shows that this outbreak took place earlier than in previous years, with the initial onset occurring quite rapidly. The amplitude panel also clearly reflects the trend of increasing numbers of notifications over time. Finally, the phase identifies where

influenza epidemics occurred later in the year, such as in 2008, 2010 and 2013, and the corresponding amplitude indicates that these years typically involved fewer cases than the previous year.

### 3. Discussion

The analysis presented in this paper provides motivation for the use of complex demodulation in the analysis of periodic infectious disease data. Complex demodulation extracts key seasonal components of interest, the amplitude and phase, in order to quantify and visualize patterns in the data. It can also be used as a tool to compare data for different regions or different diseases.

Complex demodulation complements other time series analysis methods, such as Fourier analysis and wavelet analysis. Fourier analysis is an important method for detecting the different underlying frequencies in a data set, and is used for determining the demodulator in the complex demodulation method [2]. Wavelet analysis has been successfully applied in a variety of contexts, including data for measles epidemics in England [7] and dengue hemorrhagic fever in Thailand [3]. Wavelet analysis is appropriate when the data are non-stationary, that is, when the dominant frequency changes over time, and when the location of some isolated event needs to be identified [16]. In cases where the data are stationary, such as for the present influenza data, complex demodulation offers distinct advantages in that it is suitable for extracting and visualizing variations in the amplitude and phase with respect to those of the underlying dominant frequency over time.

In this paper, we examined synthetic data with annual and biennial patterns (Figures 1 and 2) that are typical of many infectious diseases. For example, while influenza typically displays an annual pattern, diseases such as pre-vaccination measles [4] and RSV [15] often display biennial patterns in temperate regions. The results for these synthetic data sets inform the interpretation of complex demodulation applied to real data. Using synthetic data in the manner outlined above allows features in the complex demodulation amplitude and phase to be related to the known structure in the synthetic data. These features then become indicators of the structure in real data, as explained for the influenza data.

The application to national influenza notification data for Australia demonstrates how complex demodulation can be used to analyse infectious disease dynamics (Figure 3). The amplitude output clearly identifies the earlier outbreak and high number of notifications associated with the pandemic influenza strain of 2009. The increase in amplitude from 2010 onwards reflects an increase in testing practices following the pandemic in 2009. It is interesting to observe that the high-amplitude years (2009 and 2012) are followed by a dip in both the amplitude and the phase the following year, with fewer notified cases and a later epidemic. A possible driver of this behaviour is increased levels of immunity resulting from high incidence in the previous year, although we cannot rule out differences in the distribution of subtypes from year to year that may influence the timing and size of the peak in notifications.



A key consideration in complex demodulation analysis is the interpretation of the amplitude. For each of the Figures 1 and 2, the moving average of the original data set is plotted in panels (i) and (ii), to allow direct comparison with the amplitude. The complex demodulation amplitude is not the amplitude of either the data or the moving average of the data, but reflects the data relative to the demodulator.

In our analysis, we found that small irregularities in the amplitude and phase occur, when a demodulator and filter corresponding to an annual frequency are applied to data with a strong biennial frequency. We also found that increasing the difference between the heights of the small and large peaks produced more pronounced deviations in the phase output. These deviations can be removed by use of a wider moving average window; however, this will remove some desired components of the analysis, as well as more values from the beginning and end of the data set [2].

The drivers of seasonality for diseases such as influenza and RSV remain a much-studied question. While individual climate and demographic drivers have been identified in various settings [5, 6], there is likely no single driver, and a combination of multiple factors leads to higher disease incidence during winter months in temperate regions. Nevertheless, methods such as complex demodulation allow us to better identify yearly variability in these underlying seasonal patterns. In future work, we plan to apply this method to explore the timing of influenza epidemics in different regions of Australia, and investigate the potential of this method to analyse data for two respiratory infections at the same time.

## References

- [1] Australian Government Department of Health, “Number of notifications of Influenza (laboratory confirmed), Australia, 2016”, <http://www9.health.gov.au/cda/source/cda-index.cfm>.
- [2] P. Bloomfield, *Fourier analysis of time series: An introduction* (Wiley, New York, 2000).
- [3] B. Cazelles, K. Cazelles and M. Chavez, “Wavelet analysis in ecology and epidemiology: impact of statistical tests”, *J. R. Soc. Interface* **11** (2014) 1–10; doi:10.1098/rsif.2013.0585.
- [4] B. Finkenstädt and B. Grenfell, “Empirical determinants of measles metapopulation dynamics in England and Wales”, *Proc. R. Soc. Lond. B* **265** (1998) 211–220; doi:10.1098/rspb.1998.0284.
- [5] D. N. Fisman, “Seasonality of infectious diseases”, *Ann. Rev. Pub. Health* **28** (2007) 127–143; doi:10.1146/annurev.publhealth.28.021406.14412.
- [6] N. C. Grassly and C. Fraser, “Seasonal infectious disease epidemiology”, *Proc. R. Soc. B* **273** (2006) 2541–2550; doi:10.1098/rspb.2006.3604.
- [7] B. T. Grenfell, O. N. Bjørnstad and J. Kappey, “Travelling waves and spatial hierarchies in measles epidemics”, *Nature* **414** (2001) 716–723; doi:10.1038/414716a.
- [8] Y.-L. Hao, Y. Ueda and N. Ishii, “Improved procedure of complex demodulation and an application to frequency analysis of sleep spindles in EEG”, *Med. Biol. Eng. Comput.* **30** (1992) 406–412; doi:10.1007/BF02446168.
- [9] J. Hayano, J. A. Taylor, A. Yamada, S. Mukai, R. Hori, T. Asakawa, K. Yokoyama, Y. Watanabe, K. Takata and T. Fujinami, “Continuous assessment of hemodynamic control by complex demodulation of cardiovascular variability”, *Am. J. Physiol.* **264** (1993) H1229–H1238; <http://ajpheart.physiology.org/content/264/4/H1229.long>.
- [10] A. B. Hogan, R. S. Anderssen, S. Davis, H. C. Moore, F. J. Lim, P. Fathima and K. Glass, “Time series analysis of RSV and bronchiolitis seasonality in temperate and tropical Western Australia”, *Epidemics* **16** (2016) 49–55; doi:10.1016/j.epidem.2016.05.001.

- [11] P. A. Kingan, P. Bloomfield and R. S. Anderssen, “Phase drift and coherence in geomagnetic data during a magnetic storm (Dst)”, *J. Geomagn. Geoelectr.* **32** (1980) 57–65; doi:10.5636/jgg.32.57.
- [12] H. Kondo, M. Ozone, N. Ohki, Y. Sagawa, K. Yamamichi, M. Fukuju, T. Yoshida, C. Nishi, A. Kawasaki, K. Mori, T. Kanbayashi, M. Izumi, Y. Hishikawa, S. Nishino and T. Shimizu, “Association between heart rate variability, blood pressure and autonomic activity in cyclic alternating pattern during sleep”, *Sleep* **37** (2014) 187–194; doi:10.5665/sleep.3334.
- [13] E. Lofgren, N. H. Fefferman, Y. N. Naumov, J. Gorski and E. N. Naumova, “Influenza seasonality: underlying causes and modeling theories”, *J. Virol.* **81** (2007); doi:10.1128/JVI.01680-06.
- [14] I. W. Nader, J. Pietschnig, T. Niederkrotenthaler, N. D. Kapusta, G. Sonneck and M. Voracek, “Suicide seasonality: complex demodulation as a novel approach in epidemiologic analysis”, *PLoS ONE* **6** (2011) e17413; doi:10.1371/journal.pone.0017413.
- [15] V. E. Pitzer, C. Viboud, W. J. Alonso, T. Wilcox, C. J. Metcalf, C. A. Steiner, A. K. Haynes and B. T. Grenfell, “Environmental drivers of the spatiotemporal dynamics of respiratory syncytial virus in the United States”, *PLoS Pathog.* **11** (2015) e1004591; doi:10.1371/journal.ppat.1004591.
- [16] M. B. Priestley, “Wavelets and time-dependent spectral analysis”, *J. Time Series Anal.* **17** (1996) 85–103; doi:10.1111/j.1467-9892.1996.tb00266.x.
- [17] T. Sasai, M. Matsuura and Y. Inoue, “Change in heart rate variability precedes the occurrence of periodic leg movements during sleep: an observational study”, *BMC Neurol.* **13** (2013); doi:10.1186/1471-2377-13-139.
- [18] J. B. Thomas, *An introduction to statistical communication theory* (Wiley, New York, 1969).
- [19] C. Viboud, W. J. Alonso and L. Simonsen, “Influenza in tropical regions”, *PLoS Med.* **3** (2006) 468–471; doi:10.1371/journal.pmed.0030089.
- [20] F. Yasuma and J. Hayano, “Respiratory sinus arrhythmia: why does the heartbeat synchronize with respiratory rhythm?”, *Chest* **125** (2004) 683–690; doi:10.1378/chest.125.2.683.