

Multifractal signatures of infectious diseases

A. M. Holdsworth¹, N.K.-R. Kevlahan² and David J.D. Earn^{2,3}

¹ *Department of Earth & Atmospheric Sciences, University of Alberta,
Edmonton, Alberta, Canada*

² *Department of Mathematics & Statistics, McMaster University, Hamilton,
Ontario, Canada*

³ *M.G. DeGroot Institute for Infectious Disease Research, McMaster
University, Hamilton, Ontario, Canada*

Key words: infectious disease, multifractal, wavelet.

Incidence of infection time series data for the childhood diseases measles, chicken pox, rubella, and whooping cough are described in the language of multifractals. We explore the potential of using the wavelet transform maximum modulus (WTMM) method to characterize the multiscale structure of the observed time series and of simulated data generated by the stochastic SEIR epidemic model. The singularity spectra of the observed times series suggest that each disease is characterized by a unique multifractal signature, which distinguishes that particular disease from the others. The wavelet scaling functions confirm that the times series of measles, rubella and whooping cough are clearly multifractal, while chicken pox has a more monofractal structure in time. The stochastic SEIR epidemic model is unable to reproduce the qualitative singularity structure of the reported incidence data: it is too smooth and does not appear to have a multifractal singularity structure. The precise reasons for the failure of the SEIR epidemic model to reproduce the correct multiscale structure of the reported incidence data remain unclear.

1. Introduction

Accurate modelling of the transmission of vaccine-preventable childhood infectious diseases is of great importance as morbidity and mortality rates continue to be significant, particularly in some developing nations¹. Our understanding of the origin of recurrent outbreaks^{2,3} and changes in the period between successive outbreaks⁴⁻⁶ has improved substantially in the last few decades. Further developments in our understanding of these dynamics will enhance our ability to identify models that can guide the design of effective control and eradication strategies⁷.

In this paper we focus on two fundamental questions related to the statistical analysis of epidemiological times series. First, is it possible to identify a disease from a statistical analysis of incidence time series data alone? Secondly, does the

standard susceptible-exposed-infectious-recovered (SEIR) epidemic model capture the multiscale time structure of the observed incidence data?

Power spectra (based on the Fourier transform) are traditionally used to characterize the frequency content of a signal, but they provide no information about the frequency content at a particular time. That is, Fourier methods work well for stationary time series, but not for non-stationary signals, where the frequency content changes over time. Scalograms based on the wavelet transform provide simultaneously both time and frequency information, which is important information for epidemiological data where the frequency content is typically complex and non-stationary in time.

Wavelet analysis can also characterize the smoothness of time series by using the wavelet transform maximum modulus (WTMM) technique^{8–10} to construct the singularity spectrum associated with the fractal or multifractal structure of the data. The singularity spectrum is a particularly useful tool for analyzing and comparing time series with irregular (possibly chaotic) multiscale structure. The WTMM method has been used to analyze a wide variety of time series data from biological systems, such as electrocardiographic (ECG) signals¹¹, human gait recordings¹², electroencephalographic (EEG) signals¹³, and functional magnetic resonance imaging (fMRI) time series¹⁴. In particular, it has been suggested that the width and peak value of the singularity spectrum of ECG signals are influenced by disease and ageing, and may therefore have diagnostic value. Indeed, in a recent article Chiu et al¹⁵ claim that a particular heart drug can actually restore the normal singularity spectrum of heart beat time series in patients with advanced congestive heart failure! These investigations suggest that singularity spectra constructed using the WTMM method are a valuable tool for understanding and classifying the statistics of complex biological time series.

A number of studies have already used wavelet based techniques to analyze recurrence in epidemiological time series^{16,17}. They used local wavelet power spectra to investigate patterns of recurrence in the incidence of infection time series for childhood diseases, however they did not look for any associated fractal structure.

In this paper we use wavelet-based multifractal analysis to characterize and understand the incidence of infection time series for a number of important childhood diseases. We also investigate whether synthetic data generated by the stochastic SEIR epidemic model can be made to match qualitatively the singularity spectrum (and hence multifractal structure) of the observed time series data for common childhood diseases. In brief, our aims are to determine whether infectious disease time series can be characterized by “multifractal signatures” and, if so, whether the standard stochastic SEIR model can reproduce these signatures. We show that each infectious disease does indeed appear to be characterized by a unique multifractal signature, but we have not been able to reproduce these signatures with the stochastic SEIR model. We believe that this is the first study to apply the language of multifractal analysis to epidemiological time series.

Section 2 introduces the wavelet based multifractal formalism, section 3 describes the epidemiological time series and stochastic SEIR model. In section 4 the main results are presented and a brief discussion and conclusions are given in section 5.

2. Wavelet theory and data analysis

The wavelet transform of a signal $s(t)$ is given by

$$\tilde{s}(a, b) = \int_{-\infty}^{+\infty} s(t) \overline{\psi}_{a,b}(t) dt \quad (2.1)$$

where the ‘‘mother wavelet’’ $\psi(t)$ has been shifted to time b and dilated or compressed to scale a ,

$$\overline{\psi}_{a,b}(t) = \frac{1}{a^{1/p}} \overline{\psi}\left(\frac{t-b}{a}\right). \quad (2.2)$$

In this definition, $p=2$ gives L^2 normalization, which is used when calculating wavelet power spectra, while $p=1$ gives L^1 normalization, which is used when measuring local regularity. A large amplitude wavelet coefficient $\tilde{s}(a, b)$ indicates that at time b the signal has significant variation at frequency $1/a$.

The choice of analyzing wavelet (mother wavelet) is guided by the application, as well as by the structure of the data to be analyzed. Complex-valued wavelets, such as the Morlet wavelet, are ideal for capturing a signal’s oscillatory behaviour (e.g., local wavelet power spectra) as they provide information about both amplitude and phase. Real valued wavelets, such as the Gaussian family of wavelets, return only a single component making them well suited to measure local regularity (i.e., local singularity strength). The appropriate choice of mother wavelet will be discussed separately for the computation of the wavelet power spectrum and the measurement of local regularity.

Power Spectra

The wavelet power spectrum, also known as the scalogram, is defined as

$$P(a, b) = |\tilde{s}(a, b)|^2, \quad (2.3)$$

and the total energy of a signal s is its wavelet power spectrum integrated over all scales and times,

$$E = \int_{-\infty}^{\infty} \int_0^{\infty} P(a, b) da db. \quad (2.4)$$

To compute scalograms we adapted the on-line wavelet toolbox provided by Torrence and Compo¹⁸, which includes a guide for wavelet spectral analysis. Before taking the wavelet transform of the time series data, we normalize the data by subtracting its mean and dividing by its standard deviation. It can be shown that for Gaussian white noise with mean zero and variance one the expectation of $P(a, b)$ is one. Thus, using the previous normalization, $P(a, b)$ directly measures the power of the scalogram relative to white noise. This provides a useful diagnostic for noisy data since we can say, with 95% confidence, that the part of the wavelet power spectrum above the $P(a, b) = 2$ contour is significant (i.e., *not* due simply to random white noise fluctuations). We plot this 95% confidence contour on all scalogram plots to identify the most significant features. Note that shifting by the mean and normalizing by the standard deviation also

makes it easy to compare the periodic structure of different data sets (provided the sampling rate is also taken into account).

We approximated continuous wavelet transforms using the fast Fourier transform (FFT). If the data are not periodic, the FFT introduces edge effects (discontinuities at the edge of the time series). To reduce this problem, zeros were added to both ends of the time series before transforming (this is known as *zero padding*). Two contours identifying the so-called *cone of influence* of the *edge effects* are plotted on the scalograms to indicate which areas in position and scale may be influenced by non-periodicity of the data; between these contours, edge effects are considered negligible.

Consistent with Grenfell *et al.*¹⁷ we chose the Morlet wavelet,

$$\psi(t) = c_{\sigma} \pi^{-1/4} e^{-\frac{1}{2}t^2} e^{i\omega_0 t}, \quad (2.5)$$

to compute scalograms for each of the data sets. For the Morlet wavelet with $\omega_0 = 6$ the value of the Fourier period is $\lambda \simeq 1.03$, so scale and period are nearly equivalent¹⁸. This allows us to easily plot period versus time (rather than scale versus time). More importantly, Heisenberg's uncertainty principle means there is a tradeoff between localization in frequency and localization in time. The Morlet wavelet provides excellent frequency resolution at the expense of temporal resolution. This trade-off is near-optimal for scalograms, but better temporal resolution is needed for analyzing local singular structure, as we explain in the following section.

Singularity Spectra

The decay of wavelet coefficients is determined by the local regularity of the signal, and therefore the local regularity of a signal may be determined by measuring the rate of decay of the wavelet coefficients. This is the basis of the WTMM method.

The Lipschitz (or Hölder) exponent, α , is a measure of the local regularity of a function. A function, $f(t)$, satisfies a Lipschitz condition of order α at a point, t , if there is a non-negative real number k such that

$$|f(t + \Delta t) - f(t)| \leq k|\Delta t|^{\alpha} \quad \text{as } \Delta t \rightarrow 0. \quad (2.6)$$

f being Lipschitz of order 0 is equivalent to being bounded. If $k = 0$ then the Lipschitz condition is equivalent to the ordinary definition of continuity. The local Hölder exponent $\alpha(t)$ is defined to be the maximum exponent for which the above condition holds, i.e., $\alpha(t) = \sup\{\alpha_0 : f \text{ is Lipschitz of order } \alpha_0 \text{ at the point } t\}$. The function $f(t)$ is said to be singular at the point t if $\alpha(t) < 1$ and the strength of the singularity is greater if $\alpha(t)$ is further from 1.

If $-1/2 < \alpha < 0$ the data are *persistent* or positively correlated, with long-term memory effects, while if $-1 < \alpha < -1/2$ the data are *anti-persistent* or negatively correlated ($\alpha = -1/2$ implies white noise, i.e., the temporal autocorrelation function is a δ -function).

When a persistent time series increases/decreases from t_{n-1} to t_n then it is expected to increase/decrease from t_n to t_{n+1} . Conversely, for an anti-persistent time series an increase is expected to be followed by a decrease. The smoothness of a function as measured by its local Hölder exponent is summarized in table 1.

Hölder exponent α	Singularity type	Example
$\alpha > 1$	continuous, differentiable	Smooth curve
$\alpha = 1$	continuous, differentiable almost everywhere	
$0 < \alpha < 1$	continuous, non-differentiable	Brownian motion: $\alpha = 1/2$
$-1 < \alpha \leq 0$	discontinuous, non-differentiable	Heaviside function: $\alpha = 0$ Gaussian noise: $\alpha = -1/2$
$\alpha \leq -1$	not locally integrable	Dirac pulse: $\alpha = -1$

Table 1. Interpretation of Hölder exponent α in terms of the local regularity of a function. Note that if $f(t)$ has exponent α then df/dt has exponent $\alpha - 1$ and $\int_0^t f(u) du$ has exponent $\alpha + 1$. If f is Lipschitz α uniformly (for the same k) in a neighbourhood of a point t , and $\alpha > n$ (n a positive integer), then f is n -times continuously differentiable in this neighbourhood⁹.

A signal is said to have a *multifractal* structure when the Hölder exponent varies in time. In contrast, a *monofractal* signal has the same Hölder exponent at each time point in the signal. The spectrum of singularities of the entire signal can be estimated using the Wavelet Transform Maximum Modulus (WTMM) Method⁹ described below.

A WTMM⁹ is a point (a_0, b_0) in scale-time space at which $|\tilde{s}(a_0, b)|$ attains a strict local maximum at a fixed scale $a = a_0$. This implies, in particular, that

$$\left. \frac{\partial \tilde{s}(a, b)}{\partial b} \right|_{(a_0, b_0)} = 0. \quad (2.7)$$

The existence of a singularity at a point $b = b_*$ means that there is a sequence of local wavelet maxima at each scale that converges to the point b_* as scale $a_0 \rightarrow 0$. Only the largest amplitude WTMM in each interval of size a_0 is retained at each scale a_0 , and these WTMM are connected across scales to form the WTMM lines. The rate of decay of the wavelet moduli along the WTMM lines with decreasing scale estimates the pointwise Lipschitz regularity. If $|\tilde{s}(a, b)|$ has no modulus maxima at fine scales, then f is locally regular.

The distribution of singularities is described by the *singularity spectrum*, $D(\alpha)$, which represents the proportion of Lipschitz α singularities that occur at any time for a given scale a . The regularity of a signal is thus characterized by the regularity of its subsets.

Define a set $\{b_n(a)\}_{n \in \mathbb{Z}}$ to be the temporal positions of the local maxima at a fixed scale a . Now, define a partition function Z :

$$Z(a, q) = \sum_n |\tilde{s}(a, b_n)|^q. \quad (2.8)$$

This function measures the sum at a power q of all the aforementioned local modulus maxima. The wavelet itself defines the shape of the partitions and the scale parameter dictates the size. WTMM are used to indicate how the partitions should be taken at each scale. The scaling exponent $\tau(q)$ measures the asymptotic decay of $Z(a, q)$ at fine scales a for each $q \in \mathbb{R}$.

$$\tau(q) = \liminf_{a \rightarrow 0} \frac{\log Z(q, a)}{\log a}. \quad (2.9)$$

The scaling exponent $\tau(q)$ is the Legendre transform of the singularity spectrum $D(\alpha)$ ⁹. Jaffard¹⁰ generalized the result of⁸ which relates the scaling exponent, $\tau(q)$, to the singularity spectrum.

Suppose the support of $D(\alpha)$ is $\Lambda = [\alpha_{min}, \alpha_{max}]$. Let ψ be a wavelet with $n > \alpha_{max}$ vanishing moments. If f is self-similar then:

$$\tau(q) = \min_{\alpha \in \Lambda} \left(q \left(\alpha + \frac{1}{2} \right) - D(\alpha) \right). \quad (2.10)$$

Computing the derivative of equation 2.10 reveals $q(\alpha) = \frac{dD}{d\alpha}$. From this computation and using the fact that $\tau(q)$ is at a minimum we derive that $D(\alpha)$ is a convex function, and $\tau(q)$ is an increasing and convex function. For the Legendre transform to be invertible, $D(\alpha)$ must be convex. Details of the proof are given in¹⁰. Note that the negative of the scaling function is used in the computation giving a *concave* spectrum,

$$D(\alpha) = \min_{q \in \mathbb{R}} q \left(\alpha + \frac{1}{2} \right) - \tau(q). \quad (2.11)$$

$D(\alpha)$ is the fractal dimension of the set with Holder exponent α . If the set of points where the signal is Lipschitz/Holder α is an empty one, by convention $D(\alpha) = -\infty$.

A closer look at equation 2.10 shows that the maximum or peak of the singularity spectrum occurs at $q = 0$,

$$-\tau(0) = \max_{\alpha} f(\alpha). \quad (2.12)$$

The right side of $D(\alpha)$ is computed from negative q , and the left from positive values of q .

Classifying a signal as either monofractal or multifractal is an important, but delicate, aspect of multifractal analysis. In principle, the singularity spectrum of a true monofractal should be a single point, $D(\alpha_0) = 1$. However, the wavelet based multifractal formalism often generates spurious data points in the singularity spectrum which cause the singularity spectrum of a monofractal to have a finite (although small) width. This may lead to the false conclusion that a signal is a multifractal, when it is in fact a monofractal.

Notice from equation 2.10 that there is an important *linear* relationship between $\tau(q)$ and the Hurst exponent¹⁹, $h = \alpha_0 + 1/2$, for monofractal signals,

$$\tau(q) \approx qh - 1. \quad (2.13)$$

Therefore, linear behaviour of $\tau(q)$ (and the narrow width of the singularity spectrum) indicates the presence of a monofractal, while non-linear behaviour (and a wide singularity spectrum) indicates multifractality.

Numerical implementation of the WTMM method

The numerical analysis tools required to implement the WTMM method are modifications of codes which are readily available from Wavelab²⁰, an open source wavelet toolbox for signal processing.

We suggested above that real wavelets are an appropriate choice for examining the local regularity of a signal. To choose the best wavelet for the WTMM method

amongst the real wavelets available we consider various intrinsic properties of the mother wavelet.

A function $\psi(t)$ is said to have p vanishing moments if

$$\int_{-\infty}^{+\infty} t^k \psi(t) dt = 0 \quad (2.14)$$

for $0 \leq k < p$. Wavelets with n vanishing moments can only detect the regularity, α , of f for $\alpha < n$. Theorem 6.5 of Mallat's book⁹ proves that choosing a Gaussian wavelet guarantees that all maxima lines propagate to the finest scales. The family of Gaussian wavelets includes all derivatives of the Gaussian function. They have infinitely many vanishing moments and the n^{th} derivative of a Gaussian can measure Lipschitz exponents up to order $\alpha < n$.

It is advantageous to choose a wavelet with a high number of vanishing moments to measure higher orders of regularity, but increasing the number of vanishing moments also increases the number of WTMM lines in the cone of influence²¹. The presence of many lines makes it more difficult to track the WTMM lines and accurately detect the singularities present in the time series. Therefore, the number of vanishing moments should be kept to a minimum consistent with the expected regularity of the signal to be analyzed.

The childhood infection time series are highly oscillatory with many isolated singularities. Therefore, choosing a wavelet with minimal effective compact support increases the resolution in our analysis of the singularity structure because the larger the effective compact support is, the more wavelets there are that intersect a particular singularity. As the order of the Gaussian derivative increases the number of wavelet oscillations increase. To find the correct balance between having enough zero crossings and a minimal compact support, a range of Gaussian wavelets were tested on the infectious disease time series.

3. Data sets

Our study compares the multifractal singularity structure of incidence time series for several childhood infectious diseases (measles, chicken pox, rubella, and whooping cough, from several geographic locations) with synthetic data produced by the stochastic SEIR model.

Reported incidence time series

The time series used in this study can be found at the International Infectious Disease Data Archive (IIDDA), an online resource for infectious disease data²². In the rare instances where data points were missing, the values were interpolated with cubic splines so as to minimize any effects on the singularity spectrum. Any negative interpolations were set to zero. We analyzed data for four common childhood infectious diseases (measles (meas), rubella (Rub), whooping cough (WC) and chicken pox (CP)) from four Canadian provinces (British Columbia, Saskatchewan, Manitoba and Ontario), two American cities (New York and Baltimore) and two British cities (London and Liverpool).

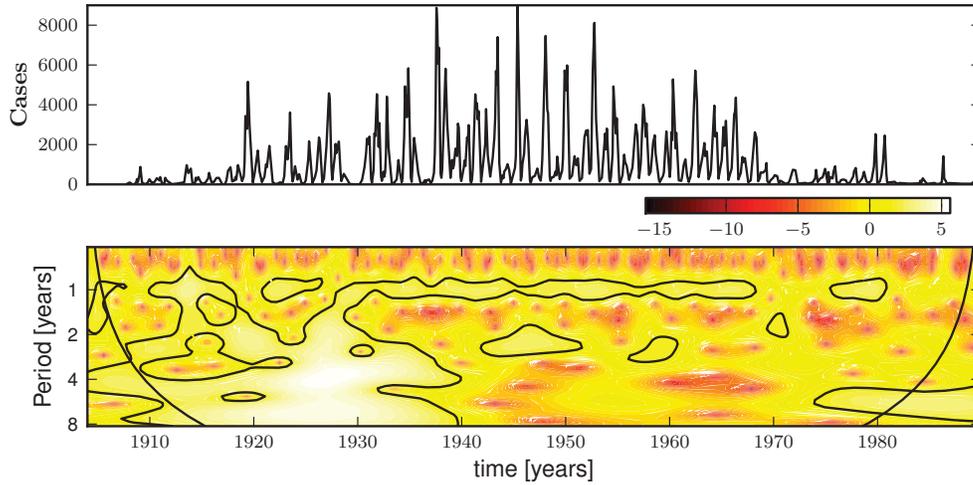


Figure 1. (a) Reported monthly measles cases in Ontario, 1904–1989. (b) Corresponding wavelet power spectrum computed using the Morlet wavelet. The 95% confidence level (compared to Gaussian white noise) is indicated by black contours. The cone of influence indicates that part of the wavelet power spectrum that is contaminated by end effects.

Figure 1 shows (a) measles incidence in Ontario (1904–1989) and (b) the corresponding wavelet power spectrum. High amplitude wavelet coefficients are isolated by dark contours and reveal a distinctive period one recurrence in the time series. An annual rise in infections is attributed to an increase in contact rates at the beginning of the school year^{3,4}, while the presence of significance contours at non-integer values of the period indicates non-seasonal recurrence²³. After mass measles vaccination was introduced in the late 1960s, the number of cases dropped dramatically. A distinct decrease in the number of high amplitude wavelet coefficients is apparent from the scalogram.

A strong yearly recurrence of chicken pox is unmistakable in figure 2. These monthly data were collected from 1928–1972 in New York City.

The incidence of whooping cough, as reported in Ontario from 1904–1989, is plotted in figure 3 (a). The wavelet scalogram in figure 3(b) demonstrates period 3-5 year recurrence after 1945.

Figure 4(a) shows monthly incidence of rubella in Ontario from 1929-1989. The scalogram illustrates complex patterns of recurrence with both seasonal and non-seasonal peaks.

Simulated incidence time series

The WTMM method was also applied to incidence data generated by the standard stochastic SEIR epidemic model^{7,24}. The SEIR model divides the host population into compartments containing susceptible (S), exposed (E), infectious (I), and recovered (R) individuals. Susceptibles have no immunity and can become infected upon contact with an infectious individual. Exposed individuals have been infected but are not yet infectious. Infectious individuals can infect susceptibles

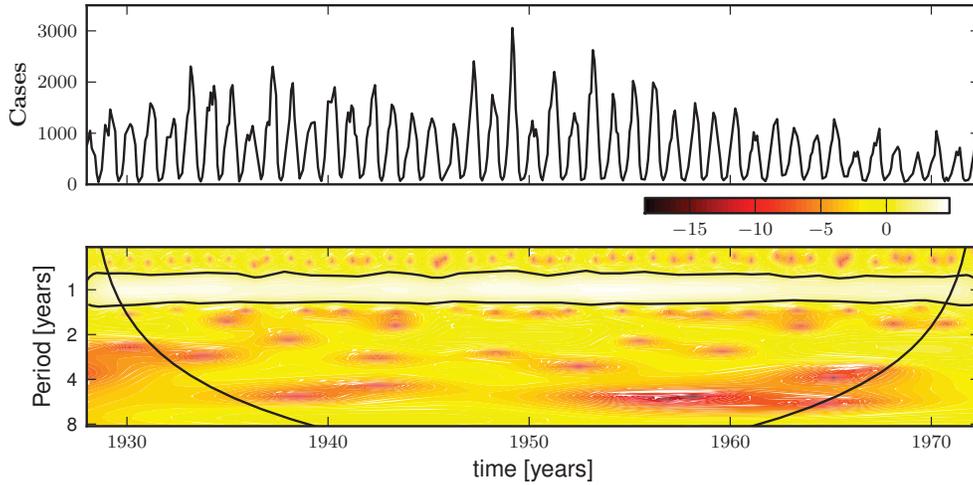


Figure 2. (a) Reported monthly chicken pox cases in New York City, 1928–1972. (b) Corresponding wavelet power spectrum as in figure (1).

who they contact. Recovery is assumed to entail lifelong immunity. Assuming a mass-action contact process, the mean-field in the large population size limit²⁵ is governed by the standard deterministic SEIR model^{7,26}, which is specified by the following set of differential equations.

$$\dot{S} = \nu(1 - p) - \beta SI - \mu S \quad (3.1a)$$

$$\dot{E} = \beta SI - \sigma E - \mu E \quad (3.1b)$$

$$\dot{I} = \sigma E - \gamma I - \mu I \quad (3.1c)$$

$$\dot{R} = \nu p + \gamma I - \mu R \quad (3.1d)$$

The total population size is $N = S + E + I + R$. ν is the birth rate, which varies seasonally in reality²⁷ but is usually assumed constant or very slowly varying^{4,28}. p is the proportion of individuals who are vaccinated before encountering infectious individuals. μ is the *per capita* death rate (from “natural” causes; disease-induced mortality is negligible for the diseases we consider here). If $\nu = \mu N$ then N remains constant (which was true in our simulations). β is the *transmission rate*, which is typically time-varying for childhood infections (as a result of the aggregation of children in schools in term-time³). In our simulations we used sinusoidal seasonal forcing,

$$\beta = b_0(1 + b_1 \cos(2\pi t)). \quad (3.2)$$

Thus, b_0 is the mean transmission rate and b_1 is the amplitude of seasonality. σ is the (constant) rate at which exposed individuals become infectious (so the mean latent period is $1/\sigma$). γ is the (constant) rate of recovery, so the mean infectious period is $1/\gamma$. The basic reproduction number, the average number of secondary infections caused by an infectious individual in a wholly susceptible population,

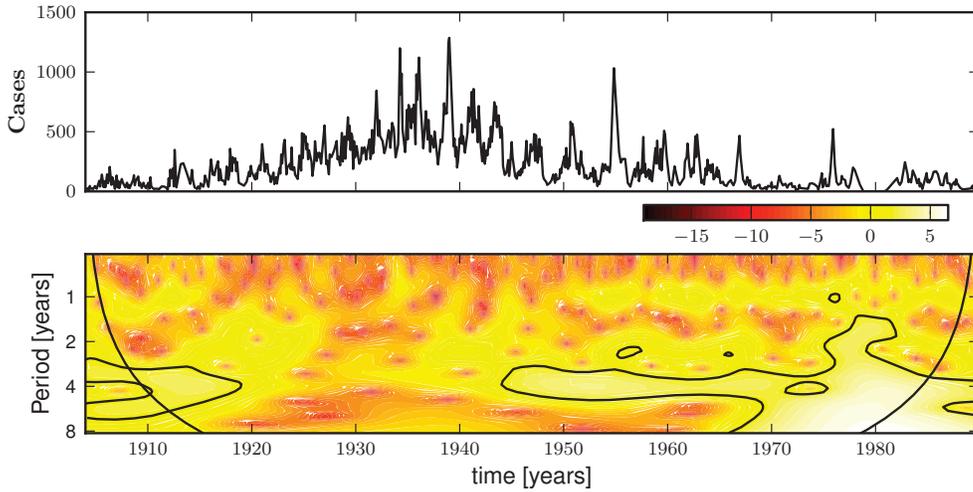


Figure 3. (a) Reported monthly whooping cough cases in Ontario, 1904–1989. (b) The corresponding wavelet power spectrum as in figure (1).

is⁷

$$\mathcal{R}_0 = \frac{\beta\sigma}{(\gamma + \mu)(\sigma + \mu)}. \quad (3.3)$$

We implemented the stochastic SEIR model using the standard Gillespie algorithm²⁹ with event rates given by each of the terms in equations (3.1). While the SEIR model ignores many aspects of real demographic and epidemiological interactions, it nevertheless successfully captures many features of real epidemics^{7,28}.

Figures 5 and 6 show simulation time series and wavelet power spectra for comparison with the reported incidence in the earlier figures. Figure 5 was generated with parameter values appropriate for measles ($\mathcal{R}_0 = 17$, $1/\sigma = 8$ days, $1/\gamma = 5$ days, amplitude of seasonal forcing $b_1 = 0.08$) and a population of two million. The scalogram in Figure 5(b) exhibits the same period one recurrence observed for the Ontario measles data in figure 1, but shows significant period 2 recurrence which is not present in the Ontario data and fails to capture the complexity of recurrence over larger intervals of time.

Figure 6 is based on a simulation of chicken pox dynamics ($\mathcal{R}_0 = 10.5$, $1/\sigma = 8$ days, $1/\gamma = 5$ days, $b_1 = 0.08$) and $N = 10$ million. Again, significant period one coefficients reveal an annual pattern of recurrence in Figure 6(b). A similar pattern was observable from the reported incidence of chicken pox in New York City shown in figure 2.

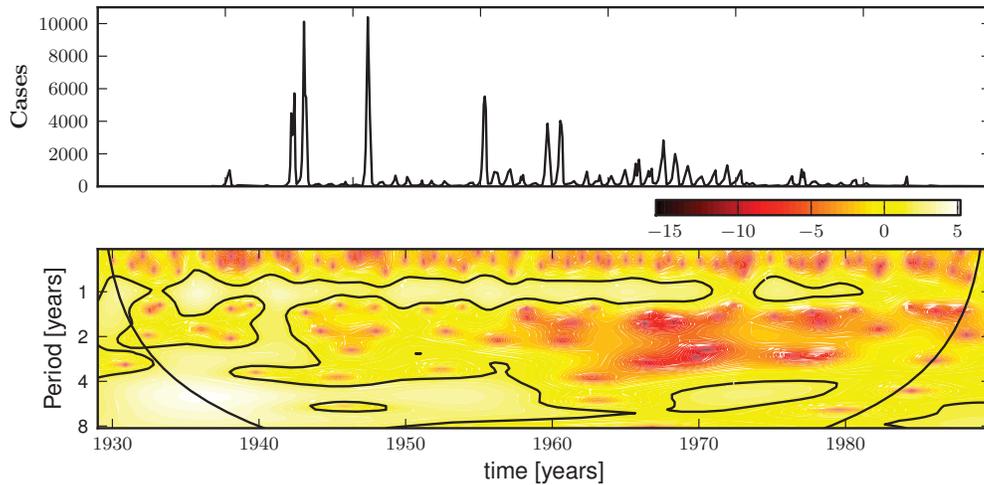


Figure 4. (a) Reported monthly rubella cases in Ontario, 1929–1989. (b) The corresponding wavelet power spectrum as in figure (1).

4. Results

(a) Analysis of incidence of infection time series data

To select the optimal wavelet for the time series we computed the singularity spectrum $D(\alpha)$ using Gaussian analyzing wavelets of increasing order with $q \in [-5, 5]$. Figure 7(a) shows the spectra computed from the monthly Ontario measles incidence. The figure is a representative example of convergence results for measles data, but the location of the peak varies with geographic location. The fourth-order Gaussian wavelet was the lowest order Gaussian wavelet for which we observed convergence, and therefore was the optimal choice.

Figure 7(b) shows that $D(\alpha)$ for weekly reported chicken pox in Ontario is also convergent, having similar peak locations and overall shapes for all orders of wavelets tested. All of the available chicken pox, whooping cough and rubella data also produced convergent spectra and confirmed that the fourth order wavelet is an appropriate choice for comparisons of the observed time series.

Figure 8 shows that the distribution of singularities is nearly identical for the chicken pox time series from British Columbia, Saskatchewan, Manitoba, and Ontario. These results suggest that the nature of the disease itself may determine the shape of the singularity spectrum. We propose that this characteristic shape is the multifractal *signature* of the disease.

Figure 9 compares the singularity spectra for measles incidence in several widely separated geographic locations. This figure shows that measles incidence appears to have a distinctive singularity structure, regardless of location. All of the measles singularity spectra share the same qualitative features: they are all smooth on the top with a wide base (characteristic of a true multifractal statistical structure in time). This distinguishes the measles spectra from the chicken pox spectra which have pointed tops with a narrow base (suggestive of

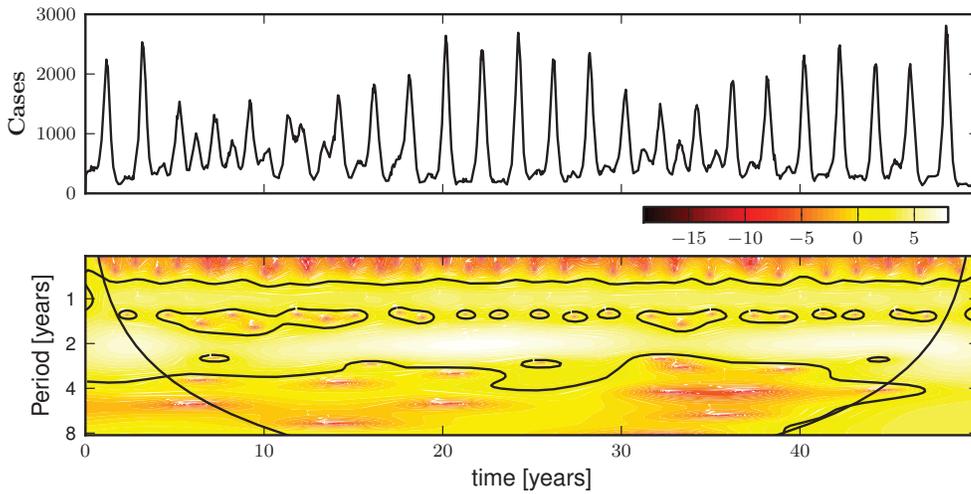


Figure 5. (a) Simulated weekly measles incidence, generated by the stochastic SEIR model with basic reproduction number $\mathcal{R}_0 = 17$, mean latent period $1/\sigma = 8$ days, mean infectious period $1/\gamma = 5$ days, seasonal amplitude $b_1 = 0.08$, and a population of $N = 2$ Million. (b) Corresponding wavelet power spectrum as in figure (1).

an approximately monofractal statistical structure in time). The scaling functions produced by the chicken pox data sets are more linear than that of measles. This indicates that chicken pox incidence is approximately monofractal while measles incidence has a true multifractal structure (recall that $\tau(q)$ is a straight line for a true monofractal). Indeed, measles incidence appears to be more singular in structure than the chicken pox incidence, i.e., the peaks of the measles singularity spectra are generally located at smaller values of α .

Figure 10 shows the same data as Figure 9, but the weekly Liverpool and London time series have been aggregated four-weekly to approximate a monthly reporting interval. This temporal aggregation shifts the peak of the spectrum to the left and narrows the base of the spectrum (for both cities). It is evident from a comparison of Figures 9 and 10 that changes in the reporting interval can have non-negligible effects on the singularity spectrum. Comparisons of different data sets should be made using the same (or a similar) reporting interval.

Extending the analysis to rubella and whooping cough reveals that each disease is characterized by a unique singularity structure as evinced by the qualitative differences between their respective singularity spectra. Figure 11 illustrates these differences between singularity spectra computed from monthly reported incidence. The qualitative shape of the whooping cough spectrum starkly contrasts the narrow pointed shape of the chicken pox spectrum and their associated scaling functions confirm that the former exhibits a broader multifractal structure. These comparisons, together with the very similar chicken pox spectra from different geographical locations, support the idea that each disease is characterized by a unique multifractal *signature*.

Comparisons of the spectra resulting from higher frequency (weekly) incidence from each of the diseases confirm the presence of such signatures and an example

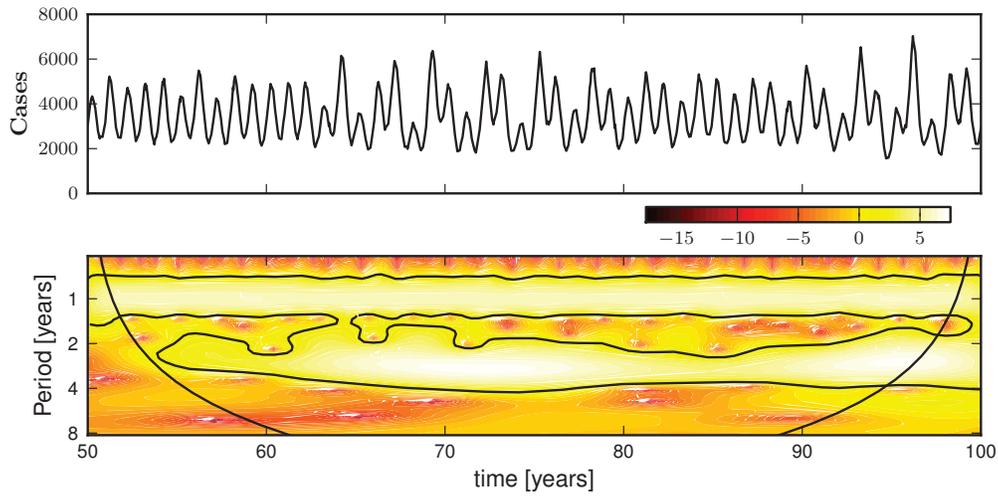


Figure 6. (a) Simulated weekly chicken pox cases, generated by the stochastic SEIR model with basic reproduction number $\mathcal{R}_0 = 10.5$, mean latent period $1/\sigma = 8$ days, mean infectious period $1/\gamma = 5$ days, seasonal amplitude $b_1 = 0.08$, and a population of $N = 10^7$. (b) Corresponding wavelet power spectrum as in figure (1).

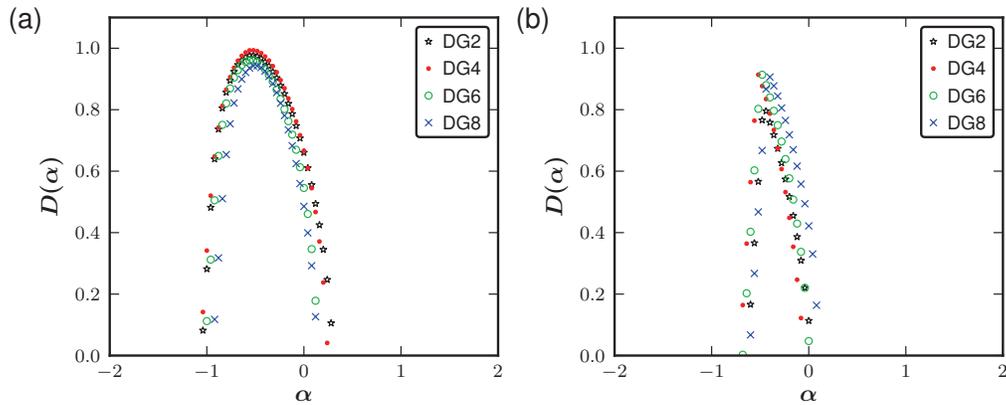


Figure 7. Wavelet choice: the multifractal singularity spectra computed from (a) measles incidence of infection data reported monthly in Ontario Canada from 1904–1989, and (b) the number of chicken pox cases reported weekly in Ontario from 1925–1958. Gaussian wavelets were tested against the data using an increasing number of vanishing moments. The order of the wavelet and thus the number of vanishing moments of the wavelet is given by the number of derivatives of the Gaussian (DG) function.

is presented in figure 12. This higher frequency (weekly) whooping cough data produces a more distinctly multifractal spectrum than was observed in the

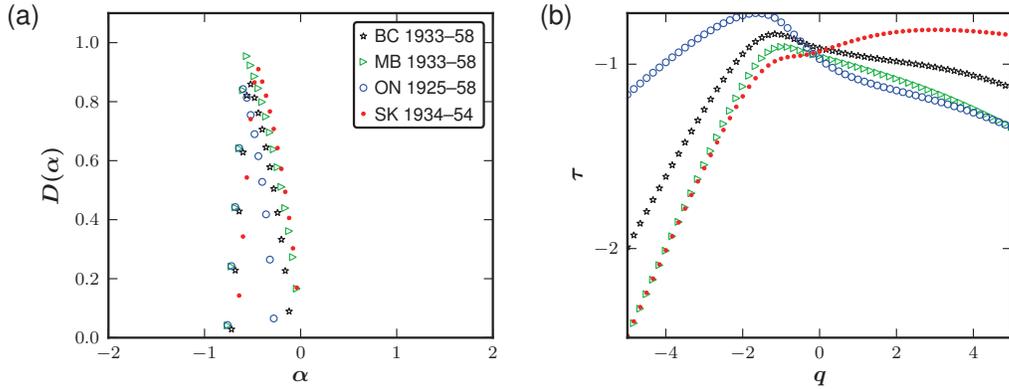


Figure 8. (a) The singularity spectra for weekly chicken pox incidence in British Columbia (BC), Manitoba (MB), Ontario (ON) and Saskatchewan (SK). (b) The associated scaling functions.

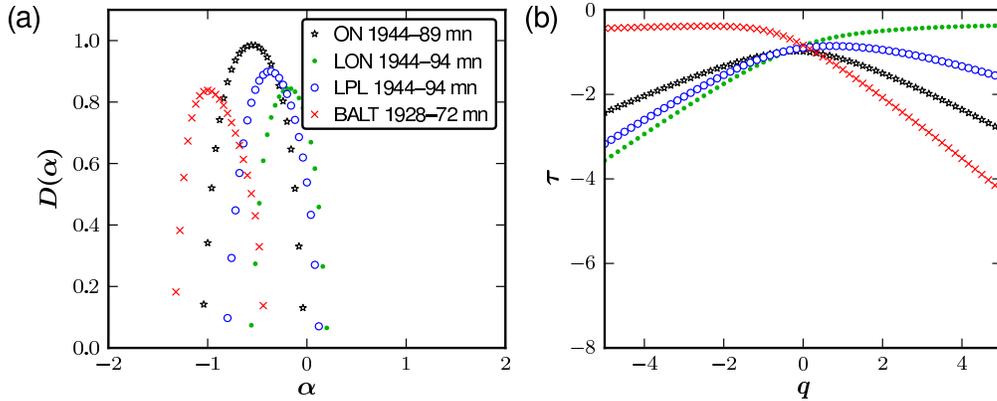


Figure 9. (a) Singularity spectra computed from reported measles incidence in Ontario Canada (ON), London England (LON), Liverpool England (LPL), and Baltimore (BALT) in the United States. (b) The associated scaling functions.

monthly case. Again, each disease appears to have a unique singularity structure that can be identified qualitatively.

We propose that the singularity spectrum provides a distinctive signature that may be used to characterize different diseases based solely on a statistical analysis of incidence time series. An accurate numerical model should be able to reproduce this multifractal signature, at least qualitatively. In section b we investigate whether the stochastic SEIR epidemic model, which is the standard mathematical model for the childhood diseases we have examined, is capable of reproducing the multifractal structure we have detected in the incidence data.

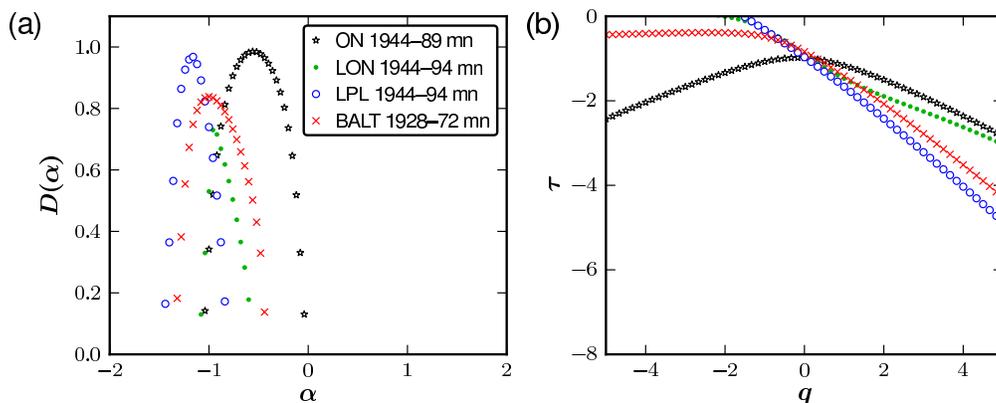


Figure 10. Effects of temporal aggregation of data on singularity spectra (a) and scaling functions (b). The weekly measles incidence data analyzed in Figure 9 were aggregated four-weekly to approximate monthly data.

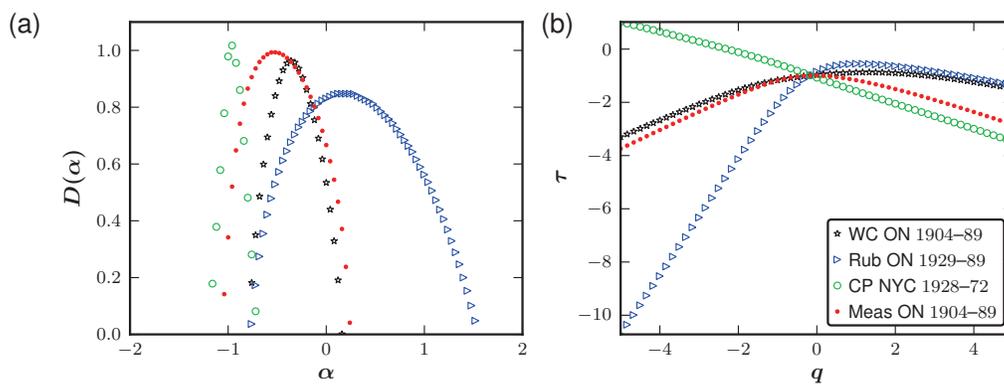


Figure 11. (a) Singularity spectra computed from monthly incidence for several different diseases: whooping cough (WC), rubella (Rub), chicken pox (CP) and measles (Meas). (b) Scaling functions.

Previous work^{4,23,28} has demonstrated that substantial changes in susceptible recruitment rates (determined by birth and vaccination rates) induce dynamical transitions. We therefore divided several time series into segments of approximately constant recruitment and recomputed the singularity spectra for each segment separately (we did this for whooping cough in London and measles in London and Liverpool). Our investigation into the effects of these dynamical transitions is restricted by the limitations of the WTMM method, which requires large numbers of data points to accurately measure the local regularity of a time series. For this reason, we considered only weekly data for this analysis.

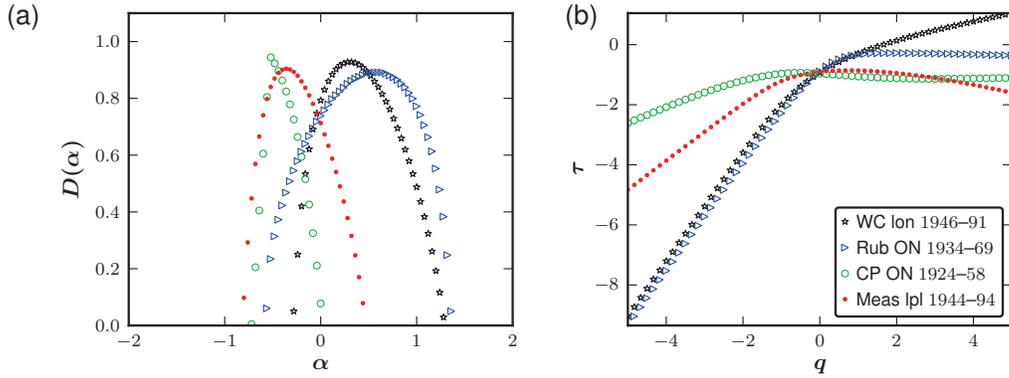


Figure 12. (a) Singularity spectra for weekly reported cases of whooping cough (WC), rubella (Rub), chicken pox (CP) and measles (Meas). (b) Scaling functions.

Changes in London whooping cough incidence were strongly influenced by the introduction of whole cell vaccination in 1957³⁰. Figure 13 shows the singularity spectrum for each period of approximately constant susceptible recruitment. The main qualitative difference is that the peak location for the full time series is more negative (i.e., more singular) than the peak locations of those spectra generated by the divided regions. Although the peak locations vary slightly, the qualitative signature of the data is similar for all three time periods.

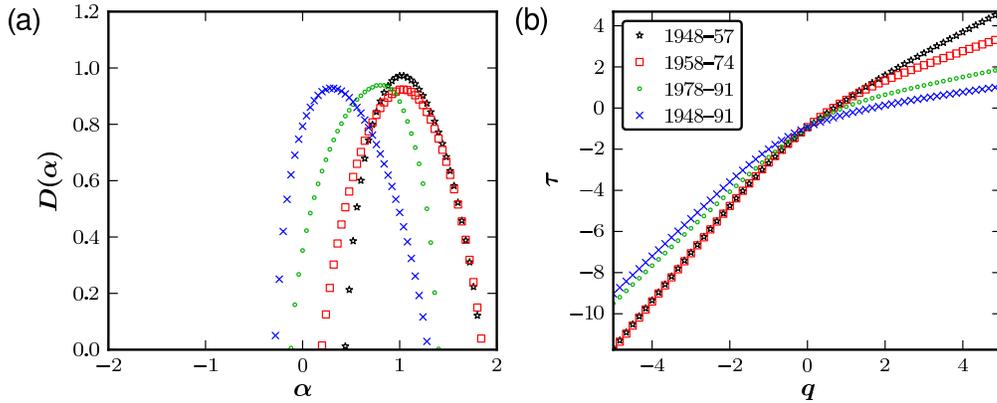


Figure 13. (a) Singularity spectrum for whooping cough in London England 1948–1991. The data are split into sections of roughly constant recruitment as described in²³. (b) The scaling function.

Figure 14 shows the singularity spectra for weekly measles incidence in (a) London and (b) Liverpool. The partition of the data was determined by the time of introduction of measles vaccine. For Liverpool, the spectra for

post-vaccination data lies further to the right (i.e., smoother) than for the pre-vaccination era. This suggests that the introduction of the vaccine caused the disease dynamics to become more regular. For London, the pre-vaccination spectrum is much broader (i.e., more multifractal) than the post-vaccination spectrum, but the peak location changes very little. This indicates that the introduction of the vaccine made the time series more monofractal (i.e., regularly irregular).

The pre-vaccination spectra from London and Liverpool are qualitatively different. The reasons for these differences are unclear, though we note that the birth rate was much higher in Liverpool and that this induced an annual cycle of epidemics in Liverpool and a biennial cycle in London⁴.

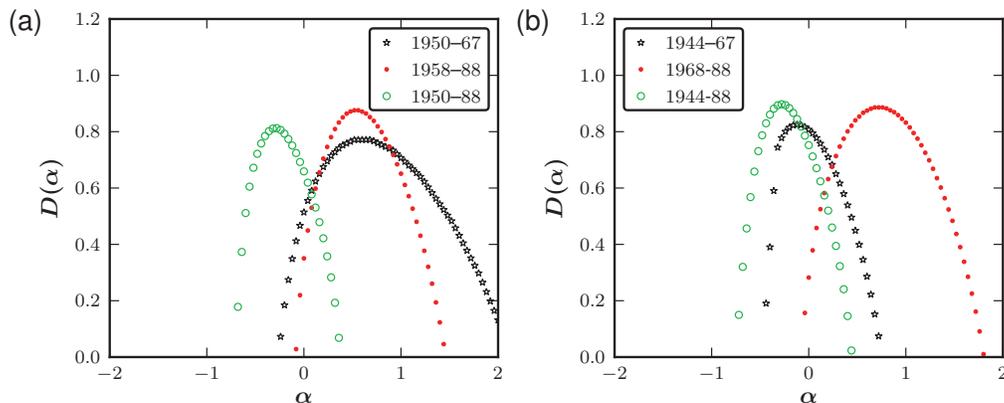


Figure 14. (a) Singularity spectrum for measles time series taken from London England 1950-1988. (b) Singularity spectrum for measles time series taken from Liverpool England 1944-1988. The data is split into sections of roughly constant recruitment as described by Bauch²³.

(b) Singularity spectra of SEIR epidemic simulations

The WTMM method was applied to incidence times series generated by the stochastic SEIR epidemic model with parameters typical of measles in the pre-vaccination era ($\mathcal{R}_0 = 17$, $1/\sigma = 8$ days, $1/\gamma = 5$ days) and a population of two million people. For comparison with the reported data, the optimal analyzing wavelet is found by computing the singularity spectra using increasing orders, n , of Gaussian analyzing wavelets.

Figure 15 shows singularity spectra for simulated measles incidence with seasonal forcing amplitude $b_1 = 0.08$. The spectrum shifts monotonically to the right as n increases from 2 to 16. The same result was found for simulations with other seasonal amplitudes b_1 . The spectrum clearly does not converge with increasing order of analyzing wavelet, and much of the DG16 spectrum extends to $\alpha \geq 1$, suggesting large subsets of the signal are in fact continuous and differentiable.

The optimal wavelet analysis was repeated for simulated chicken pox incidence with a population size of ten million ($\mathcal{R}_0 = 10.5$, $1/\sigma = 8$ days, $1/\gamma = 5$ days) in figure 16. The regularity of the simulated chickenpox data varies significantly over

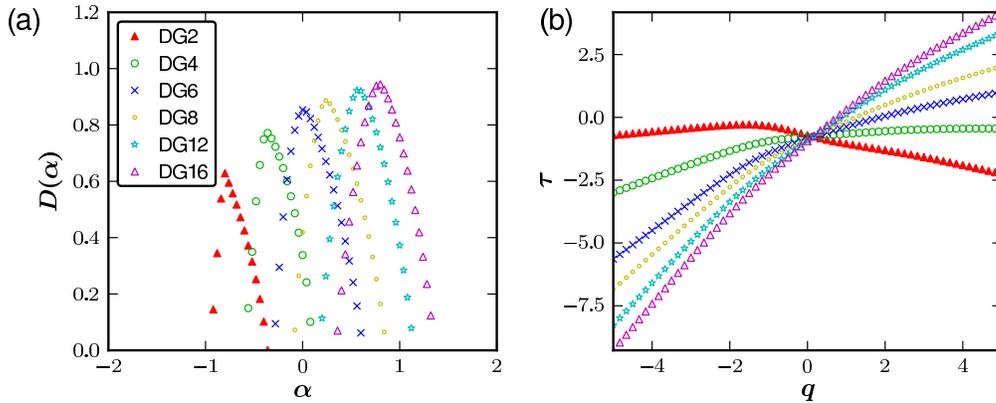


Figure 15. (a) Singularity spectra computed from measles simulations using wavelets from the Gaussian family. The amplitude of seasonal forcing was $b_1 = 0.08$ with a population of two million people. Derivatives of the Gaussian function are denoted DG and the order, or the number of derivatives taken, is the same as the number of vanishing moments of the wavelet. (b) The associated scaling functions.

time, so we have deliberately selected one of the less smooth regimes. Although the simulation spectra seem to share the pointed peak characteristic of real chicken pox data, the simulation data are much smoother. As for measles, the location of the peak of the singularity spectrum shifts monotonically to the right as the order of the Gaussian analyzing wavelet increases.

Since a wavelet of order n is capable of detecting singularities of order $\alpha < n$, the spectrum of a fractal signal is expected to converge to a unique spectrum once the highest order singularities present in the signal are resolved. Lack of convergence suggests that the simulated data are smoother than the analyzing wavelets. For all of the simulated data the scaling function becomes more linear with increasing n , which suggests that higher order singularities are not present in the signal. The linearity of the scaling function, $\tau(q)$, usually indicates a monofractal structure, but the lack of convergence provides evidence that the data are dominated by smooth sub-intervals.

Mallat⁹ proved that smooth perturbations of a multifractal signal introduce a bias in the singularity spectrum and proposed that this bias be detected by varying the order of the analyzing wavelet, n . The presence of smooth sub-intervals can inhibit the tracking of WTMM lines (discussed in section 2) causing the spectrum to vary with the order of the analyzing wavelet.

Numerical tests of SEIR simulation data for a range of model parameters have confirmed the presence of smooth sub-intervals in the simulated time series. This behaviour is qualitatively different from the singularity spectra produced by the reported data shown in Figure 7, which exhibits convergence with n . The time series generated by the stochastic SEIR model produces fewer high amplitude wavelets coefficients at the finest scales, which further demonstrates that the simulated incidence data are significantly smoother than the reported incidence data.

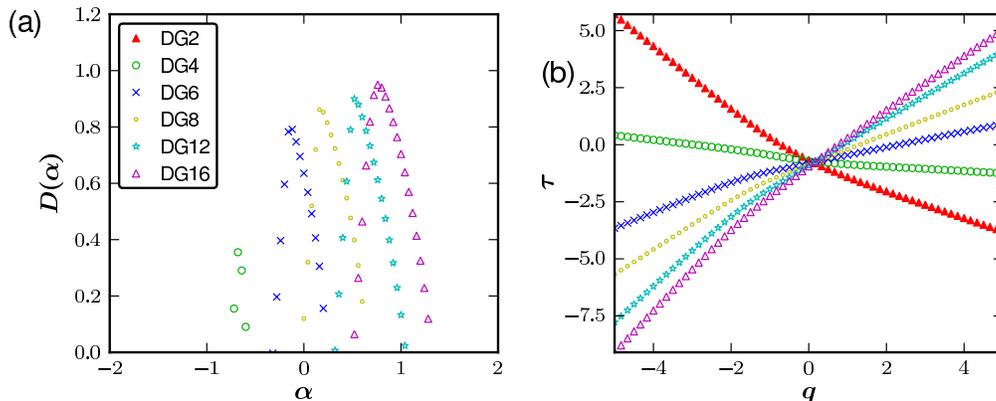


Figure 16. (a) Singularity spectra for SEIR simulations with chicken pox parameters. The amplitude of seasonal forcing was $b_1 = 0.08$ with a population of ten million people. Derivatives of the Gaussian function are denoted DG and the order, or the number of derivatives taken, is the same as the number of vanishing moments of the wavelet. (b) The associated scaling functions.

Overall, our analysis strongly suggests that the time series produced by the stochastic SEIR epidemic model are only weakly singular and are unlikely to be multifractal or even monofractal. This contrasts sharply with the reported data, which are clearly multifractal and strongly singular. It appears that the stochastic SEIR epidemic model does not capture the singular multiscale time structure that characterizes the reported incidence data.

5. Discussion and conclusions

The WTMM multifractal formalism was used to analyze time series of reported cases of measles, chicken pox, whooping cough and rubella from a variety of geographic locations. A characteristic multifractal singularity spectrum was identified for each disease. Multiple data sets from different locations corresponding to the same disease produce singularity spectra with qualitatively similar shapes, which distinguish them from spectra associated with the other diseases (with some exceptions, e.g. measles in London versus Liverpool).

Table 2 shows that weekly incidence data for a given disease at different geographical locations have similar characteristics, while Table 3 shows that weekly incidence data for different infectious diseases differ significantly in both in their irregularity and their degree of multifractality. The peak Hölder exponent α ranges from -0.6 (indicating discontinuous data) to 0.6 (indicating continuous but non-differentiable data). The degree of multifractality, measured by the width of the singularity spectrum, ranges from 0.7 to 2.0. In fact, the chicken pox data seem to be approximately monofractal (all others are clearly multifractal).

Visual comparisons of time series and conventional (periodicity) spectra of infectious disease incidence or mortality suggest that weekly or monthly counts of cases or deaths encode essential characteristics of different pathogens. However,

Disease	Peak	Width	Type
Chicken pox	-0.6 ± 0.1	0.7 ± 0.1	approximately monofractal
Measles	-0.5 ± 0.4	1.3 ± 0.1	multifractal

Table 2. Signatures of measles and chicken pox based on properties of the singularity spectra of their time series data (weekly incidence data). Results are averages from four geographic locations.

Disease	Peak	Width	Type
Chicken pox	-0.6	0.7	approximately monofractal
Measles	-0.4	1.3	multifractal
Rubella	0.6	2.0	multifractal
Whooping cough	0.3	1.6	multifractal

Table 3. Distinguishing infectious diseases based on properties of the singularity spectra of their time series data (weekly incidence data).

converting such visual impressions into a formal statistic is not straightforward and has not been attempted to our knowledge. The approach we have presented in this paper suggests that infectious diseases have *multifractal signatures*, which are relatively easy to compute and provide a useful new way to describe infectious disease data. These signatures might in certain cases permit a disease to be identified purely on the basis of a statistical analysis of its reported incidence or mortality time series. This approach could be used, in principle, to confirm the cause of historical epidemics that have been identified by analysis of ancient DNA³¹, to identify the causative agent of historical epidemics from which direct evidence cannot be obtained, or to help isolate the causes of transitions in disease dynamics that correlate with changes in the associated multifractal signature. Moreover, the WTMM method can be used as a new way of statistically validating models: accurate numerical simulations should reproduce the multifractal signatures of the diseases they are intended to model.

We used the WTMM technique to analyze simulated incidence data generated by the stochastic SEIR epidemic model. The simulated time series generated by the stochastic SEIR model were much smoother than the observed data for all of the diseases tested. In fact, by analyzing the data with Gaussian wavelets of increasing order we found that the singularity spectra do not even converge. This lack of convergence suggests that the simulated data generated by the stochastic SEIR model are dominated by smooth sub-intervals and do not capture the full multiscale structure of the real incidence data.

In particular, we found that real measles data are characterized by a broad (multifractal) singularity spectrum, but after testing a range of model parameters we concluded that the stochastic SEIR model could not reproduce such a spectrum, even qualitatively. Increasing the amplitude of seasonal forcing, b_1 , did improve the fit of the model's spectrum somewhat, but did not produce convergent spectra. We investigated the effect of parameter changes in the SEIR model and considered the effect of imperfect reporting in an attempt to better match the qualitative statistical properties of the real data, but without success.

The precise reasons behind the SEIR model's inability to produce multifractal, singular incidence data remain unclear. Possible reasons include stochastic fluctuations in the fundamental parameters of the disease (such as transmission rate), the existence of spatial hierarchies involving the interaction of urban centres of different sizes, or time varying differences between the actual incidence rate and the reported incidence rate. It is also possible that the dynamics of the stochastic SEIR model do not adequately capture the richness of the dynamical system governing the actual infectious disease transmission process.

Our analysis confirms that multiscale wavelet analysis offers powerful new tools for classifying infectious diseases based on their incidence time series, and for qualitatively comparing and fitting models to reported data. Accurate mathematical models of infectious disease epidemics are an important tool for public health decision makers, allowing them to successfully plan control and eradication campaigns. We have shown that, as measured by the singularity spectrum, the most popular epidemic model appears to miss some important multiscale time structure clearly present in the reported data. We hope that the results will lead to improved epidemic models, and a better understanding of the mechanisms underlying the spread of infectious disease.

Acknowledgements

We are grateful for support from the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

- [1] Muller CP, Kremer JR, Best JM, Dourado I, Triki H, Reef S. Reducing global disease burden of measles and rubella: Report of the WHO Steering Committee on research related to measles and rubella vaccines and vaccination, 2005 [Article]. *VACCINE*. 2007 JAN 2;25(1):1–9.
- [2] Bartlett MS. Stochastic population models in ecology and epidemiology. vol. 4 of Methuen's Monographs on Applied Probability and Statistics. London: Spottiswoode, Ballantyne & Co. Ltd.; 1960.
- [3] London W, Yorke JA. Recurrent outbreaks of measles, chickenpox and mumps. I. Seasonal variation in contact rates. *American Journal of Epidemiology*. 1973;98(6):453–468.
- [4] Earn DJD, Rohani P, Bolker BM, Grenfell BT. A simple model for complex dynamical transitions in epidemics. *Science*. 2000;287(5453):667–670.
- [5] Alonso D, McKane AJ, Pascual M. Stochastic amplification in epidemics. *J R Soc Interface*. 2007;4:575–582.
- [6] Stone L, Olinky R, Huppert A. Seasonal dynamics of recurrent epidemics. *Nature*. 2007;446(7135):533–536.
- [7] Anderson RM, May RM. *Infectious Diseases of Humans: Dynamics and Control*. Oxford: Oxford University Press; 1991.

- [8] Bacry, E, J Muzy, Arneodo A. Singularity spectrum of fractal signals from wavelet analysis: exact results. *Journ of Statistical Physics*. 1993;70.
- [9] Mallat S. *A Wavelet Tour of Signal Processing*. Academic Press; 1998.
- [10] Jaffard S. Multifractal Formalism for Functions. Part II: Self-Similar Functions. *SIAM Journal on Mathematical Analysis*. 1997 Jul;28(4):971–998. Available from: <http://epubs.siam.org/sam-bin/dbq/article/28300>.
- [11] Ivanov PC, Amaral LA, Goldberger AL, Havlin S, Rosenblum MG, Struzik ZR, et al. Multifractality in human heartbeat dynamics. *Nature*. 1999;399:461–465.
- [12] West BJ, Latka M. Fractional Langevin model of gait variability. *J NeuroEngineering and Rehabilitation*. 2005;2:24–33.
- [13] Song IH, Lee SM, Kim IY, Lee DS, Kim SI. In: *Mutifractal Analysis of Electroencephalogram Time Series in Humans*. Lecture Notes in Computer Science. Springer-Verlag; Berlin; 2005. p. 921–926.
- [14] Shimizu Y, Barth M, Windischberger C, Mose E, Thurner S. Wavelet-based multifractal analysis of fMRI time series. *Neuroimage*. 2004;22:1195–1202.
- [15] Chiu KM, Chan HL, Chu SH, Lin TY. Carvedilol can restore the multifractal properties of heart beat dynamics in patients with advanced congestive heart failure. *Autonomic Neurosci Basic and Clinical*. 2007;132:76–80.
- [16] Bauch CT. The Role of Mathematical Models in Explaining Recurrent Outbreaks of Infectious Childhood Diseases. *Mathematical Epidemiology*. 2008;p. 297 – 319.
- [17] B T Grenfell JK O N Bjornstad. Travelling waves and spatial hierarchies in measles epidemics. *Nature*. 2001;414:716–723.
- [18] Torrence C, Compo GP. A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society*. 1998;4(3):415–447. Available from: <http://www.inference.phy.cam.ac.uk/mackay/PhD.html>.
- [19] Hurst HE. Long-term storage capacity of reservoirs. *Trans Am Soc Civ Eng*. 1951;116:770–808.
- [20] Donoho D, Maleki A, Shahram M. WaveLab; 2007. www-stat.stanford.edu/~wavelab.
- [21] Mallat S, Hwang WL. Singularity detection and processing with wavelets. *IEEE Trans Inf Th*. 1992;38:617–643.
- [22] IIDDA. The International Infectious Disease Data Archive; <http://iidda.mcmaster.ca>.

- [23] Bauch C, Earn D. Transients and attractors in epidemics [Article]. Proceedings of the Royal Society of London Series B–Biological Sciences. 2003 AUG 7;270(1524):1573–1578.
- [24] Bailey NTJ. The Mathematical Theory of Infectious Diseases and its Applications. 2nd ed. New York: Hafner Press; 1975.
- [25] Kurtz TG. Relationships between stochastic and deterministic population models. Lecture Notes in Biomathematics. 1980;38:449–467.
- [26] Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London Series A. 1927;115:700–721.
- [27] He D, Earn DJD. Epidemiological effects of seasonal oscillations in birth rates. Theoretical Population Biology. 2007;72:274–291.
- [28] Earn DJD. Mathematical epidemiology of infectious diseases. In: Lewis MA, Chaplain MAJ, Keener JP, Maini PK, editors. Mathematical Biology. vol. 14 of IAS/ Park City Mathematics Series. American Mathematical Society; 2009. p. 151–186.
- [29] Gillespie DT. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational Physics. 1976;22:403–434.
- [30] Rohani P, Earn D, Grenfell B. Opposite patterns of synchrony in sympatric disease metapopulations [Article]. SCIENCE. 1999 OCT 29;286(5441):968–971.
- [31] Bos KI, Schuenemann VJ, Golding GB, Burbano HA, Waglechner N, Coombes BK, McPhee JB, DeWitte SN, Meyer M, Schmedes S, Wood J, Earn DJD, Herring DA, Bauer P, Poinar HN, Krause J. A draft genome of *Yersinia pestis* from victims of the Black Death. Nature. 2011;478:506–510.