Published in final edited form as:
J Geogr Syst. 2006 September ; 8(3): 227–252. doi:10.1007/s10109-006-0027-8.

A swash–backwash model of the single epidemic wave

Andrew. D. Cliff and
Department of Geography, University of Cambridge, Downing Place, Cambridge, CB2 3EN, UK

Peter Haggett
Institute for Advanced Studies, University of Bristol, Royal Fort House, Bristol, BS8 1UJ, UK, E-mail: p.haggett@bristol.ac.uk

Abstract

While there is a large literature on the form of epidemic waves in the time domain, models of their structure and shape in the spatial domain remain poorly developed. This paper concentrates on the changing spatial distribution of an epidemic wave over time and presents a simple method for identifying the leading and trailing edges of the spatial advance and retreat of such waves. Analysis of edge characteristics is used to (a) disaggregate waves into ‘swash’ and ‘backwash’ stages, (b) measure the phase transitions of areas from susceptible, *S*, through infective, *I*, to recovered, *R*, status ($S \rightarrow I \rightarrow R$) as dimensionless integrals and (c) estimate a spatial version of the basic reproduction number, R_0. The methods used are illustrated by application to measles waves in Iceland over a 60-year period from 1915 to 1974. Extensions of the methods for use with more complex waves are possible through modifying the threshold values used to define the start and end points of an event.

Keywords

Spatial diffusion; Waves; Epidemics; Measles; Iceland

1 Introduction

The shape of epidemics has fascinated both epidemiologists and mathematicians for several generations. As early as 1840, Farr had effectively fitted a normal curve to smoothed quarterly data of deaths from smallpox in London and, in a series of papers between 1909 and 1918, Brownlee fitted various Pearson curves to epidemic data on a wide range communicable diseases including measles and scarlet fever. Bailey (1975, pp. 9–19) provides a useful historical review of these early papers. The deterministic and stochastic models used to describe such wave-like phenomena have recently been drawn together by Daley and Gani (1999). But, whereas the temporal structure of waves has received detailed treatment, the spatial form taken by epidemic waves as they travel over geographical space has received less attention. Exceptions to this generalization come in the early work of Morrill (1968, 1970) who reported on the changing shape of spatial diffusion waves at different stages of the spread process. Much of our own earlier work has also attempted to restore the time/space imbalance (e.g. Cliff et al. 1981, 1986), and the spatial characteristics of travelling waves are now occasionally incorporated into epidemic models (e.g. Abramson et al. 2003; Grenfell et al. 2001; Zhao and Wang 2004).

In this paper we concentrate on the changing spatial distribution of the wave over time. We look first at a method for disaggregating waves using spatial extent rather than number of

E-mail: adc2@cam.ac.uk.

Europe PMC Funders Author Manuscripts

reported cases as our criteria. Their spatial spread is measured in terms of (a) leading edges as new areas are infected and (b) following edges as infected areas become clear of infection through either spatial retreat or decay in situ. Statistical analysis of the time distribution of the two edges allows the wave to be divided into distinctive swash and backwash stages and permits the phase transition of areas from susceptible, through infected, to recovered status to be measured as a series of dimensionless integrals. The terms swash and backwash are based upon a hydrological analogy with the distinctive stages of waves on a beach, and are fully defined in Sect. 3. The methods are applied to a set of historical epidemic waves, the 14 measles waves that ran through the population of Iceland in the 60-year period from January 1915 to December 1974. Finally, we discuss the results and consider their significance for forecasting and for devising early-warning systems.

2 Epidemics as spatial diffusion waves

Descriptions of many thousands of outbreaks of infectious diseases are extant in the epidemiological literature. Many of these vividly describe the process of spread of the disease as a spatial diffusion wave, one usually starting in one or a few locations and spreading out through various contagious mechanisms to infect an ever wider area. This spread process is set into reverse as populations recover and previously-infected areas are cleared. A summary of the mechanisms involved and examples of such spread are given elsewhere (Haggett 2000; Cliff et al. 2000). A wide selection of epidemics at different geographical scales, from local outbreaks to global pandemics, and for over 50 infectious diseases and over several centuries of record has recently been published as a World Atlas of Epidemic Diseases (Cliff et al. 2004).

From these many descriptions, both qualitative and quantitative, some understanding of the spatial structure of a typical epidemic can be built up. Figure 1 presents an idealized epidemic wave in the spatial domain with infection being introduced into an island (shaded). It suggests that a wave's evolution over time may go through several distinctive spatial forms, conceived here as falling into five stages:

- **a.** *Onset*: The introduction of an infectious agent (virus, bacteria, etc.) into a new area with a susceptible population which is open to infection. Typically, this initial invasion occurs either in one or in a small number of locations.
- **b.** *Youth*: A very active stage marked by the rapid spread of infection from the original point of introduction to the main population centres. Historical studies suggest that this may occur through a mix of both local (neighbourhood) diffusion and by long-range (cascade) diffusion.
- **c.** *Maturity*: In this central stage, the epidemic rages at its highest intensity with strong clusters of cases widely spread through the susceptible population. In Fig. 1c this is shown as affecting the whole island. Here the epidemic is at its maximum intensity with strong contrasts in infection density between the various sub-areas.
- **d.** *Decay*: The peak is usually followed by a longer period with rapidly declining intensity (and thus fewer reported cases) and less rapid spatial contraction. Typically, the spatial structure tends to be less coherent than in earlier stages with a scatter of low-intensity infected areas.
- **e.** *Extinction*: The last throws of the epidemic wave with a final scatter of cases. Spatially, such cases often tend to be found in the less accessible areas.

The sequence in Fig. 1 is not intended as a sufficient summary of spatial epidemic behaviour. Rather it serves as a visual marker against which the realism of the outputs from quantitative models can be checked and refined.

The two-dimensional pattern shown in Fig. 1 has obvious parallels with the normal and logistic curves in the temporal domain. The spatially skewed shape, with the loss of the left-hand tail, is probably due to multiple infectives triggering the epidemic, while the augmented right-hand tail may be due to the slowing down of infective–susceptible contacts arising from both the increasing presence of recovereds in the population and the impact of public health measures.

3 Methods of wave decomposition

In order to test whether this speculative sketch of wave behaviour is robust, we need to develop some quantitative measures of wave patterns and to test them against examples of real epidemic waves. In this first section we modify standard statistical methods to provide some elementary measures of spatial form and then illustrate their application through a small hypothetical example. We apply these methods to real waves drawn from the historical record of epidemics in the next section.

3.1 Analysis of a hypothetical wave

We begin with the observation that a single epidemic wave in any large geographical area is a composite of the waves for each of its constituent sub-areas. That is, the composite wave at the larger geographical scale (say, a country) can in principle be broken down into a series of multiple waves for its constituent sub-areas (say, its regions) at the smaller geographical scale. We say, in principle, because in practice data for sub-areas is often ephemeral and sub-area records can be hard to obtain.

Assuming that both the sub-areas and the time periods are discrete, we use the following notation:

A Area covered by an epidemic wave in terms of the number of sub-areas infected where a_i is a sub-area in the sequence 1, 2,…, a_i, …, A.

T Duration of an epidemic wave, defined in terms of a number of discrete time periods, t_j, in the sequence 1, 2, …, t_j, …, T.

Q Total cases of a disease recorded in a single epidemic wave measured over all sub-areas and all time periods. Thus q_{ij} is the number of cases recorded in the cell formed by the i-th sub-area and the j-th time period of the $A \times T$ data matrix.

To illustrate the method, we assume in Fig. 2 a simple epidemic wave where $A = 12$, $T = 10$ and $Q = 122$. Thus in Fig. 2a we begin with a 12×10 space–time data matrix in which 122 recorded cases of a disease are distributed to simulate an array typical of an epidemic wave. Note that, whereas this overall wave is continuous (no time periods with zero cases), for individual sub-areas the record may be discontinuous with one or more time periods with zero cases. We make use of these ‘internal’ zero cells as a measure of eddies within the overall wave later in the paper. The overall wave is given by the marginal summation of cases for each time period.

The data in matrix (a) can be analyzed in both the time and space domains. Figure 2b shows the cases as a time series using the column sums, Sum q. The time series of the number of areas infected appears in Fig. 2c and is given by Sum a. In this exemplar, the case distribution is more peaked and more left-skewed than the distribution of areas. The spatial narrative of the outbreak is shown in (d) with a series of ten maps, one for each of the time periods. The shading on the map is the same as that in the original data matrix.

For any one of the rows in the data matrix in Fig. 2a, two cells can be identified which mark the 'start cell' and the 'end cell' of a recorded outbreak; if the infection only lasts for one time period, the start and end cell are the same. Figure 3 analyzes these start and end cells. In Fig. 3a, the 12×10 matrix is re-arranged so as to position the start cells in an ascending temporal order. This line of cells (dark shading) defines the position of the leading edge (LE) marking the start of the epidemic wave in the different sub-areas. To the left and above this line lies a zone of cells (light shading) which have yet to be infected and thus may be regarded as areas to which the epidemic has yet to spread.

Equally, the 12×10 matrix can be organized as in Fig. 3b, so as to arrange the end cells in ascending temporal order. This line of cells (dark shading) defines the position of the trailing edge or following edge (FE), marking the completion of the epidemic waves in each of the different sub-areas. To the right and below this line lies a zone of cells which have ceased to be infected and which thus may be regarded as areas which have recovered from infection.

Both the edges, LE and FE, can be combined as in Fig. 3c to define cells which are in a susceptible, S, infected, I and recovered, R, states. The resulting graph may be regarded as a phase transition diagram. It has two roles: first, it defines the boundaries of the two critical phase shifts from susceptible to infective status ($S \rightarrow I$) and from infective to recovered status ($I \rightarrow R$) and second, it integrates the three phases, S, I and R, as areas within the graph. Here, we use S, I and R in the conventional way adopted in epidemiological models (e.g. Bailey 1975) but with the essential difference that our population is made up of areas (S_A, I_A and R_A) rather than individuals. Thus all measurements are in terms of cells a_{ij} and the numbers refer to the proportion of the total area occupied by each state, so that

$$S_A + I_A + R_A = 1.00 = A. \quad (1)$$

The diagram draws on the geomorphological concept of the hypsometric integral introduced by Strahler (1952) in his analysis of height distribution of erosional topography within river catchments. The epidemiological significance of the integral values is discussed later in this paper.

3.2 Statistical analysis of wave characteristics

The method followed in our simple 12×10 example was essentially geometrical and it was used largely for its pedagogic purpose. In practice and when studying large epidemic waves, standard statistical analysis of the distribution of cases (Q), areas (A), and the two edges (LE and FE) can be substituted to yield the same results more economically. The arithmetic mean $\bar{t}$ of all four characteristics can be computed as a time-weighted mean. For cases, Q, the equation is

$$\bar{t}_Q = \frac{1}{n} \sum_{t=1}^{T} t\, q_t, \quad (2)$$

where q_t is the number of cases notified at time t and $n = \sum q_t$ for all t. The first time period, say month, of an epidemic is coded as $t = 1$. The subsequent months of the epidemic may then be coded serially as $t = 2, 3, \ldots, T$, where T is the number of monthly periods from the beginning to the end of the epidemic (i.e. $T = 10$ in our example). The time-weighted means yield not only a useful measure of central tendency but of the velocity of the wave in terms of time to infection. Similar measures can be made for areas, leading edges and following edges. The values for $\bar{t}_Q$ and $\bar{t}_A$ are plotted on the time series in Fig. 2b and c, and $\bar{t}_{\mathrm{LE}}$ and $\bar{t}_{\mathrm{FE}}$ on Fig. 3d. They show the epidemic wave for cases was faster than that for areas; this

implies a relatively spatially concentrated epidemic. Higher moments about the mean can also be computed to yield valuable measures of epidemiological behaviour. The r-th central moment about $\bar{t}$ for cases, q, may be written as

$$m_r = \frac{1}{n}\sum_{t=1}^{T}[(t-\bar{t})^r q_t]. \quad (3)$$

Using Eq. 3 we may then define further measures of velocity as

$$s = \sqrt{m_2} \quad (4)$$

and

$$b_1 = \frac{m_3^2}{m_2^3} \quad \text{and} \quad b_2 = \frac{m_4}{m_2^2}. \quad (5)$$

The quantity s defined in Eq. 4 is the familiar standard deviation of the frequency of cases against time, while b_1 and b_2 in Eq. 5 are the Pearson measures of skewness and kurtosis, respectively. Skewness indicates the tendency of the distribution to peak before or after the mean, and kurtosis a gauge for whether the distribution clusters more closely or less closely than would be expected under the normal curve. Further details are given in Cliff et al. (1986, pp. 200–201). Relatively fast epidemic waves are characterized by a small (low) average time to infection $\bar{t}_Q$, a small standard deviation, a positive value for the coefficient of skewness and a large value of the coefficient of kurtosis. Conversely, relatively slow epidemic waves are characterized by a large (high) value of $\bar{t}_Q$ and s, a negative value for b_1 and a small value for b_2. A qualitative interpretation of the case distribution associated with fast and slow moving epidemic waves, and to which the moment values refer, is given by Trevelyan et al. (2005, p. 286, Table 6). Gilg (1973) has used these parameters to study the spread of an epizootic wave in England and Wales, and Cliff and Haggett (1981) have studied changing wave velocities for measles in Iceland.

3.3 *S*, *I* and *R* integrals

The means of the two wave edges can be used to generate estimates of the susceptible, infected and recovered integrals within the phase diagram of Fig. 3c. The susceptible areas integral in the spatial domain can be measured as

$$S_A = \frac{(\bar{t}_{\mathrm{LE}} - 1)}{T} \quad (6)$$

while infected areas integral in the spatial domain measured as

$$I_A = \frac{\bar{t}_{\mathrm{FE}}}{T} - S_A. \quad (7)$$

and the recovered areas integral as

$$R_A = 1 - (S_A + I_A). \quad (8)$$

All three integrals are dimensionless numbers with values in the range [0, 1]. Equations 6–8 can clearly be rearranged to obtain expressions for $\bar{t}_{LE}$ and $\bar{t}_{FE}$ in terms of I_A, S_A and R_A.

3.4 Swash and backwash stages

The means of the two edges, $\bar{t}_{LE}$ and $\bar{t}_{FE}$, can also be used to throw light on the changing spatial structure of the epidemic wave over time. Figure 3d uses the data in the 12×10 example in (b) and (c) to plot the changing position of the leading-edge cells and following-edge cells over time. In the early stages newly infected sub-areas are being added and in the later stages sub-areas are subtracted as their local outbreaks come to an end and no further cases are recorded. The shaded areas show the number of cells added and subtracted and the heavy line the net balance of areas which is positive in the earlier time period and negative in the later period. The positions of the two means are shown on this graph. The median position between them, λ_t say, broadly separates the early-positive from the later-negative stages.

Using the results, we define in Fig. 3e a two-stage swash–backwash model of the spatial spread of a single epidemic wave. The terms ‘swash’ and ‘backwash’ are borrowed from coastal geomorphology (Chorley et al. 1984) where the dynamics of waves breaking on a shoreline are described as going through two distinctive stages. In the swash (or onset) stage the wave rushes up the beach occupying a larger area of sand with the direction of the wavefront determined by the direction of the approaching wave trains in respect of the shoreline. In the backwash (or withdrawal) stage the wave retreats down the beach relinquishing the area formerly occupied but with the direction of retreat determined by the slope of the beach.

In epidemiology, the context is clearly different but the analogy of a wave occupying and then retreating from an infected area remains useful. As set out in Fig. 3e, two stages are envisaged. In the swash stage, the spatial extent of infected areas is generally increasing in area. This can be operationally defined as the period between the start of an outbreak ($t = 0.5$)[1] and the midway position, λ_t, between the means of the leading and following edges. The second backwash stage is one in which the spatial extent of infected areas is generally reducing in area. This can be operationally defined as the period between the midway position, λ_t, the leading and following edges and the end of the outbreak ($t = T + 0.5$). The two stages are directly linked to the leading and following edges and thus to the three integrals defined earlier. In fast-moving epidemic waves with high initial rates of infection, the swash stage will be shorter and the infected integral I_A smaller. Conversely in slow-moving waves with low initial rates of infection, the swash stage will be longer and the recovered integral R_A smaller.

3.5 Spatial versions of the basic reproduction number, R_0

The basic reproduction number[2], R_0, is one of the most useful parameters used to characterize mathematically infectious disease processes. R_0 is defined as the ratio between an infection rate (β) and a recovery rate (γ):

[1]The 0.5 and the $T + 0.5$ in Stage IV, later, arise because time is measured as a series of integers (1, 2, 3, …), whereas the means are measured on a continuous scale.
[2]Sometimes called the basic reproduction rate or ratio.

$$R_0 = \frac{\beta}{\gamma}. \quad (9)$$

In terms of cases, Q, R_0 is thus the average number of secondary infections produced when one infected individual is introduced into a host virgin population. Methods for estimating the basic reproduction number for infectious diseases are considered by Dietz (1993), and an example of their use is given in Watts et al. (2005).

We noted earlier that the S, I and R integrals define the boundaries of the two phase shifts from susceptible to infective status ($S \rightarrow I$) and from infective to recovered status ($I \rightarrow R$). This raises the prospect of defining a spatial version of R_0. In spatial terms, A, the spatial reproduction number, R_{0A}, is the average number of secondary districts produced from one infected district in a virgin area. But the integral S has parallels to β in that a small value indicates a very rapid spread while the integral R has parallels to γ in that a small value indicates a very rapid recovery. Since both terms are inversely related to their power we suggest that their complement might be substituted in estimating a spatial version of R_0, namely

$$R_{0A} = \frac{1 - S_A}{1 - R_A}. \quad (10)$$

Such a spatial reproduction number would measure the propensity of an infected district to spawn other infected districts in later time periods. In effect, it would provide an indicator of the tendency of an infected district to produce secondaries. Values over unity imply a tendency to spread and calibrate the velocity of such spread (the larger the value, the greater the rate of spread). For our small 12×10 example, the value of R_{0A} = 1.14 indicating a relatively low rate of spread.

4 Application to historical epidemic data

So far the analysis has been conducted in terms of a small, idealized epidemic wave and methods have been put forward to allow its measurement and analysis. But how far do the methods work when applied to actual epidemic waves of infectious diseases? These may be several orders of magnitude larger and considerably more complex in their spread. To conduct a more realistic test we follow in the classic footsteps of Bailey (1953) in using outbreaks of the highly contagious human virus disease, measles, for our epidemic modelling. The full rationale for the use of measles in such modelling has been dealt with elsewhere (Cliff et al. 1993, pp. 4–9).

The specific example chosen for analysis is that of measles outbreaks on the island of Iceland over a 60-year period from January 1915 to December 1974. Iceland has long been recognized as having exceptional epidemiological records, and these have been subject to a series of studies over the years. The summary characteristics of the 14 waves analyzed here are given in Table 1. They show that wave duration was generally less than 2 years, with the spatial extent tending to be widespread; roughly nine out of ten of the 50 districts became infected in each wave. Typically, some 6,000 cases were reported for each wave with, on average, a rate of around 40 per 1,000 the Icelandic population becoming infected. The gap between epidemic waves was on average around 3 years. Whereas the first three characters in Table 1 do not show significant trends over time, the remaining two are highly time dependant. The five early waves in the period up to 1945 show much higher attack rates and longer gaps between waves; the nine later waves tend to have lower attack rates and are bunched closer together.

Europe PMC Funders Author Manuscripts

A detailed description of the waves based on original data in the annual volumes of Heilbrigđisskýrslur (Public Health in Iceland) and on doctors' accounts has already been published in monograph form (Cliff et al. 1981, account: 59–88; data tables: 201–229) so will not be repeated here. Figure 4 shows the time series with the original monthly data aggregated into 3-month (quarterly) format on the *x*-axis of the graph. The *y*-axis shows not the number of reported cases but the spatial extent of the epidemic waves in each quarter in terms of the number of Icelandic medical districts reporting cases. Waves I and II (in 1904 and in 1907–1908, respectively) lack specific spatial detail for analysis. So, here, we begin our record in 1915 but retain the original wave numbering and look at our first wave (Wave III) in some detail and then at comparative results for the full set of waves.

4.1 Results for a single epidemic wave (Wave III)

The first of the Icelandic measles waves to be analyzed was triggered by the arrival in Reykjavík harbour in April 1916 of a ship from Norway with a sailor still infectious with measles. It came nearly 8 years after its predecessor. A scatter of measles notifications had been reported over the intervening years (56 cases in all) but preventive measures were swiftly enforced by local doctors, secondary cases were rare and no major epidemics had occurred. Wave III, when fully developed, was a major epidemiological event with 4,944 reported cases (an attack rate of 55 per 1,000 population) and 118 deaths (a rate of 23.9 deaths per 1,000 measles cases). It lasted for 14 months, peaked in July 1916, only 4 months after onset and spread widely over the island infecting all but seven of Iceland's 50 medical districts.

The detailed pattern of its geographical spread is shown in Fig. 5. For simplicity the original monthly figures have been aggregated into 3-month groups based on conventional northern hemisphere seasons (e.g. spring = March, April, May). Cases are shown by concentric circles but Reykjavíkur district had so many more than other districts that it is shown by an open circle. Districts with 40 or more reported cases are named. Starting in Reykjavík, the disease spread rapidly into northwestern districts and by the late summer of 1916 was widely dispersed around the island. From this high point the disease contracted in both its intensity and spatial extent over the autumn and winter finally petering out in the spring of 1917 with its last major strongholds in Akureyar and Húsavíkur.

The form of the epidemic in the time domain is shown in Fig. 6. We begin with the monthly time series for the number of cases reported and for the number of districts infected: both are shown in Fig. 6a. In each case the curves are strongly skewed to the left indicating a rapid build-up of both cases and of infected districts. The monthly time series of density of reported cases per infected district is shown as an inset in Fig. 6b and confirms both early peaking based on Reykjavíkur and a late but smaller surge based on Akureyrar.

Monthly time series of the number of districts newly infected (i.e. when measles cases were first recorded) and of the number of districts where infection terminated (i.e. when measles cases were last recorded) are shown in Fig. 6c. Differences between the two series plotted in (c) are illustrated in Fig. 6d. Note how the differences between the leading and following edges follow the characteristic curves sketched in Fig. 3d and e. The mean values for the two series are used to define the swash and backwash stages of Wave III in Fig. 6d. Finally, cumulative curves for the number of newly infected and newly terminated districts used to define susceptible, infective and recovered districts appear in Fig. 6e.

4.2 Results for a train of epidemic events (Waves III–XVI)

How far are the results for the first wave (Wave III) representative of the whole 60-year sequence of waves? To give an overall picture, we summarize in Table 2 the range and

average characteristics of the wave train in terms of their spatial structure. The average for the leading edges tends to fall in the six month of an outbreak with the following wave towards the end of the year. In terms of the three integrals the susceptibles make up around one quarter, the infectives around a third and the recovered the residual space (around 40%) of the phase diagram.

That these averages are drawn from a highly volatile series of epidemic behaviour can be seen by looking at the individual traces for all 14 epidemics. Thus Fig. 7 contrasts the average monthly values for the leading and following edges against the individual values (inset graph in Fig. 7) from which the average curve has been calculated.

In the same manner, Fig. 8 presents the cumulative curves for 14 separate measles epidemic waves in Iceland over the 60-year period for both newly-infected districts and newly-terminated districts. In this instance, both the spatial and dimensional axes have been standardized to unit length to allow waves of different durations and spatial extents to be shown comparatively. The results show graphically that the cumulative curves for the newly infected districts (Fig. 8a) are steeper than those for newly terminated districts (Fig. 8b), confirming that the swash stage of these waves have collectively moved faster than the backwash stages.

Analysis at the individual wave level is given in Fig. 9 which shows the superimposed cumulative curves from the previous figure to define susceptible, infective and recovered zones when plotted on a phase diagram. The numbers are integrals which measure the area under each of the three curves as a proportion of the total $A \times T$ space for each wave. Comparative wave data are also given in Fig. 10 which plots the three dimensions of the susceptible, infective and recovered phases for each the 14 epidemic waves as a ternary diagram. It is notable that the early waves (shaded) lie broadly to the left of the later, post Second World War, waves on the susceptible phase axis of the ternary diagram. This implies that the early waves showed a more rapid transition from susceptible to infected phases than later waves.

4.3 The velocity of the leading and following waves

In Fig. 3b we presented the idea of a swash–backwash model in terms of a sine–cosine wave. But the values in Table 2 and the foregoing diagrams point to pronounced asymmetry in the Icelandic data the swash wave being shorter and steeper than the backwash wave. This would imply in turn that the velocity of the infection wave tends to be slightly faster than that of the recovery wave.

For a fast-moving wave, the average time to an infected area will be relatively small and for a slow-moving wave it is relatively large. Table 3 confirms that the leading edge is very significantly faster whether measured in months (nearly three months faster) or when standardized by the overall duration of the epidemic wave (14% faster). A fast epidemic wave will display a low dispersion, with most districts infected within a short time of one another; a slow wave will display reverse characteristics. Here the evidence is less clear with only two of the three measures of dispersion (the non-parametric measures of range and inter-quartile range, IQR) showing statistically significant differences with a faster leading edge. The third measure, the standard deviation, showed little difference possibly reflecting the influence of a few extreme observations. A fast epidemic wave will display a right-skewed distribution with the majority of cases occurring in the early phases of the epidemic and positive b_1 values. Here both leading and following edges show positive values but the differences are in the expected direction. Finally, a fast epidemic wave will show a peaked distribution with a high concentration of districts around the mean and vice versa. The standard measure of such peaking (kurtosis, b_2) shows that leading edges have, as expected,

higher values than the following edges but the differences are small and not statistically significant. Overall, the four standard measures of epidemic velocity suggest that the Icelandic waves we have studied show a distinct tendency to have relatively faster leading edges and relatively slower following edges.

4.4 Epidemiological significance of swash–backwash stages

Identification of the swash and backwash stages for the 14 waves yields some interesting results which are summarized in Fig. 11. The duration of the stages averages around 8 months for swash and 11 months for backwash. Despite its shorter length, the share of reported cases is much greater in the swash period, averaging 72%, with the balance of 28% in the longer backwash stage. The intensity of cases is much greater in the swash period with 550 cases per month, four times greater than in the backwash stage (140 per month).

The contrast between the swash and backwash stages also applies to wave continuity. One in four of all the cells within the four waves have zero readings (i.e. months in which a district which is experiencing an epidemic fails to report measles cases). Over the 60 years one quarter of all epidemic cells was zero. But the incidence of such cells is 1.5 times higher in the backwash as compared to the swash stage. We interpret this distribution as indicating that the continuity of the waves for the individual districts tends to break down as the epidemic progresses with small eddies within the overall wave.

For the 14 Icelandic waves, the average value for R_{0A} was found to be 1.28. Values for individual waves ranged from 1.88 (Wave XIV) to 0.89 (Wave IX). For 12 of the waves, $R_{0A} \geq 1$. The two unexpected results were produced by waves IX and XII which were unusual in being right-skewed, with a slow spatial build-up of infected districts.

4.5 Stationarity over time

A question arises on the stationarity of the Icelandic waves. Is there any evidence from the analysis that the epidemic waves are changing systematically over time or may they be regarded as samples drawn from a relatively stable wave population? Evidence can be gained from two sources, internal and external.

We can test the internal evidence for change by plotting some of the characteristics of each wave against its location on the time line between January 1915 and December 1974. However, a series of linear regression equations in which a wave characteristic is plotted as a function of historical time failed to show significant trends and we have to conclude that the waves show only minor changes over the 60-year period and may be regarded as being drawn from a (quasi)-stationary process.

Such internal evidence stands in contrast to the external evidence. Over this same 60 years, the population of Iceland was evolving rapidly. It more than doubled in total size from 80,000 to 200,000. Its spatial distribution changed, with major growth in the capital city, Reykjavík, and out-migration from its more remote communities. This changing epidemiological environment is illustrated in Fig. 12a which plots the decreasing gaps in time between the 14 waves. The best-fit cubic regression shows a two-thirds drop over the 60 years with an R^2 more than four times that for the strongest internal trend.

There is some evidence that these historical changes are reflected in wave characteristics. Figure 12b shows a surface plot of the time gap in months from the end of the previous epidemic wave against values for susceptible and infective phases. A long gap between waves seems to be followed by a steep swash wave with a low susceptibility integral. The infectious integral appears to be unaffected by this factor.

5 Discussion

Both the original swash–backwash model and its application to an historical set of epidemic waves for a given infectious disease raise a number of questions. We look here at three: the strengths of the swash–backwash model, its limitations and possibilities for extension.

The swash–backwash model provides an analytical tool for statistically describing the spread of infection through space and time. The method stands midway between modelling infection spread using difference or partial differential equations and descriptive mapping of disease dynamics. The method is considerably less data intensive than many other epidemic models, requiring only presence and absence of disease in each region, while detailed knowledge of the mechanics of the spread of infection (e.g. infectivity, susceptibility, contact patterns) is not needed. As such it is likely to be of greatest use in areas with fewer resources for public health surveillance or where rapid exploratory analysis is required. The eight Icelandic examples used here were computed by hand (a somewhat tedious process) but a computerized version based on morbidity data stored in readily available EXCEL tables is being prepared which should make the computation of basic wave parameters very rapid.

In the exemplar case, the threshold of the model was set to one ($q_{ij} = 1$). This is because measles waves in Iceland are discrete in the sense that, at the end of the wave, case notifications drops to zero and only a handful of imported cases are reported in inter-epidemic periods. But for many diseases the pattern may be less clear cut. For example, an epidemic wave may begin with zero and end in a low (but non-zero) endemic state: newly-introduced diseases (such as West Nile fever in the United States) provide examples where an epidemic ends in endemicity rather than extinction. Alternatively, the epidemic wave may rise from and return to an endemic state. This happens with many communicable diseases in large populations. The swash–backwash model may be readily adjusted to include such non-discrete examples by resetting the threshold so that either (a) $q_{ij} > 1$ or (b) a particular threshold rate (cases/susceptible population) is exceeded.

Extensions of the single wave model to Icelandic epidemics which are recurrent and continuous rather than recurrent but discrete is one obvious pathway. The results for measles will be compared with those for other infectious diseases which generated single-wave outbreaks in Iceland over the same historical time period. Data from Heilbrigđisskýrslur have already been assembled for four other infectious diseases (influenza, poliomyelitis, rubella and whooping cough) and work on their analysis is in hand.

By confining our immediate extension of work to Iceland we retain the basic infrastructure of data recording (the medical reporting district). This ensures that spatial parameters such as R_{0A} for one disease are directly comparable with those from another. Testing the robustness of the method by widening the results geographically, to make comparisons between different geographical areas, poses far more difficult problems and remains to be undertaken. Making comparisons across different spatial scales and between countries (each of which has its own recording infrastructure) is a classic geographical problem which has long attracted statistical research (Duncan et al. 1961). Unlike time where the unit of measurement (days, weeks, months) has a common and universal meaning, in space the units (districts, parishes, counties) are particular to the area being studied and much work remains to be done on spatial standardization problems.

So far we have confined our analysis to phase transitions within the SIR representation of the life history of infection. As such it provides a useful approach to diseases which follow this simple structure (classically, measles). We have not investigated diseases which involve more than three phases (e.g. HIV/AIDS with its five or more stages). In principle, the

swash–backwash model could be extended to diseases with more complex biological life-histories, and cross-phase statistics might allow important by-products to be generated (e.g. the residence time in different intermediate disease states). Regretfully, the paucity of space–time data which record such intermediate stages is likely to hinder progress in this area.

6 Conclusion

We have argued in this paper that, while there is a large literature on the form of epidemic waves in the time domain, models of their geographical structure and form in the spatial domain are poorly developed. Here, we have concentrated on the changing spatial distribution of the wave, and we have presented a simple method for identifying the leading and trailing edges of the spatial advance and retreat of such waves. We have used analysis of these edges both to disaggregate waves into swash and backwash stages, and to measure as dimensionless integrals the phase transitions of areas from susceptible through infective to recovered status. A spatial version of the basic reproduction ratio is generated as a useful by-product of the analysis. The methods used have been illustrated by application to a hypothetical example and then to measles waves in Iceland over a 60-year period from 1915 to 1974. The methods proposed would seem to have special relevance to areas where epidemiological data is relatively sparse but where dates for the start and finish of reported infections are available on a spatial basis. The model can be readily extended to more complex cases by modifying the threshold values used to define the start and end points of an event.

References

Abramson G, Kenkre VM, Yates TL, Parmenter RR. Travelling waves of infections in the Hantavirus epidemics. Bull Math Biol. 2003; 65:519–534. [PubMed: 12749537]

Bailey NTJ (1975) The mathematical theory of infectious diseases. Griffin, London

Chorley RJ, Schumm SA, Sugden DE (1984) Geomorphology. Methuen, London

Cliff AD, Haggett P. Methods for the measurement of epidemic velocity from time-series data. Int J Epidemiol. 1981; 11:82–89. [PubMed: 7085184]

Cliff AD, Haggett P, Ord JK, Versey GR (1981) Spatial diffusion: an historical geography of epidemics in an island community. Cambridge University Press, Cambridge

Cliff AD, Haggett P, Ord JK (1986) Spatial aspects of influenza epidemics. Pion, London

Cliff AD, Haggett P, Smallman-Raynor MR (1993) Measles: an historical geography of a major human viral disease, from global expansion to local retreat, 1840–1990. Blackwell, Oxford

Cliff AD, Haggett P, Smallman-Raynor MR (2000) Island epidemics. Oxford University Press, Oxford

Cliff AD, Haggett P, Smallman-Raynor MR (2004) World atlas of epidemic diseases. Arnold, London

Daley D, Gani J (1999) Epidemic modelling. Cambridge University Press, Cambridge

Dietz K. The estimation of the basic reproduction number for infectious diseases. Stat Meth Med Res. 1993; 2:23–41.

Duncan OD, Cuzzort RP, Duncan B (1961) Statistical geography: problems of analyzing areal data. Free Press, Glencoe

Gilg AW (1973) A study in agricultural disease diffusion: the case of the 1970–71 fowl-pest epidemic, Institute of British Geographers, Publications 59:77–97

Grenfell BT, Bjørnstad ON, Kappey J. Travelling waves and spatial hierarchies in measles epidemics. Nature. 2001; 414:716–723. [PubMed: 11742391]

Haggett P (2000) The geographical structure of epidemics. Clarendon Press, Oxford

Morrill R. Waves of spatial diffusion. J Reg Sci. 1968; 8:1–18.

Morrill R. The shape of diffusion in time and space. Econ Geogr. 1970; 46:259–268.

Europe PMC Funders Author Manuscripts

Strahler AN. Hypsometric (area-altitude) of erosional topography. Bull Geol Soc Am. 1952; 63:1117–1142.

Trevelyan B, Smallman-Raynor M, Cliff AD. The spatial dynamics of poliomyelitis in the United States: emergence to vaccine-induced retreat, 1910–1971. Ann Assoc Am Geogr. 2005; 95:269–293. [PubMed: 16741562]

Watts DS, Muhamad R, Medina DC, Dodds PS. Multiscale, resurgent epidemics in a hierarchical metapopulation model. Proc Natl Acad Sci USA. 2005; 102:1157–1162. [PubMed: 15650048]

Zhao X-Q, Wang W. Fisher waves in an epidemic model. Discrete Contin Dyn Syst. 2004; B4:1117–1128.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

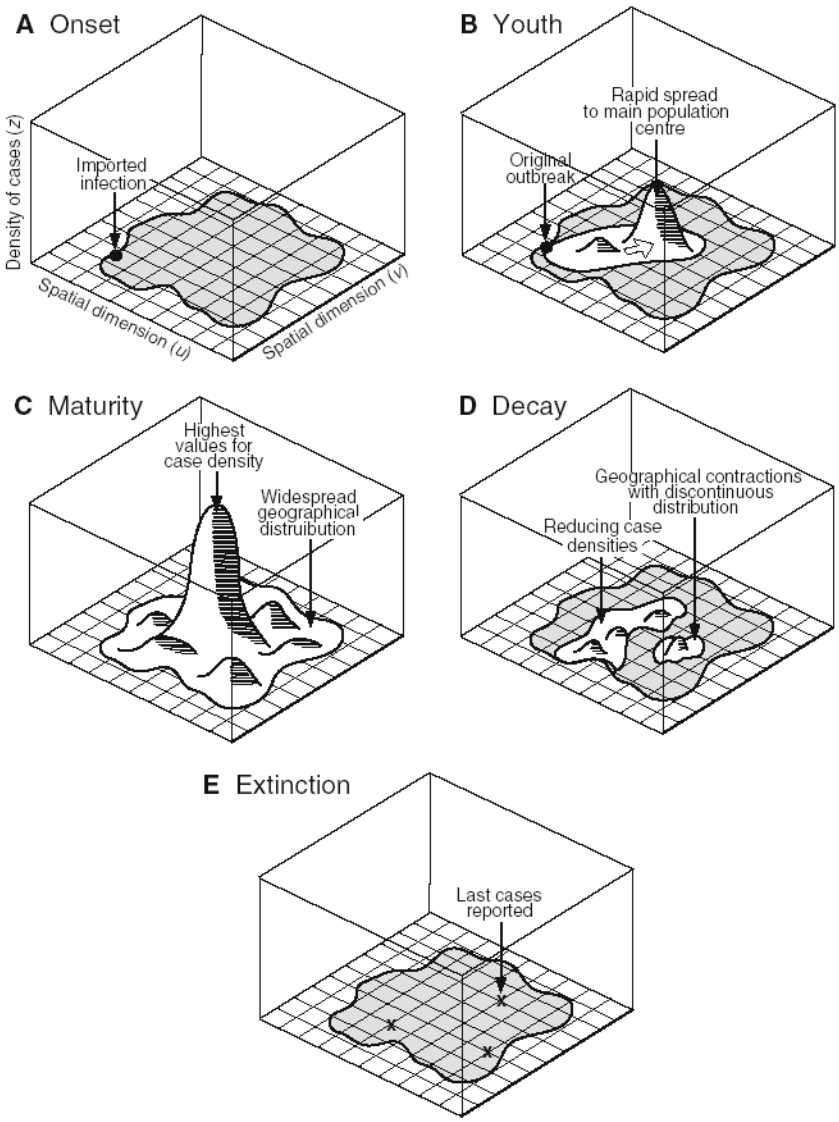

Fig. 1.

Schematic reconstruction of stages in the spread of an idealized epidemic wave. The number of cases is given in the vertical (z) direction and the spatial extent of the infection is given by the horizontal plane (u, v). *Shading* indicates the spatial extent of the island being infected

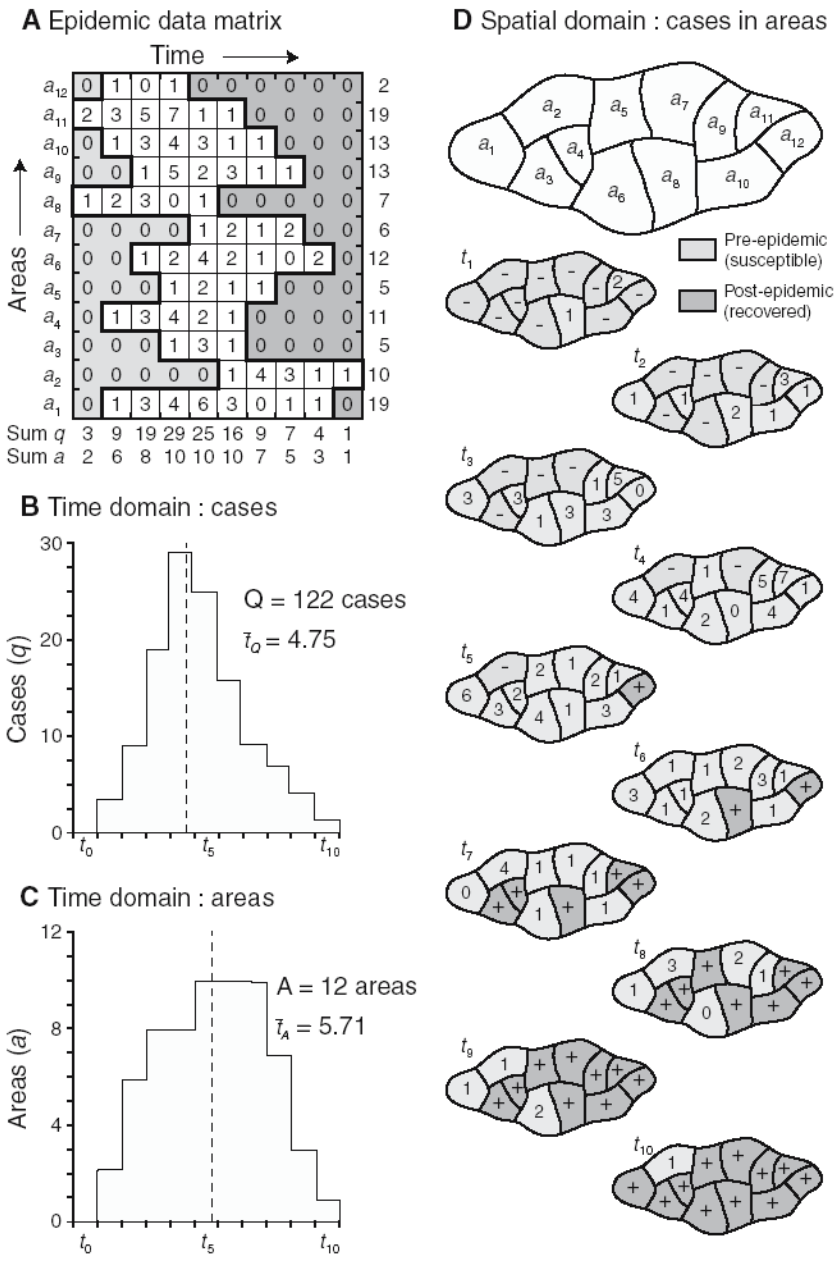

Fig. 2.

Exemplar 12×10 matrix of epidemic spread, **I. a** Original space–time data matrix with reported cases given in each cell. **b** Time distribution of cases. **c** Time distribution of infected areas. **d** Map sequence of spread over the ten time periods

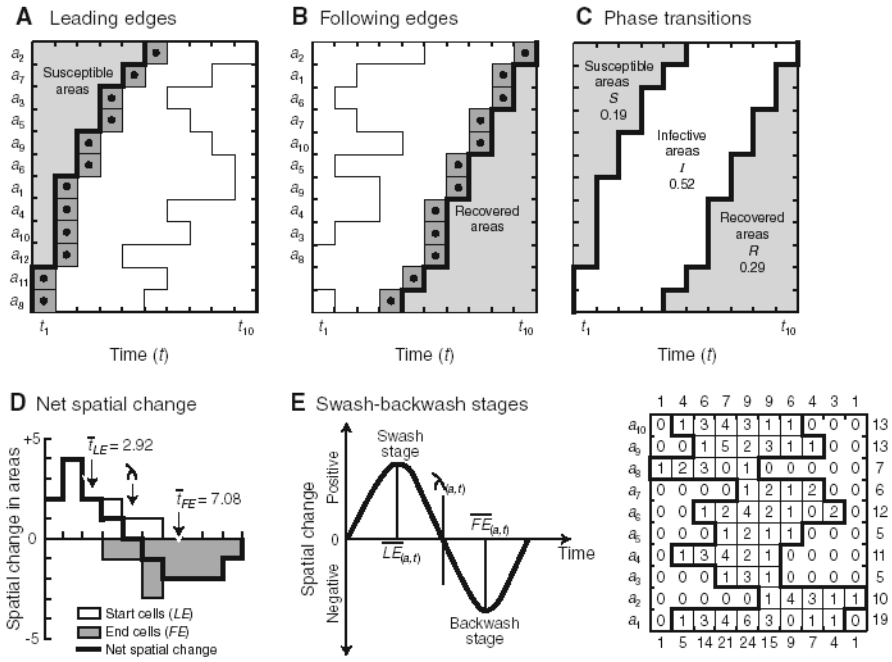

Fig. 3.

Exemplar 12×10 matrix of epidemic spread, **II. a** Space–time data matrix rearranged to order leading edges and **b** following or trailing edges. **c** Leading and following edges plotted as a phase-transition diagram with susceptible, infected and recovered integrals. **d** Distribution of leading and following edges by time. **e** Concept of a swash–backwash wave

Europe PMC Funders Author Manuscripts

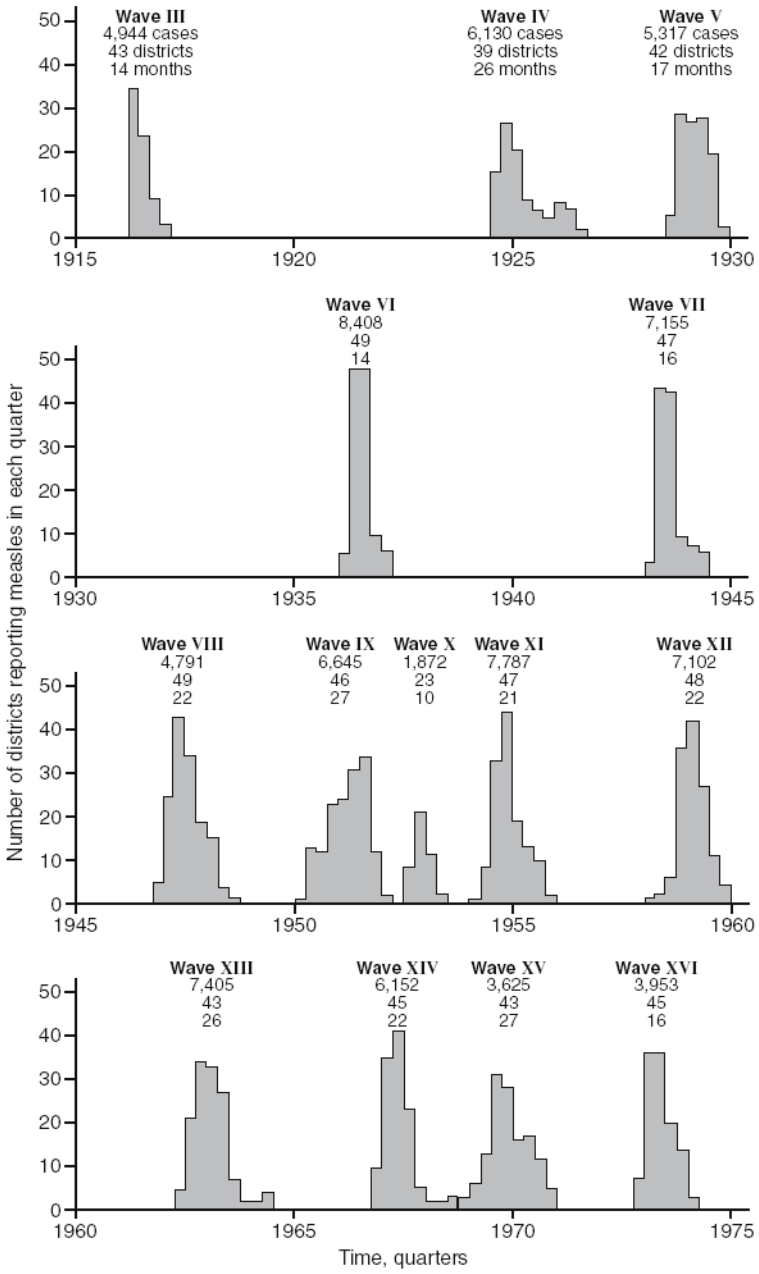

Fig. 4.

Time series of epidemic measles waves for Iceland over a 60-year period from 1915 to 1974. Monthly figures have been aggregated to give the number of districts reporting measles cases in each of the four quarters of the year. *Source*: Data from monthly series in Heilbrigđisskýslur summarized in Cliff et al. 1981 (data tables, pp. 202–229)

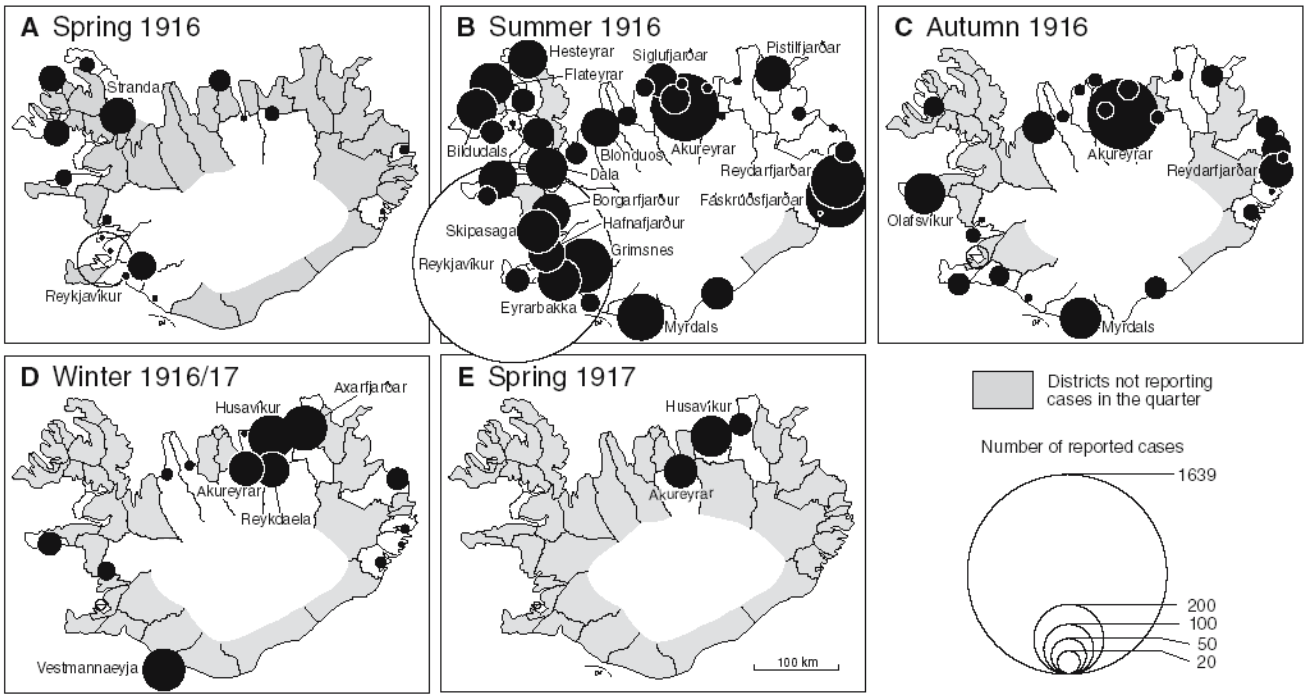

Fig. 5.

Icelandic measles Wave III. Map sequence of measles in Iceland by quarterly seasons with number of reported cases shown by *proportional circles*. The wave extended over 43 of the island's 50 medical districts over a period of 14 months from April 1916 to May 1917. *Source*: As Fig. 4

Europe PMC Funders Author Manuscripts

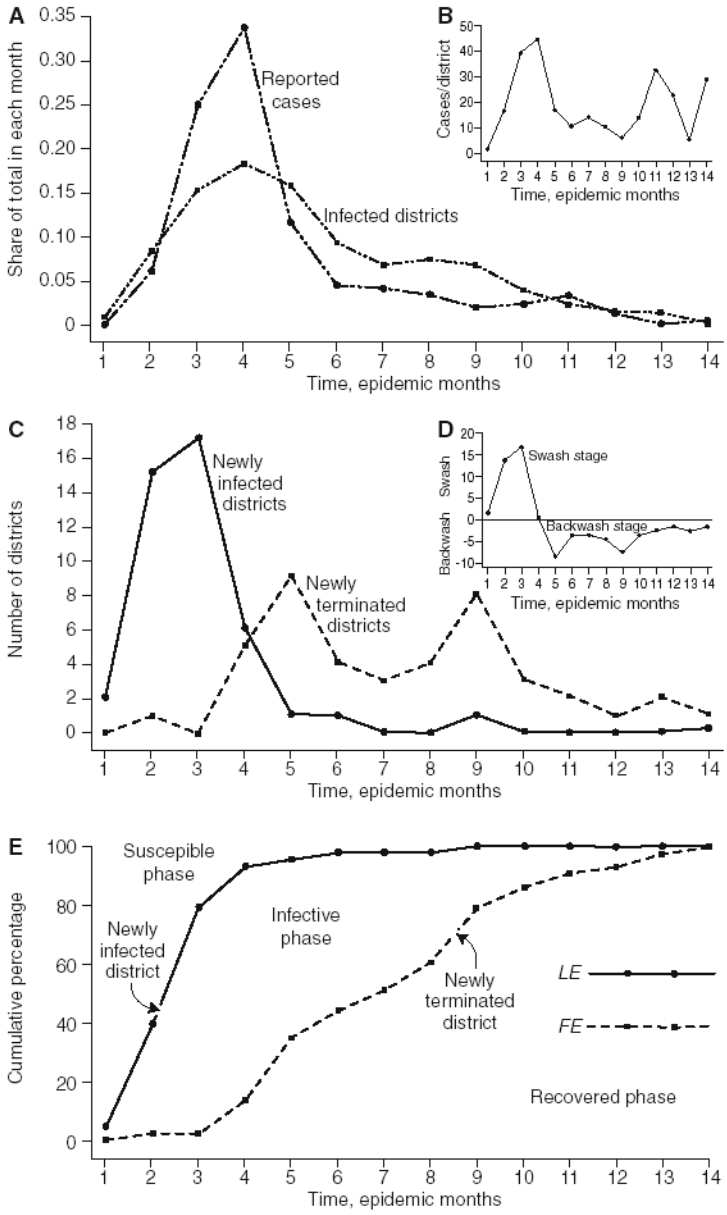

Fig. 6.

Some wave parameters for Icelandic measles Wave III. The data analyzed here refer to the outbreak mapped in Fig. 5. **a** Monthly time series for the number of cases reported and for the number of districts infected. **b** Monthly time series of density of reported cases per infected district. **c** Monthly time series of the number of districts newly infected (i.e. measles cases first recorded) and of the number of districts where infection terminated (i.e. measles cases last recorded). **d** Differences between the series plotted in **c** used to define swash and backwash stages of the epidemic. **e** Cumulative curves for the number of newly infected and newly terminated districts used to define susceptible, infective and recovered districts

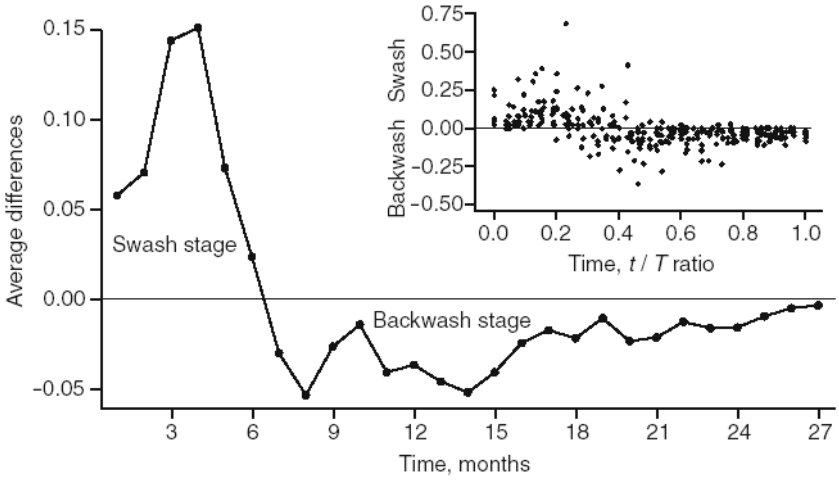

Fig. 7.

Average net values for the swash and backwash stages of the 14 Icelandic measles epidemics by month. (*Inset*) The individual values from which the average curve has been calculated are plotted on a relative time scale (*t*/*T*) to standardize for the different length of epidemic waves

Europe PMC Funders Author Manuscripts

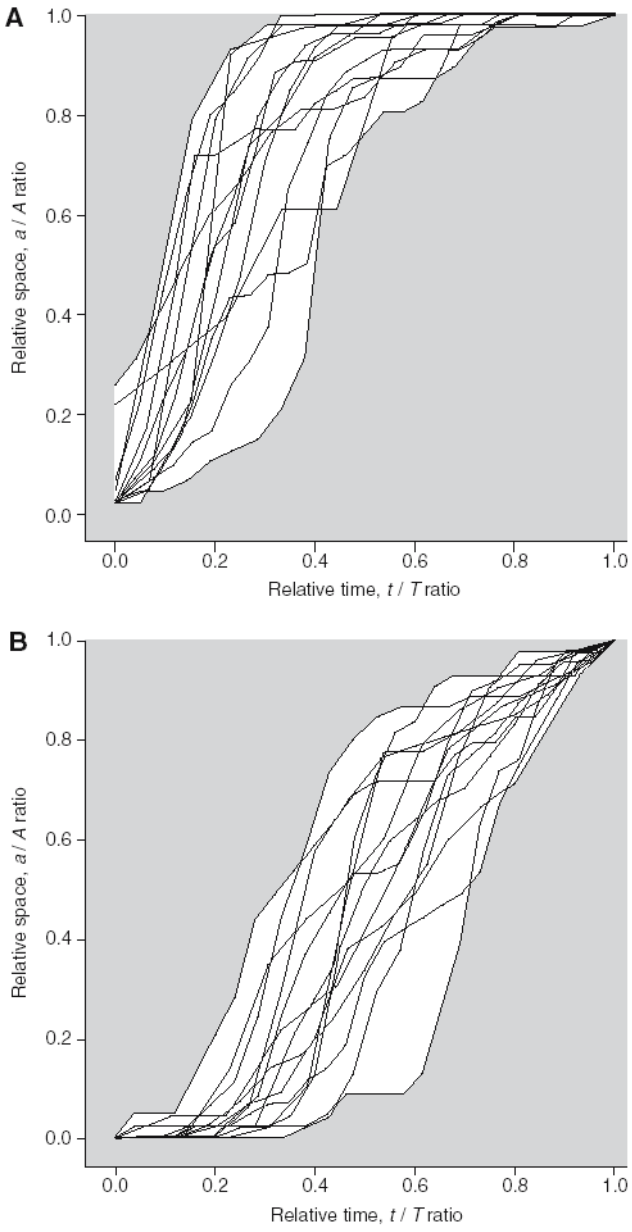

Fig. 8.

Cumulative curves for 14 separate measles epidemic waves in Iceland between 1916 and 1975. **a** Newly infected districts. **b** Newly terminated districts. In both cases the spatial and dimensional axes have been standardized to unitary length to allow waves of different time duration (t/T) and different spatial extent (a/A) to be shown comparatively

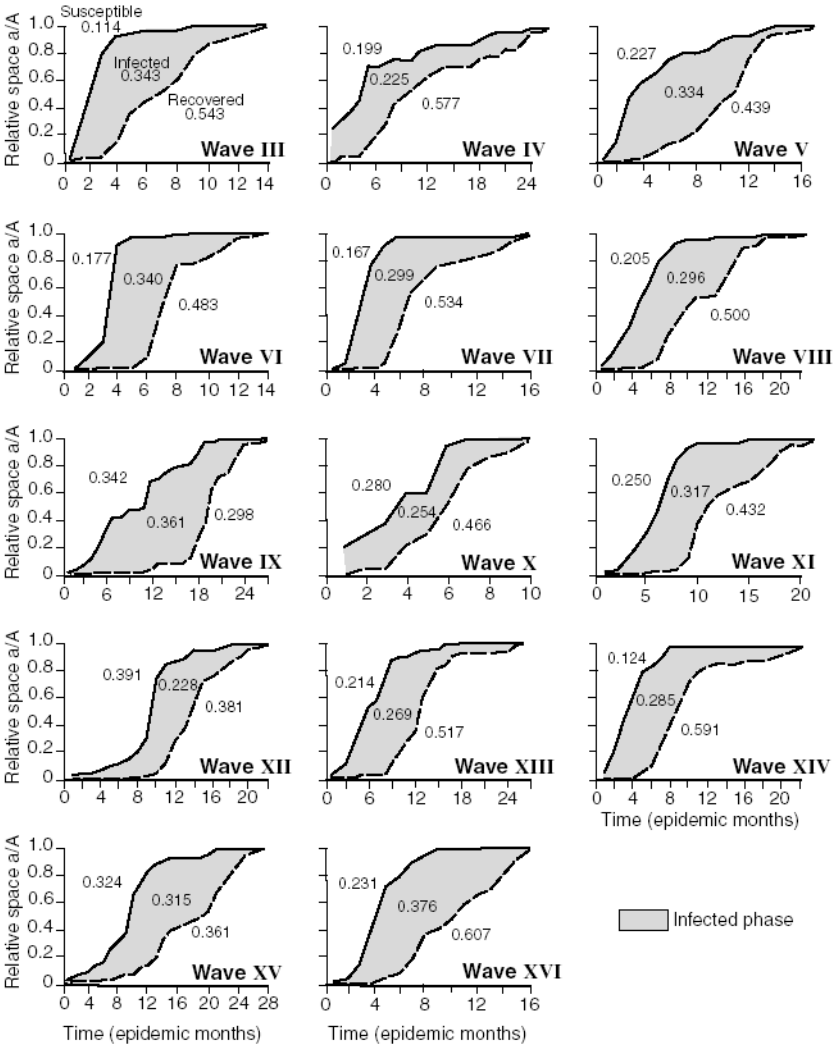

Fig. 9.

Superimposed cumulative curves from the previous figure used to define susceptible, infective and recovered phases of the individual epidemics. The numbers given are integrals which measure the area under each of the three curves as a proportion of the total. As in Fig. 8, time and space axes are measured in relative terms

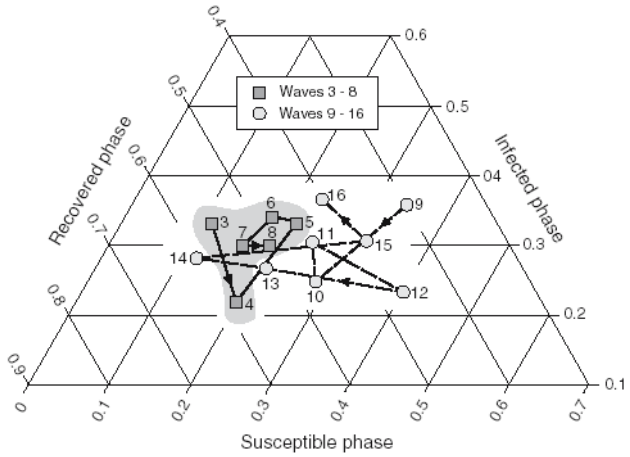

Fig. 10.

Values for the susceptible, infective and recovered phases for the 14 epidemic Icelandic measles waves plotted as a ternary diagram

Europe PMC Funders Author Manuscripts

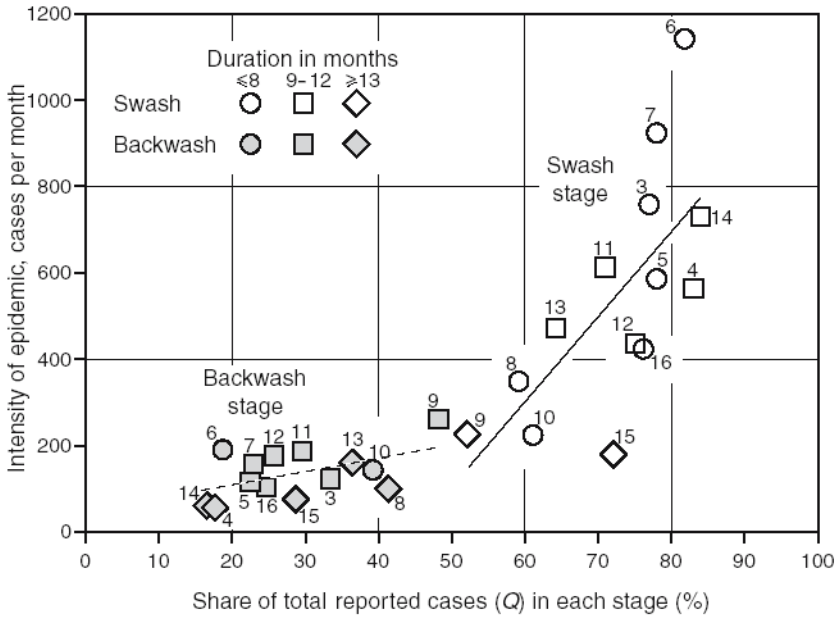

Fig. 11.

Epidemiological implications of swash and backwash stages in the 14 Icelandic measles waves. The intensity of the epidemic (*y*-axis) is plotted against the proportion of total cases recorded in each stage (*x*-axis). Linear regression lines for each of the two clusters are shown

Europe PMC Funders Author Manuscripts

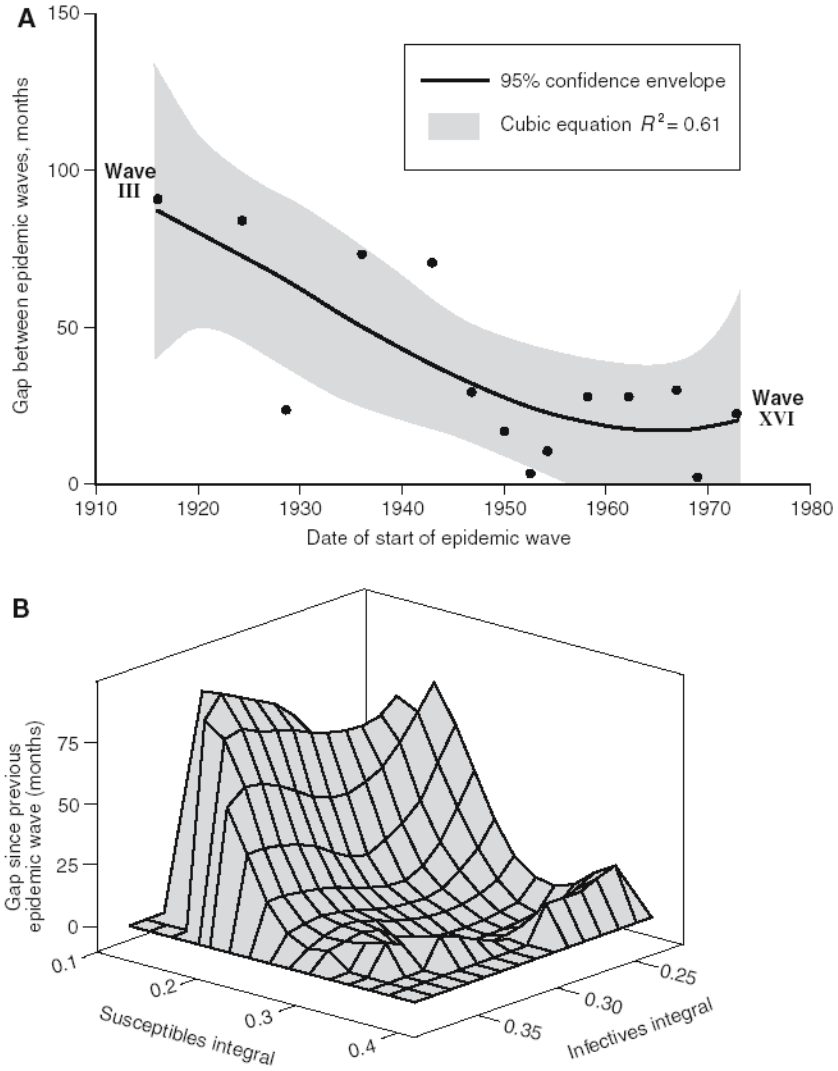

Fig. 12.

a Changes in the interval between Icelandic waves over time. The cubic regression line and 95% confidence interval are shown. **b** Surface plot of a possible control variable (time gap in months from the end of the previous epidemic wave) plotted against values for susceptible and infective phases. A long gap between waves seems to be followed by a steep swash wave with a low susceptibility integral. The infectious integral seems to be unaffected by this factor

Europe PMC Funders Author Manuscripts

Table 1

Iceland measles waves, 1915–1974: characteristics

Character	Maximum	Minimum	Mean/Median
Duration in months (*T*)	27 months (Waves IX, XV)	10 months (Wave X)	20.0/21.5 months
Spatial extent in districts (*A*)	49 districts (Waves VI, VIII)	23 districts (Wave X)	43.5/45 districts
Numbers infected, reported cases (*Q*)	8,408 cases (Wave VI)	1,872 cases (Wave X)	5,806/6,141 cases
Intensity, cases per 1,000 population	71.9 (Wave VI)	12.5 (Wave X)	42.0/43.7
Gap between epidemic waves, months	91 months (Wave III)	1 month (Wave XV)	36.0/27 months

Europe PMC Funders Author Manuscripts

Table 2

Iceland measles waves, 1915–1974: spatial characteristics in terms of the swash–backwash model

Character	Maximum	Minimum	Mean/Median
Leading edge ($\bar{t}_{LE}$)	10.4 months (Wave IX)	2.95 months (Wave III)	6.02/5.47 months
Following edge ($\bar{t}_{FE}$)	19.76 months (Wave IX)	7.44 months (Wave III)	11.83/11.31 months
Susceptibles integral (S_A)	0.395 (Wave XII)	0.139 (Wave III)	0.245/0.233
Infectives integral (I_A)	0.415 (Wave XVI)	0.260 (Wave IV)	0.347/0.347
Recovered integral (R_A)	0.541 (Wave XIV)	0.268 (Wave IX)	0.408/0.397
Swash stage	55.8% (Wave IX)	32.2% (Wave XIV)	44.6/44.4%
Backwash stage	67.8% (Wave XIV)	44.2% (Wave IX)	55.4/55.6%
Basic reproduction ratio R_0	1.88 (Wave XIV)	0.89 (Wave IX)	1.28/1.29

The values for swash/backwash stages are given in standardized form as percentages where 100 is the overall duration of the wave, *T*. Integral values are dimensionless numbers

Europe PMC Funders Author Manuscripts

Table 3

Iceland measles waves, 1915–1974: comparisons between the velocity of leading edges and following edges

Velocity parameters	Measure	Expected relationship	Leading edge (LE)	Following edge (FE)	Velocity difference confirmed?	Difference[a]
Location	Mean (months)	LE < FE	8.99	11.82	Yes	2.83***
	Mean (%)	LE < FE	44.82	59.19	Yes	14.37***
Dispersion	Standard deviation (months)	LE < FE	3.83	3.71	No	0.13
	Range (months)	LE < FE	14.57	16.64	Yes	2.07*
	IQR (months)	LE < FE	4.07	6.21	Yes	2.14**
Skewness		LE > FE	0.599	0.401	Yes	0.198
Kurtosis		LE > FE	0.460	0.456	Yes	0.004

[a]Single-tailed paired *t*-tests:

*probably significant (p=0.05),

**significant (p=0.01),

***highly significant (p=0.001)