Measles in small populations : predictability in highly stochastic systems

Q. Caudron[1,*], A. S. Mahmud[2], C. J. E. Metcalf[1], M. Gottfreðsson[3], B. T. Grenfell[1]
1 Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, 08544, USA
2 Office of Population Research, Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ, 08544, USA
3 Landspítali, Hringbraut 101, Reykjavík, Iceland
∗ E-mail: qcaudron@princeton.edu

Abstract

A standard assumption in the modelling of epidemic dynamics is that of a certain level of homogeneity. The well-known and oft-used SIR model, arguably the most important compartmental model in theoretical epidemiology, assumes that the disease being modelled is strongly immunising, directly transmitted, and has a well-defined period of infection, in addition to these homogeneity assumptions. Childhood infections, such as measles, are prime examples of diseases that fit the SIR-like mechanism, and these infections have been well studied for many systems. Populations are typically assumed to be large and well-mixed, and the data assumed to contain a high signal-to-noise ratio. Differential equations can then be used to model the disease's dynamics, generally with positive results, both in terms of explaining the underlying mechanisms for infection, and for the prediction of future time series data on disease incidence. Here, we consider a setting where populations are small and heterogeneous, and where the dynamics of infection are driven by extinction-recolonisation events. Using a TSIR model, we fit prevaccination measles incidence and demographic data in Bornholm, the Faroe Islands, and four districts of Iceland, between 1901 and 1965. The datasets for each of these countries suffers from different levels of data heterogeneity and sparsity. We explore the potential for prediction of this model : given historical incidence data and up-to-date demographic information, and knowing that a new epidemic has just begun, can we predict how large it will be ? We show that, despite a lack of significant seasonality in the incidence of measles cases, and potentially severe heterogeneity at the population level, we are able to estimate the size of upcoming epidemics, conditioned on the first time step, to within reasonable confidence. One more sentence ?

Introduction

Measles is a highly contagious and strongly immunising infection of the respiratory system [1]. Due to its extreme transmissibility, its epidemiology is conditional on the birth of susceptible individuals. As such, the temporal dynamics of measles are typically strongly oscillatory, driven seasonally by the increased contact rate amongst young children during school periods [2–4], assuming the population is large enough to sustain the disease. The critical community size, defined as the size of a population required to sustain the disease at an endemic level, is estimated to be between 250,000 and 500,000 [5–7]. In large populations, measles has been extensively studied, typically demonstrating biennial dynamics in developed countries prior to the introduction of vaccines [8,9]. These modelling efforts are typically based on a class of continuous-time systems of differential equations, such as the SIR and SEIR compartmental models. Mechanistically, these models fit well with infectious such as measles, which have a well-defined infectious period, are directly transmitted, and yield lifelong immunity to those who recover from the infection [1]. SIR-like models also assume, however, that the dynamics of the incidence of the disease are fairly smooth, and assume a certain level of homogeneous mixing between individuals in the population. In many large-populations studies, such as in [10], these assumptions hold reasonably well : the populations are large and spatially compact enough to guarantee sufficient mixing within the population and to ensure that the disease remains endemic.

In small populations, however, the dynamics of measles cases are vastly different. Susceptible individuals accumulate when measles is absent; then, driven by stochastic recolonisation, an epidemic may sweep through a large fraction of the susceptible population very quickly, only to go extinct abruptly as susceptibles become depleted. This results in very sharp, spiky epidemics, whose timing may be impossible to predict, described as Type III by Bartlett [5]. As such, methods typically used in the analysis of time-series or in dynamical systems theory are not adapted to the study of temporal changes of measles incidence in these small populations. Nonetheless, scaling analysis of the sizes and durations of measles epidemics in small populations has revealed that some level of predictability can be found within the statistics of epidemic size and duration distributions, despite the small number of epidemics observed over a given period of time [11, 12].

A discrete-time adaptation of SIR-like models was developed by Finkenstädt and Grenfell [13]. The TSIR model is a simple and computationally inexpensive system of difference equations which can be parameterised against observed incidence time-series, and is able to estimate a non-analytical time-varying contact rate. It has been successfully used in the analysis of seasonal variation of measles in several systems with large populations [10, 14–16]. Is this true ? Any other citations recommended here to show that the TSIR can reproduce dynamics that are difficult to model in continuous-time ? However, little has been done on applying the TSIR model to small populations, despite being theoretically able to tackle fast dynamics such as large epidemics in subendemic populations. Any suggested improvement to this sentence ? Also, has anything been published on small populations and TSIR ?

In this paper, we address the question of predictability of measles epidemics in small populations. First, we present data on the demographics and disease incidence in prevaccination-era Bornholm, the Faroe Islands, and four districts in Iceland. Then, we summarise the TSIR model and fit the parameters of the model to the data. After generating predictions for the evolution of each epidemic, we compare the mean predictions with the original time-series, and the predicted size of each observed epidemic. Finally, we discuss the factors which may influence the accuracy of predictions, and possible improvements to the data and methods used for improved results.

Methods

Data

Measles incidence data were obtained for Iceland, from 1901 to 1965, from [17]. This dataset consists of monthly figures for measles cases reported in 47 medical districts, originally sourced from *Heilbrigðisskýrslur* (Public Health in Iceland). Medical districts, the basic reporting unit for disease data in Iceland, are composed of *hreppar* (communes) that are roughly equivalent to English parishes or American townships. Major revisions to the boundaries of medical districts took place twice during the study period : in 1907 and 1932. Monthly incidence data for the Faroe Islands, from 1912 to 1965, were obtained from [18]; this data was originally sourced from [19]. For Bornholm, monthly measles incidence data from 1925 to 1965 were acquired from [20].

Demographic data for Iceland were obtained from [21]. Annual data on population and number of live births for the entire country were taken from [22]. Decadal population data from 1901 to 1965, for 262 municipalities, were obtained from [23]. Municipality borders changed from three to five times during the study period. In addition, many municipalities had missing data. Medical districts and municipalities were matched based on names. Several matched districts were discarded either due to missing population data, or lack of confidence in the matching of the geographical boundaries. With the data available, we were able to match four district– municipality pairs : Akureyri, Reykjavík, Hafnarfjörður, and Vestmannaeyjar. It is worth noting that matched medical district–municipality pairs may not encompass the exact same area, but one may be a (potentially partial) subset of the other.

Data on the demographics of the Faroe Islands were taken from the Statistical Yearbooks of Denmark published by Statistics Denmark [24], and from Statistics Faroe Islands [25]. Annual data on population and births from 1901 to 1965 were found in aggregated form for all of the islands in the Faroe archipelago.

Demographic data for Bornholm was collected from several publications from [24]. Annual population data for Bornholm were obtained from [26], which contains detailed statistical information collected by Statistics Denmark. Pre-1930 annual birth data were obtained from the *Ægteskaber, Fødte og Døde* (Marriages, Births and Death) available from [27]. Post-1930 annual birth data were obtained from *Befolkningsudvikling og sundhedsforhold 1901-60* (Population, Development and Health 1901–1960), from [28].

Figure 1 shows the reported incidence for Bornholm, the Faroe Islands, and four districts of Iceland.

The TSIR model

For systems with small, heterogeneous populations, epidemic dynamics are driven by stochasticity in the timing of disease recolonisations and in population mixing. As such, dynamical models such as the well established SIR model are unable to adequately represent the underlying biological processes which dictate the evolution of disease incidence over time. The time-series SIR model [13] is a discrete-time, stochastic model of disease progression written in terms of a set of difference equations. Assuming that the infection is fully immunising and that the infectious period is well-defined, then the evolution of the number of infected cases, I, can be written,

$$\mathbb{E}\left[I_{t+1}\right] = r_t\, S_t\, I_t^{\alpha}, \tag{1}$$

where S_t is the number of susceptible individuals at time t, seasonal contact rates are represented by the periodic parameter $r_t = r_{t+P}$ for P time steps per year, $0 < \alpha < 1$ is an inhomogeneity parameter, and where $\mathbb{E}\left[\,\cdot\,\right]$ denotes the expectation operator. The time step is set as the generation time of the infection. Then, the number of susceptible individuals is defined by,

$$S_{t+1} = S_t + B_{t-d} - I_t + u_t, \quad \mathbb{E}\left[u_t\right] = 0. \tag{2}$$

Here, u_t is zero-mean additive noise. B_{t-d} is the number of births born d time step prior to t, the delay d assumed due to maternal immunity, and set at four months [1].

The observed number of cases, C_t, is assumed to be underreported by a reciprocal reporting rate $\rho_t \geq 1$, such that the true number of infected cases at time t is given by $I_t = \rho_t\, C_t$. If the number of susceptible individuals S_t fluctuates around a mean $\bar{S}$ such that $S_t = \bar{S} + Z_t$, then, from equation (2), the dynamics of the susceptible individuals are given by

$$Z_{t+1} = B_{t-d} + Z_t - \rho_t\, C_t + u_t. \tag{3}$$

A major assumption made by the TSIR model is that all individuals will eventually become infected. As such, the incidence of infected cases should track births. Successive iteration of equation (3) yields,

$$Z_{t+1} = \sum_{i=1}^{t} B_{i-d} - \sum_{i=1}^{t} \rho_i\, C_i + \sum_{i=1}^{t} u_i + Z_0. \tag{4}$$

If $\rho_t = \rho$ is a constant, and u_t is small, then equation (4) reflects a linear relationship between the cumulative births and the cumulative incidence. However, as Z_{t+1} depends on Z_t, it can be shown that the reporting rate need not be a constant, and that it could be estimated using locally linear regression methods. Then, Z_t can be found as the residuals of this regression.

Fitting

The time step in the difference equations (1) and (2) is fixed at the generation time of the infection. For measles, the period of time from infection to recovery is approximately two weeks [1]. Due to the very spiky nature of the reported incidence data (whose derivatives are non-smooth due to low sampling rates), interpolation must be done such that peaks in the data are not missed or reduced. As such, a linear interpolant with an integer multiple of the number of points per year was used. This yielded 24 time points per year, thus maintaining the maximum values of the peaks in the data, and fixing the generation time at just over fifteen days.

Population and live births, assumed to be smooth, were interpolated cubically to 24 time points per year. Despite large intervals between some of the reported demographics data, Finkenstädt and Grenfell [13] report that the regression for reconstructing susceptibles is robust to pronounced changes in birth rates.

The reporting rate ρ_t was estimated using Gaussian process regression, given the births and reported cases. Unlike splines or locally-weighted regression methods, Gaussian process regressions do not optimise smoothness of the fitted values, but instead yield the best unbiased estimates, assuming the error in the underlying process is normally distributed. Once found, ρ_t is the derivative of the Gaussian process prediction for the cumulative number of births, with respect to the cumulative number of cases, and Z_t are the residuals of the regression.

The mean number of susceptibles $\bar{S}$ was estimated marginally by profiling the likelihood of the logarithmic form of equation (1) :

$$\ln\left(\mathbb{E}\left[I_{t+1}\right]\right) = \ln(r_t) + \ln\left(\bar{S} + Z_t\right) + \alpha \ln\left(I_t\right), \tag{5}$$

after which the seasonal contact rates r_t were estimated conditionally on $\bar{S}$. The inhomogeneity parameter was fixed at $\alpha = 0.97$, as in [16], implying a small, nonlinear inhomogeneity, yet not significantly impacting transmission dynamics between large and small epidemics.

Predictions

Using the TSIR model as defined by the system of equations (1, 2), predictions for epidemic dynamics were made by sampling the incidence I_{t+1} from a binomial distribution :

$$I_{t+1} \sim \mathrm{Bin}\left(S_t,\, 1 - \mathrm{e}^{-\lambda}\right), \tag{6}$$

where the number of Bernoulli trials is given by the number of susceptible individuals, S_t. The probability of a successful infection is defined by the force of infection, $\lambda = r_t\, I_t^{\alpha}$, cumulated over one biweek period. This assumes that each individual spends an exponentially-distributed period of time in the infected class.

Due to the abundance of zeros in the incidence time-series, initial conditions cannot simply be taken as the point (I_0, S_0). Instead, each epidemic must be simulated independently, with initial conditions given by the data at the time that the epidemic begins. For each epidemic, we fix the number of infected cases and of susceptible individuals as per the data, and allow the simulation to continue until the next epidemic begins. Thus, we always simulate the same number of epidemics as given by the incidence data, where each epidemic is simulated conditioned on the data available at the beginning of that epidemic. Is this clear, or does it need rewording ?

In order to clearly establish when this time is, a sensitivity threshold must be set. Let $\tau \in \mathbb{Z}^+$ define the number of reported infected cases necessary for any particular biweek period to be considered part of an epidemic. A choice of $\tau = 1$ ensures that all available non-zero data is used. However, many potential epidemics go extinct before propagating through the population, especially in highly heterogeneous populations. As such, using $\tau = 1$ would cause a large number of strongly overestimated epidemics, which in turn would deplete the susceptible pool, and underestimate future epidemics. As such, we treat τ as a sensitivity parameter, and fit it by selecting the sensitivity threshold which yields the highest correlation between the mean predicted epidemic traces and the incidence data, as defined by Pearson's R^2. Then, the first point in a sequence of time steps whose incidence is greater than or equal to the threshold is considered the beginning of that epidemic. This is a simplification; we use a convolution to be more robust in detecting when epidemics start. Need to look into exactly how much impact this has on epi detection timing; if none, then this is correct. If significant, need to expand on this section.

Results

Dynamics

After fitting parameters as described above, predicted epidemic time-series were generated for each of the six localities, using the sensitivity thresholds reported in Table 1. The mean of fifty thousand simulations, and the inferred seasonalities, are shown in Figure 2, with their respective 95% confidence intervals. The reported zero-corrected correlation coefficient is simply Pearson's R^2 computed between the mean prediction and the observed incidences, with points where both time-series are zero left out, to reduce inflation of the correlation due to the large number of zeroes in the time-series. Overall, good agreement is generally found with the observed data, with the highest correlation being in Hafnarfjörður, a small district about ten kilometres from Reykjavík. The worst fit is found in the Faroe Islands, by a significant margin; predictions here are characterised by a number of failed extinctions, general overestimation of epidemic sizes and durations, except for the single, large observed epidemic, which is significantly underestimated.

A significant number of predicted epidemics have a right shoulder, where the model predicts that epidemics take longer to go extinct than those observed. Depending on locality, many of these shoulders are small (Akureyri, Hafnarfjörður, and Vestmannaeyjar). For other localities, predicted epidemics may fail to go extinct entirely, demonstrating cyclical behaviour until the beginning of the next epidemic (Bornholm, the Faroe Islands, and Reykjavík). This may indicate that populations are strongly heterogeneous, and that the inhomogeneity parameter, fixed at $\alpha = 0.97$ for these simulations, is an overestimate. Thoughts on this ? Inferred values are between 0.8 and 0.92, but resulting predictions are fairly poor, with epidemics not going extinct. Model behaviour with α is not obvious – a smaller value should overall decrease epidemic size, yet it also seems to reduce extinction probability. Anything worth mentioning about this here ?

The inferred seasonalities have wide distributions, demonstrated by their large confidence intervals. This is potentially due to the highly stochastic nature of measles recolonisations into their respective localities, which is the primary driver for when epidemics occur. This is in contrast to the seasonality inferred in studies of large populations, such as that of England and Wales in [13], where significant seasonal trends were found, and matched well with school-based contact times.

What else should be said here about Figure 2 ?

Predictability in Epidemic Sizes

Rather than considering a point-wise comparison between the predicted and observed epidemic time-series, a potentially more robust measure of predictability is the total number of infected cases that a particular epidemic will generate. We define the size of an epidemic as the sum of reported cases C_t for observed data, or I_t/ρ_t for predicted data, from the first time point in an epidemic to the time point before the next epidemic begins. Figure 3 shows the mean predicted epidemic size for each observed epidemic for the six localities. Several of these localities show a strong linear relationship, with near-zero intercepts and gradients around one. Again, the highest correlation between predicted and observed epidemic sizes is found in Hafnarfjörður, with a correlation coefficient of $R^2 = 0.89$. Is it worth just picking out random parts of this plot to discuss ? Not quite sure what to say here.

Any more to be said in this section, about Figure 3 ?

Discussion

These are just notes. Need to summarise what we did, and then :

Predictions on epidemic sizes can be made with a significant level of certainty, despite sparse demographic data for all localities, mismatching incidence and demography information in Iceland, and strong spatial barriers to population mixing in the Faroe Islands.

- Sensitivity parameter τ : do we miss any small epidemics ? Does having a value much greater than zero strongly affect the usefulness of this approach ? (couldn't use this live as we'd have to wait until τ cases are detected before predicting). Maybe elaborate in Methods on convolution peak detection used for epi detection
- Improving the data : get spatial borders for Iceland, get disaggregated incidence data for Faroe
- Improving the method : trajectory matching instead of linearly interpolating over incidence data, creating TSEIR model ?
- Drill home "model-friendly" aspects : Bornholm is a single island (no geographical limitations to mixing) and single population in terms of incidence. Things should work nicely, and they do. Faroe is five populations which we observe in aggregate form, with strong spatial boundaries hindering population mixing. It shouldn't work well, and it doesn't. Iceland is in between, because we're observing several populations somewhat independently (mixing is involved), and we have somewhat matched demographics (but borders probably don't always match), so some places are expected to be better than others. Hafnarf comes out best, potentially because of boundary matching. Vestmann should be very good too, as it's a small island off the coast of Iceland – small enough to have its own municipality and medical district, but no more, so we know they match. However, it has a very small population, so mixing may be reduced.
- Does this method work better on other systems (UK) ? Why ? Is it because of the quality of the data, or because of the assumptions made in the model ?
- Are there other methods that could work better on this type of data ? Relate back to Rhodes and Anderson

Coauthors - Other points

Haven't mentioned anything about reporting rates. Not sure they're really very important, as they're just the glue between births and cases for susceptible reconstruction. Any thoughts ?

Acknowledgements

Q. Caudron acknowledges funding from the Department of Homeland Security contract HSHQDC-12-C-00058.

References

1. Anderson RM, May RM (1991) Infectious Diseases of Humans: Dynamics and Control. Oxford University Press.

2. London WP, Yorke JA (1973) Recurrent outbreaks of measles, chicken-pox and mumps. I. Seasonal variation in contact rates. American Journal of Epidemiology 98: 453–468.

3. Fine PE, Clarkson JA (1982) Measles in England and Wales. I: An analysis of factors underlying seasonal patterns. International Journal of Epidemiology 11: 5–14.

4. Schenzle D (1984) An Age-Structured Model of Pre- and Post-Vaccination Measles Transmission. IMA Journal of Mathematics Applied in Medicine and Biology 1: 169–191.

5. Bartlett MS (1957) Measles periodicity and community size. Journal of the Royal Statistical Society: Series A (General) 120: 48–70.

6. Black FL (1966) Measles endemicity in insular populations: critical community size and its evolutionary implication. Journal of Theoretical Biology 11: 207–11.

7. Keeling MJ, Grenfell BT (1997) Disease extinction and community size: modeling the persistence of measles. Science 275: 65–67.

8. Bolker BM, Grenfell BT (1995) Space, persistence and the dynamics of measles epidemics. Philosophical Transactions of the Royal Society of London: Series B 348: 309–320.

9. Grenfell BT, Harwood J (1997) (Meta)population dynamics of infectious diseases. Trends in Ecology and Evolution 12: 395–399.

10. Bjornstad ON, Finkenstadt BF, Grenfell BT (2002) Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. Ecological Monographs 72: 169–184.

11. Rhodes CJ, Anderson RM (1996) Power law governing epidemics in isolated populations. Nature 381: 600–602.

12. Rhodes CJ, Anderson RM (1996) Disease extinction and community size: modeling the persistence of measles. Philosophical Transactions of the Royal Society of London: Series B 351: 1679–1688.

13. Finkenstädt BF, Grenfell BT (2000) Time series modelling of childhood diseases: a dynamical systems approach. Journal of the Royal Statistical Society: Series C (Applied Statistics) 49: 187–205.

14. Grenfell BT, Bjornstad ON, Finkenstädt BF (2002) Dynamics of measles epidemics: scaling noise, determinism, and predictability with the TSIR model. Ecological Monographs 72: 185–202.

15. Glass K, Xia Y, Grenfell B (2003) Interpreting time-series analyses for continuous-time biological models – measles as a case study. Journal of Theoretical Biology 223: 19–25.

16. Metcalf CJE, Munayco CV, Chowwell G, Grenfell BT, Bjørnstad ON (2010) Rubella metapopulation dynamics and importance of spatial coupling to the risk of congenital rubella syndrome in Peru. Journal of the Royal Society Interface 8: 369–376.

17. Cliff AD, Haggett P, Ord JK, Versey GR (1981) Spatial Diffusion: An Historical Geography of Epidemics in an Island Community. Cambridge University Press.

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

19. Lancaster HO (1991) Expectations of Life: A Study in the Demography, Statistics, and History of World Mortality. Springer.

20. Denmark. Medical Report for the Kingdom of Denmark, 1927–1968, National Health Service of Denmark, Copenhagen.

21. Iceland. Statistics Iceland
`www.statice.is`.

22. Iceland. Statistics Iceland : Births by months 1853–2012
`www.statice.is/Statistics/Population/Births-and-deaths`.

23. Iceland. Statistics Iceland : Population by municipalities 1901–1990
`www.statice.is/Statistics/Population/Municipalities`.

24. Denmark. Danmarks Statistik
`www.statistikbanken.dk`.

25. Faroe Islands. Hagstova Føroya
`www.hagstova.fo`.

26. Denmark. Danmarks Statistik : Population 1 January by islands
`www.statbank.dk/statbank5a/SelectVarVal/Define.asp?MainTable=BEF4&PLanguage=1`.

27. Denmark. Danmarks Statistik : Ægteskaber, Fødte og Døde
`www.dst.dk/pukora/epub/upload/20304/aefodo1921-1925.pdf`
`www.dst.dk/pukora/epub/upload/20305/aefodo1926-1930.pdf`.

28. Denmark. Danmarks Statistik : Befolkningsudvikling og sundhedsforhold
`www.dst.dk/pukora/epub/upload/19335/befsund.pdf`.

Figure Legends

Tables

Locality	τ
Akureyri	19
Bornholm	15
Faroe Islands	15
Hafnarfjörður	8
Reykjavík	18
Vestmannaeyjar	7

Table 1. Sensitivity thresholds τ for each locality, fit by maximising the correlation between the mean simulated epidemic time-series and the reported incidence data.

Figures

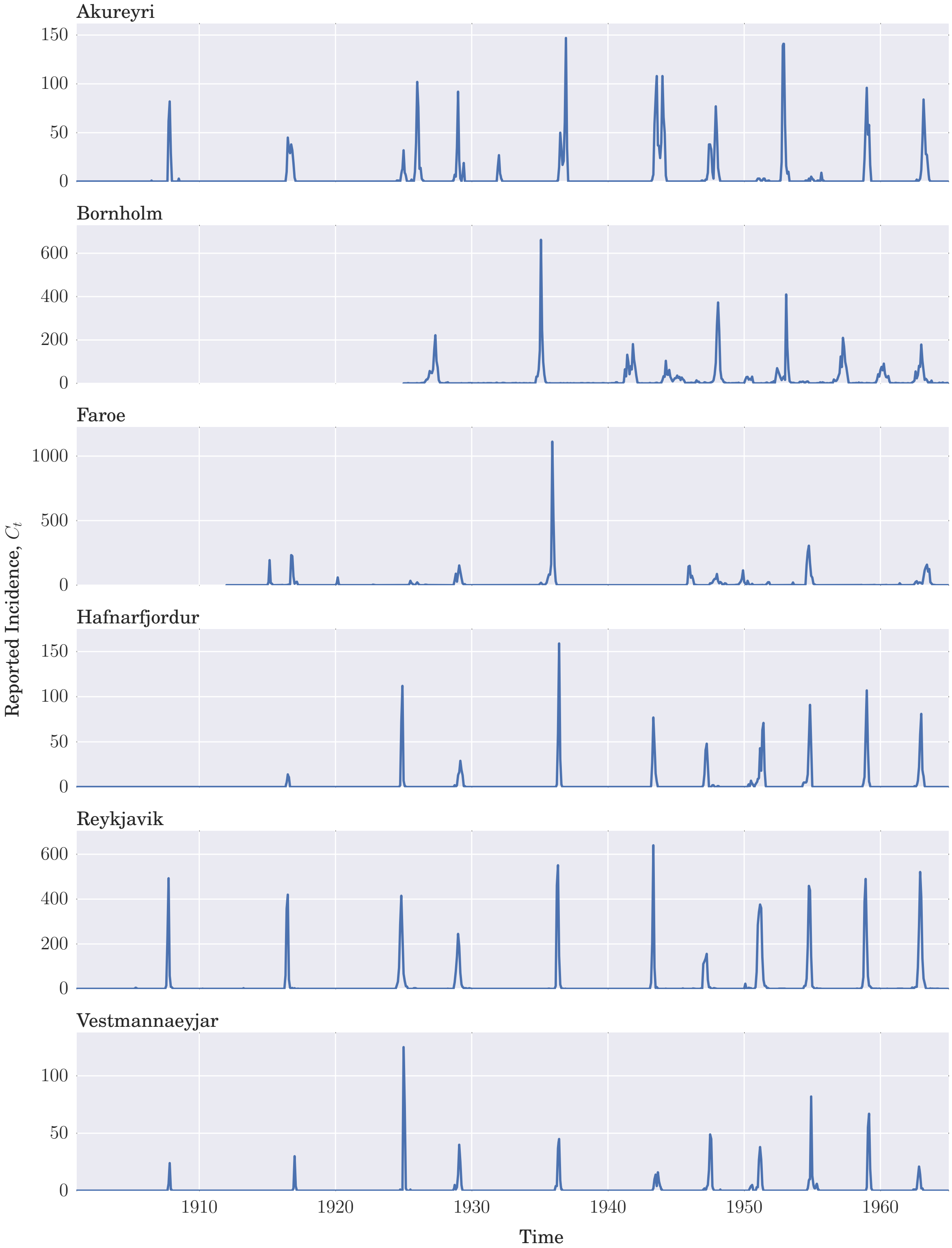

Figure 1. Reported incidence for Bornholm, the Faroe Islands, and four localities in Iceland. High temporal synchronicity can be seen in the Icelandic localities.

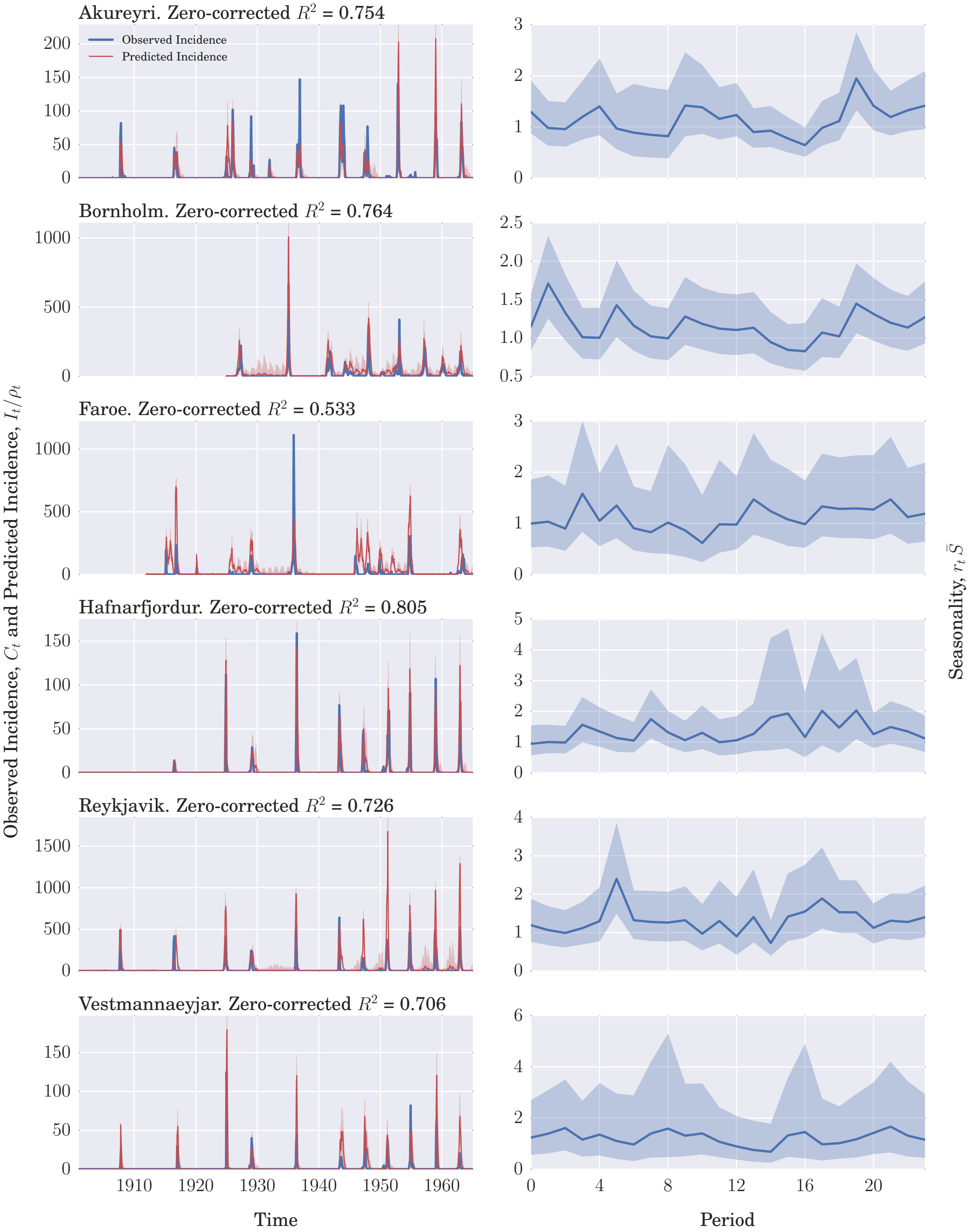

Figure 2. Predicted incidence and inferred seasonal trends. For the predicted time-series, the mean value of incidence simulations is plotted as a dark red line, with 95% confidence intervals given in light red. Seasonality is plotted as a function of the biweek, with 95% confidence intervals in light blue.

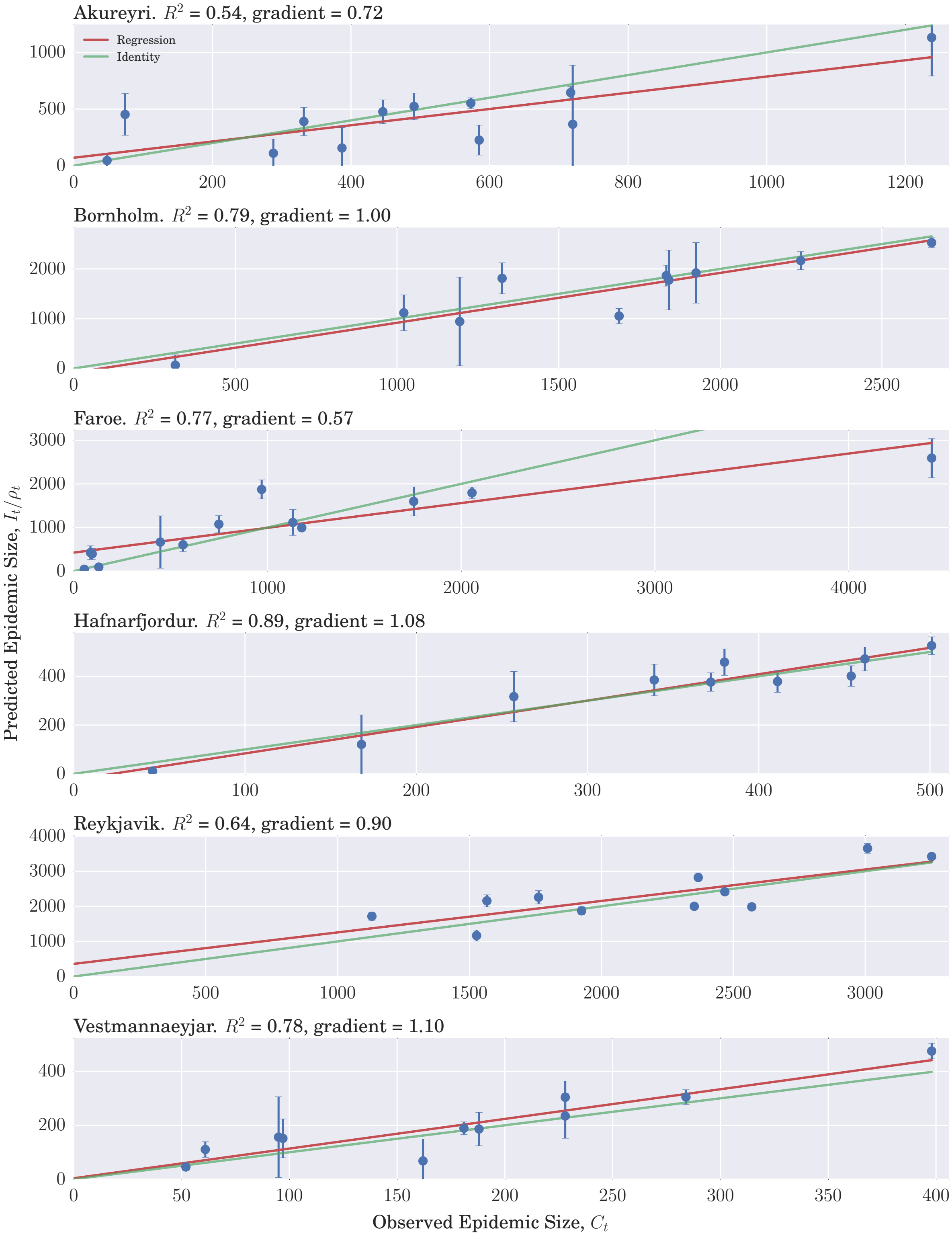

Figure 3. Predictability of epidemic sizes. The predicted size of each epidemic as a function of its observed size.