Wavelets and excess disease models for analysis of time series data

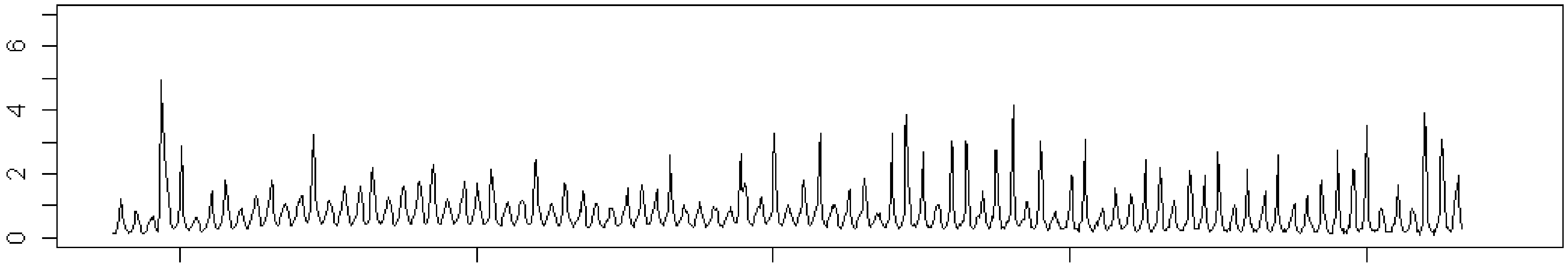
Dan Weinberger

Fogarty International Center

National Institutes of Health

DO NOT DISTRIBUTE

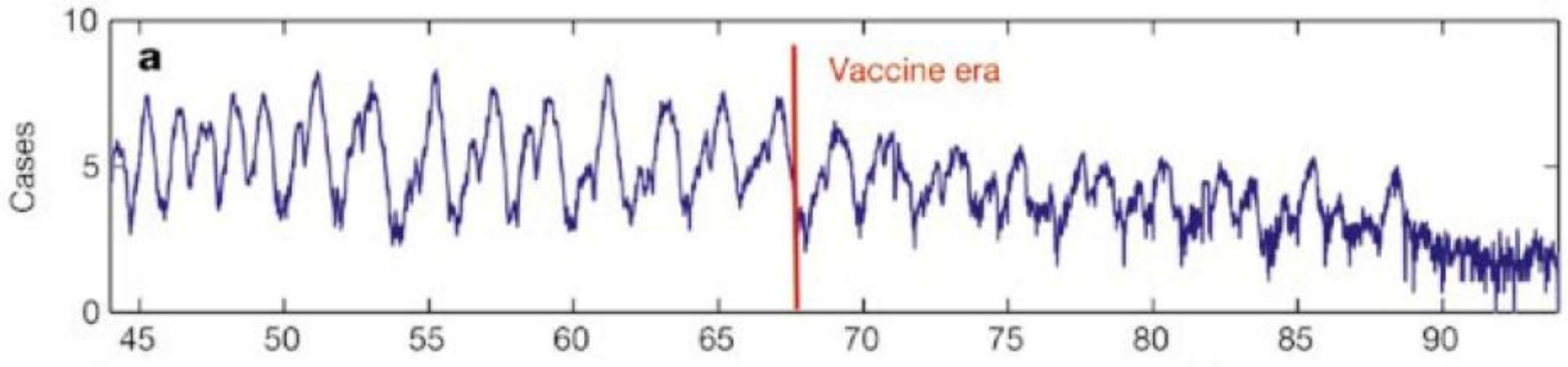
Analyzing time series data



- **Wavelets**: evaluate timing of peaks and dominant frequency
- **Regression models**: estimate seasonal baseline and calculate excess incidence

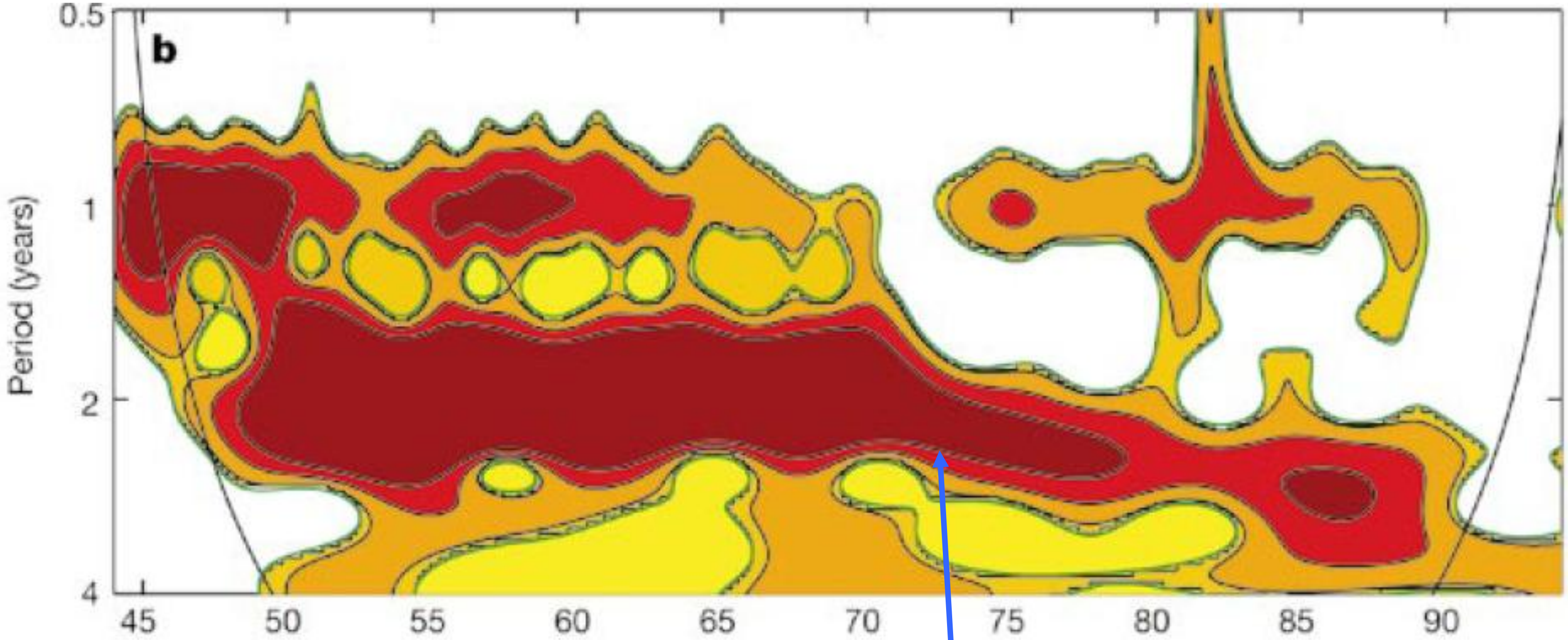
Part 1: Wavelets

Motivating example: Measles

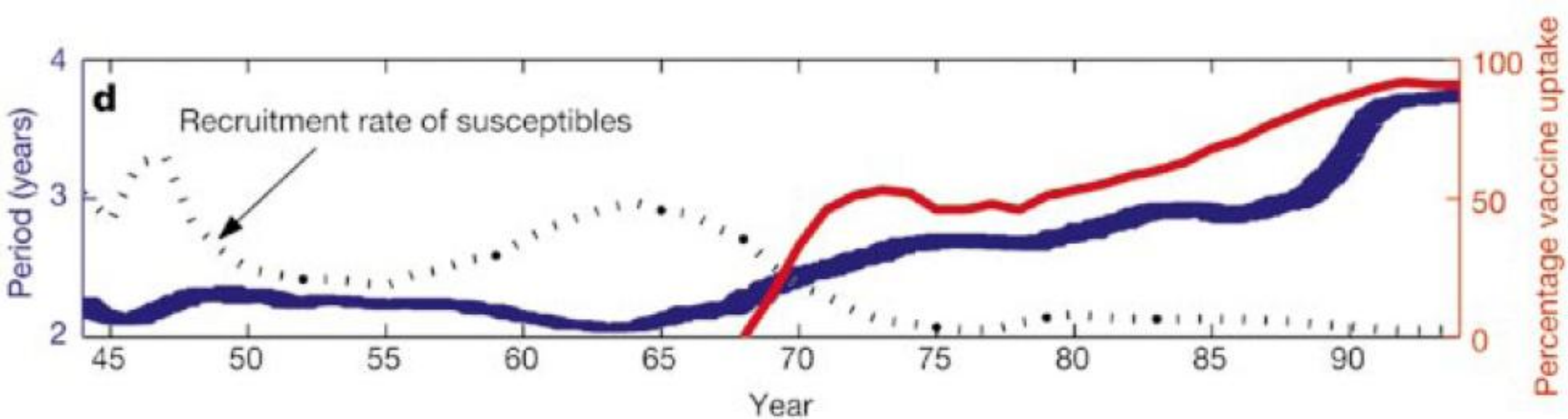


Does the frequency of the measles epidemics change after vaccination?

Figures from Grenfell et al, Nature 2001



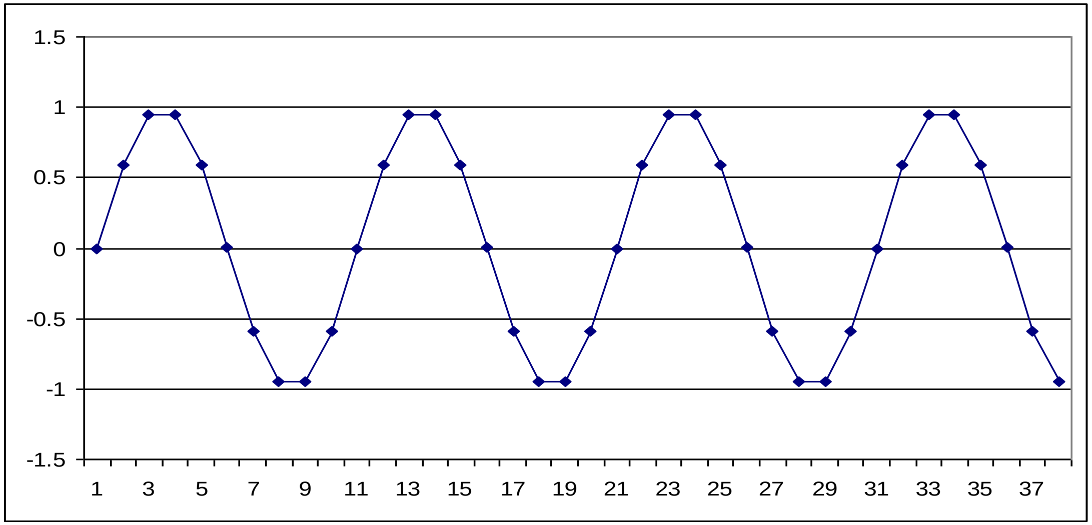
Shift in dominant frequency



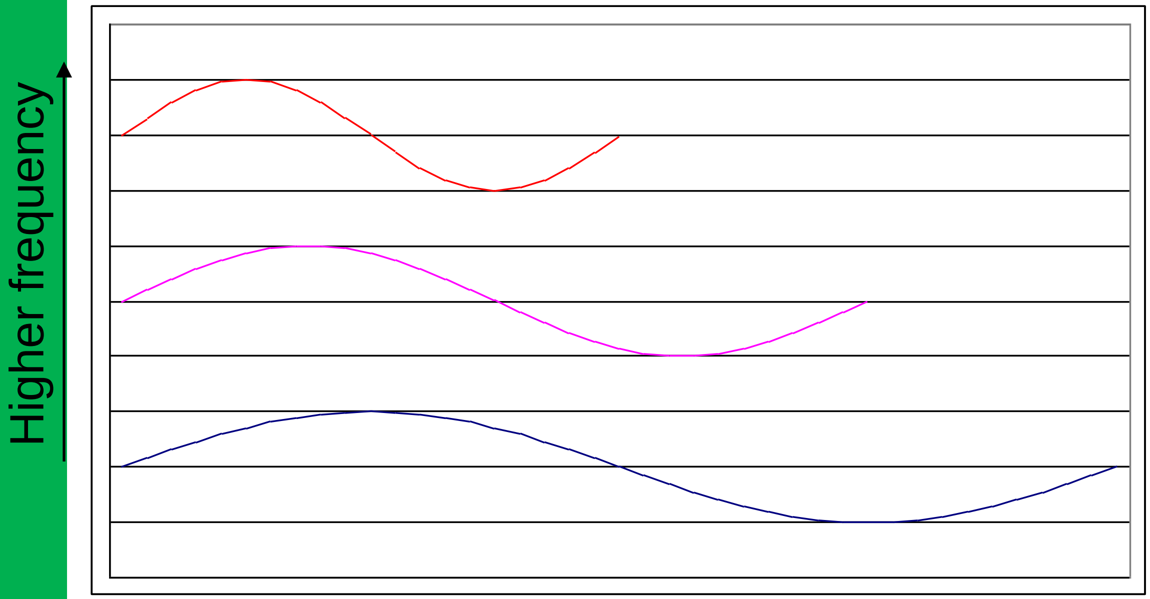
Wavelets: a powerful solution

- Identification of the dominant frequencies in a series (ie annual, monthly…) at each specific time
- Can be used to determine the “phase” of these cycles
- Can also be used as a filter to smooth a series (remove high frequency noise)
- And many other applications…

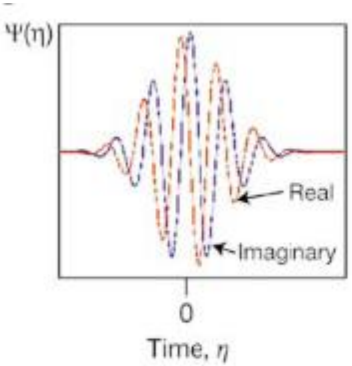
Basic concepts



Sample time series
(wave with 10 unit cycle
=sin(2*π/10*t)

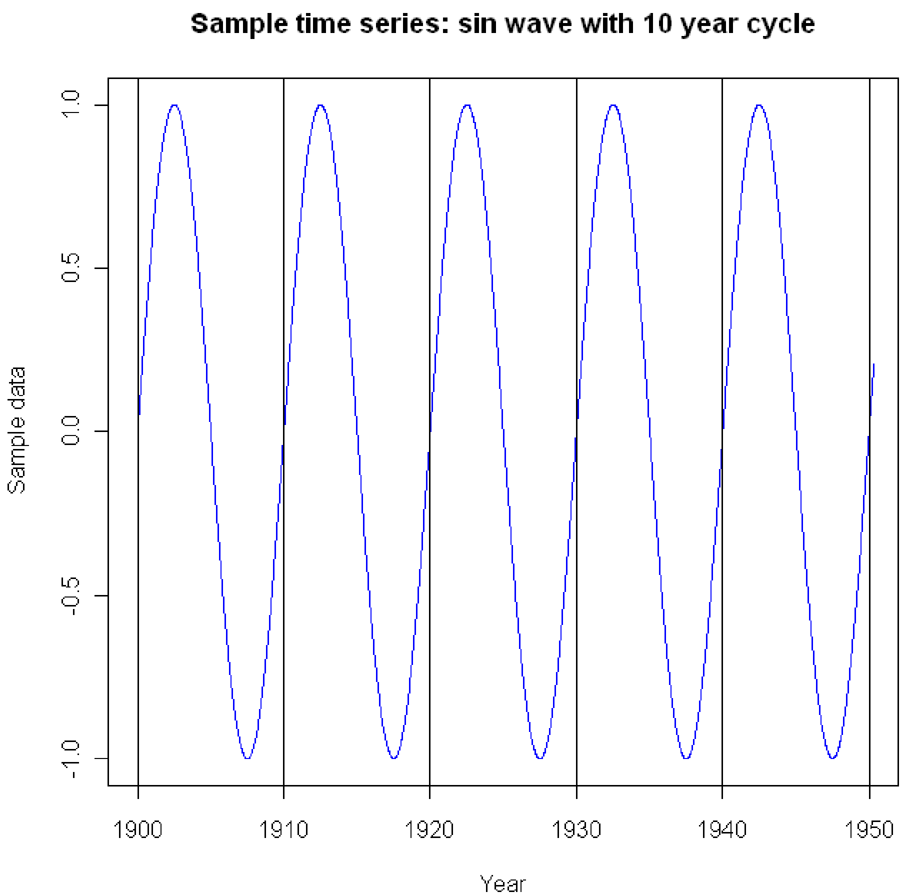


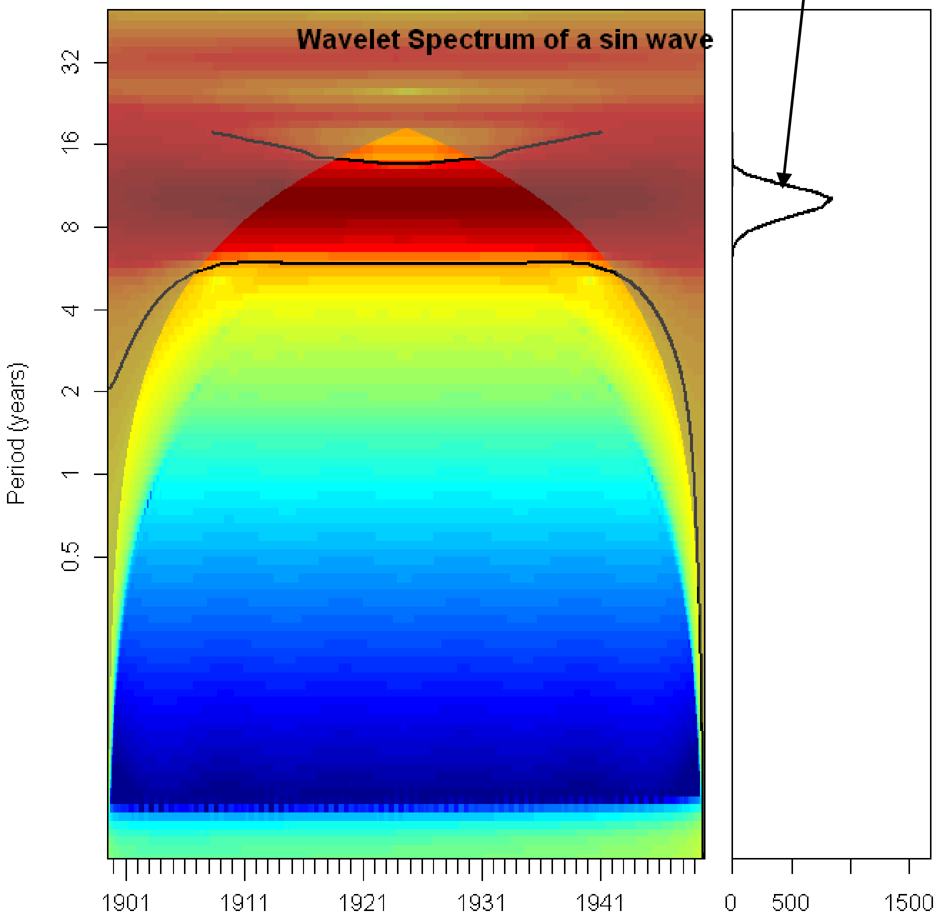
-Wavelets:little waves of a specific shape
-"slide" wavelet along time series to determine strength of correlation
-repeat, while shrinking and expanding the wavelet
-Can use different shapes of wavelets for different situations



Wavelet spectrum of a sine wave

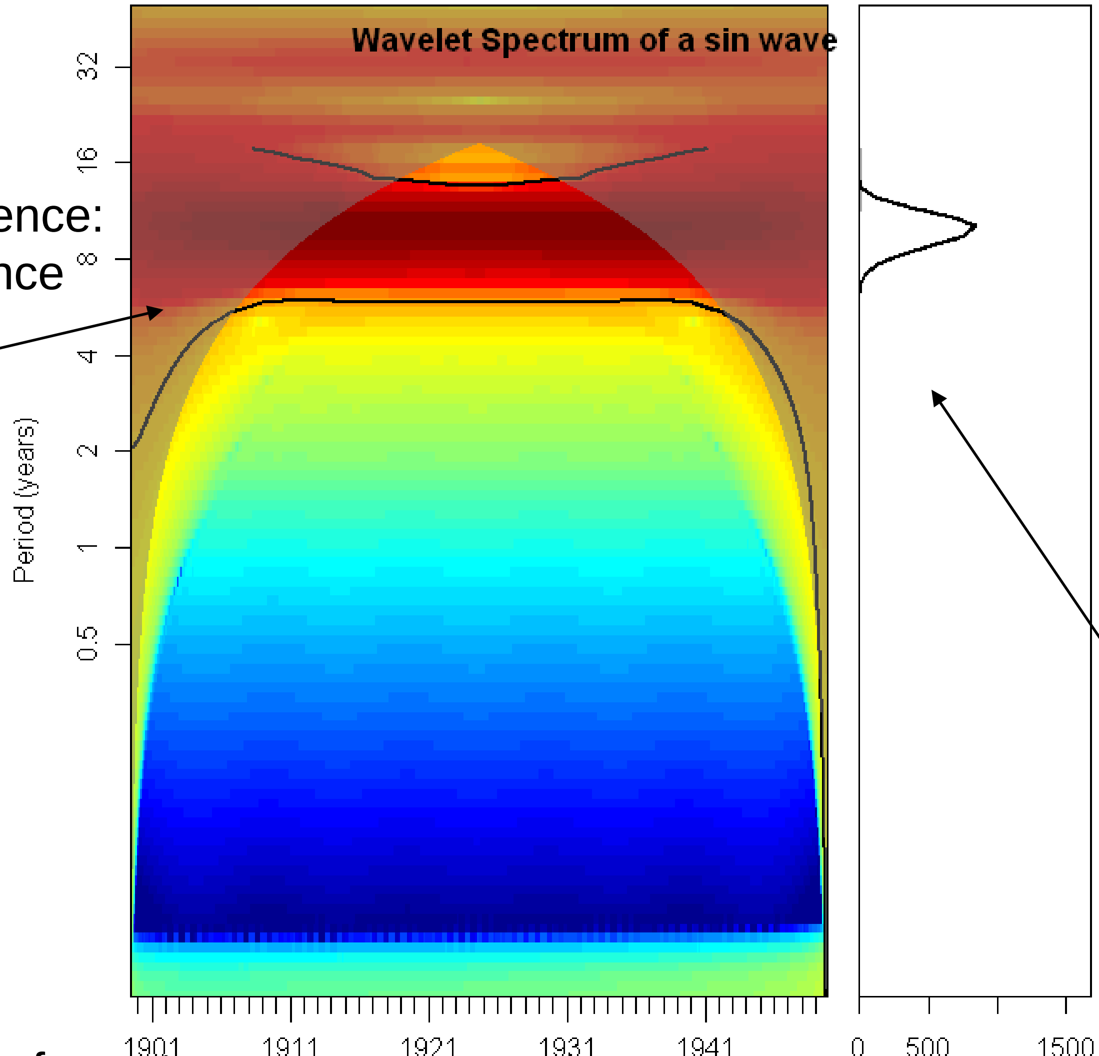
10 year cycle dominates





Wavelet spectrum of a sine wave

Grey area=
Cone of influence:
Less confidence
in this region

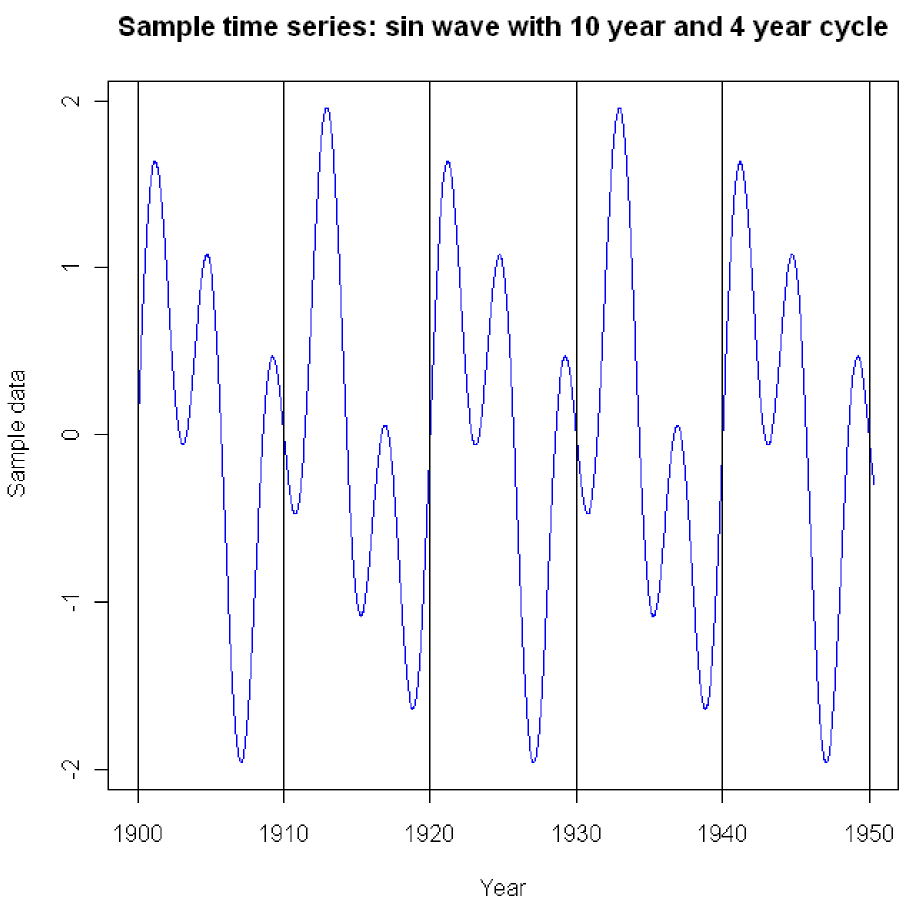


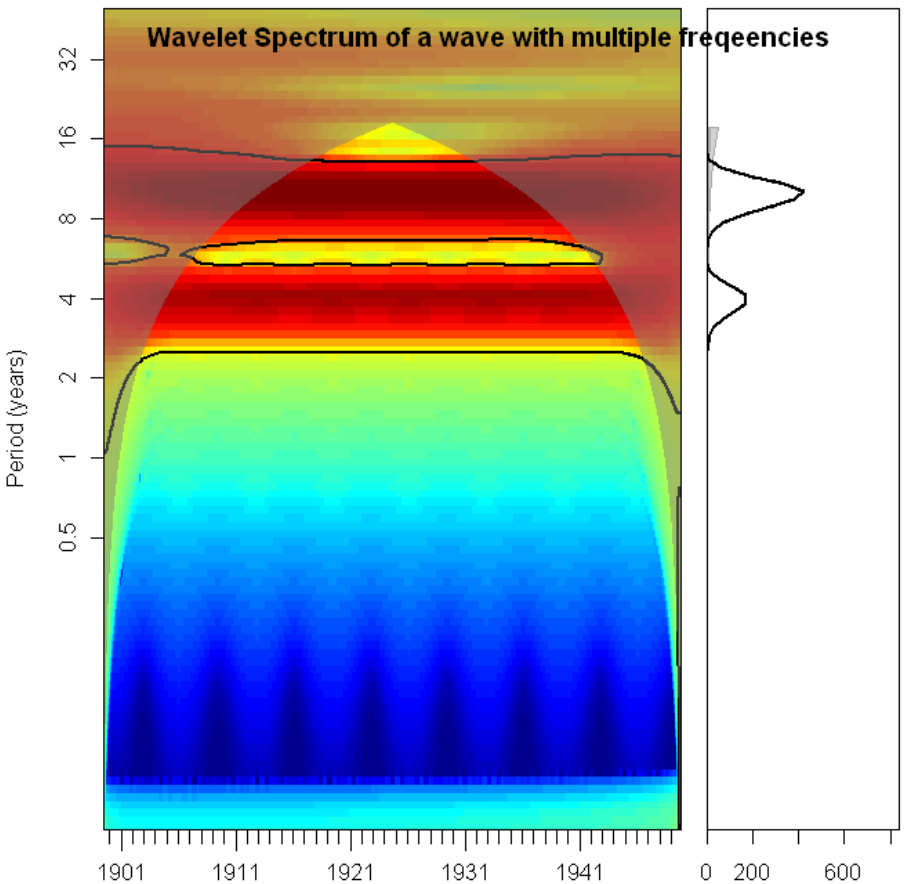
"Global wavelet"=
Average across entire
time period

Color=power of spectra:
Red=higher amplitude at that
frequency and time

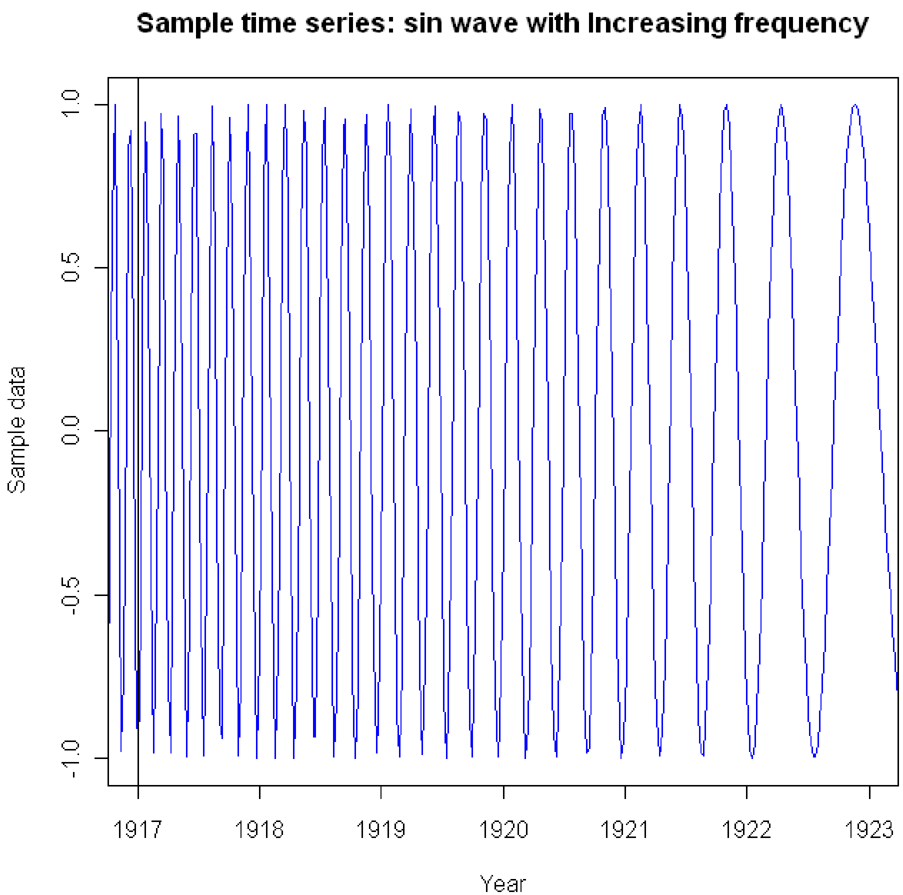
Significance tested by a
permutation test

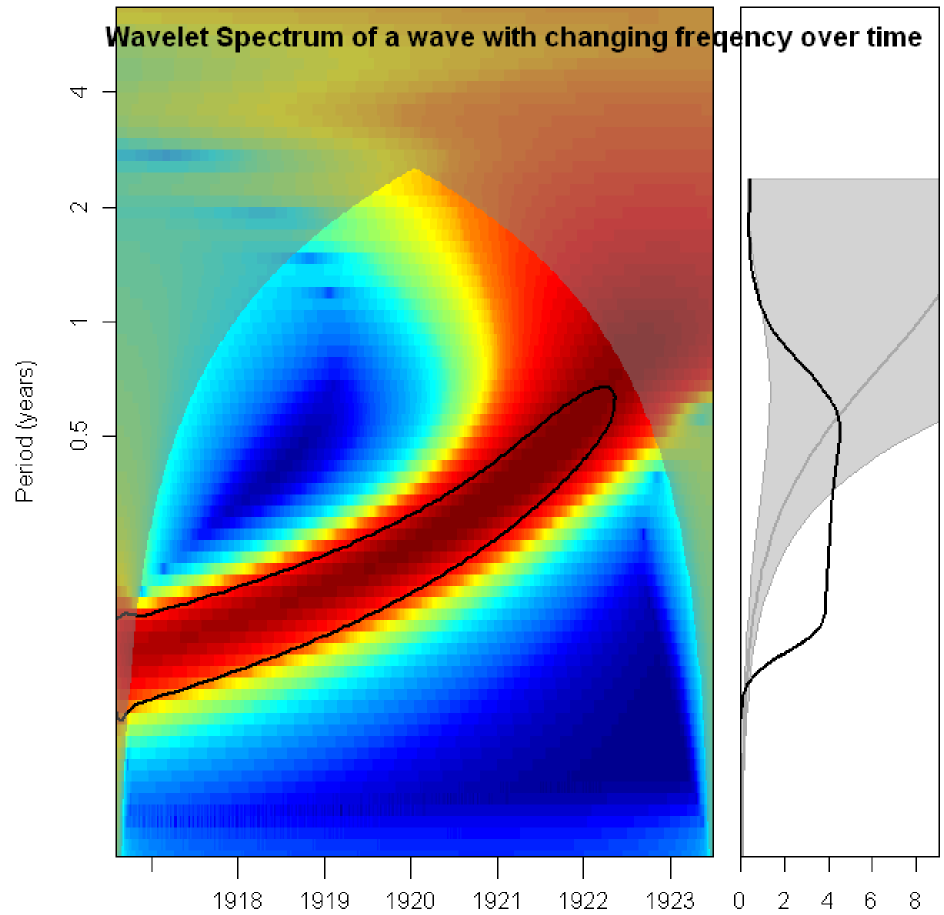
Multiple frequencies





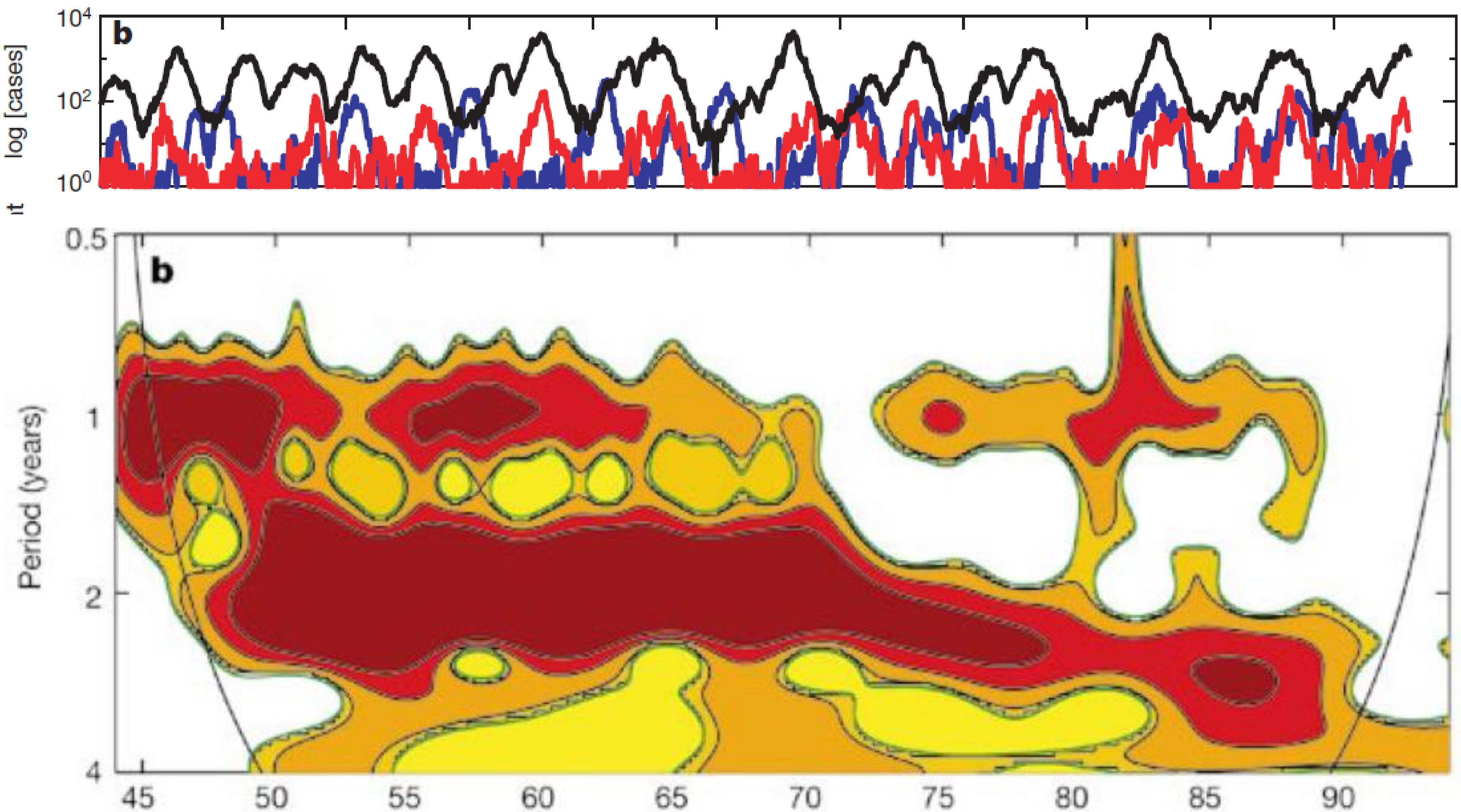
Wavelet with changing frequencies



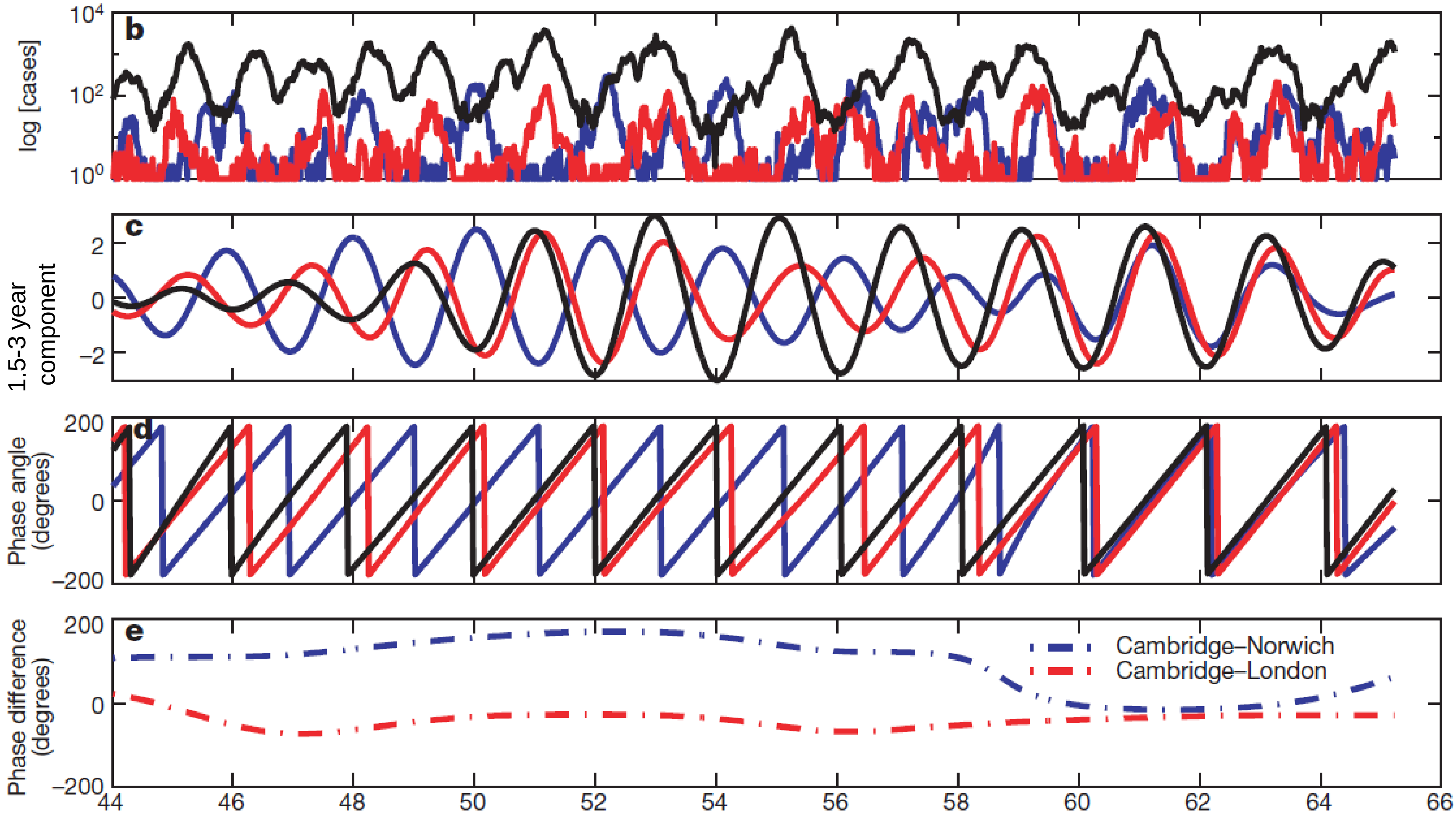


Interpretation: Wavelength increases from ~0.25 to ~0.5 (from 4 cycles/year to 2 cycles/year)

Example: epidemic timing



Example: Using wavelets to extract phase (timing) information



Applying wavelets to “real” data

- Step 1: remove any long term trends from the data (calculate baseline using spline of summer months and then divide by baseline)
- Step 2: Square root or log-transform the data
- Step 3: Use transformed data in wavelet transform, evaluate spectra, extract phase data

Note: it is important to have a complete time series without missing data for the wavelets. Need to have relatively long time series since accuracy of wavelets is poor at the beginning and end of the time series

Part 2: Excess Disease models

Part 2: Calculation of excess mortality for seasonal diseases

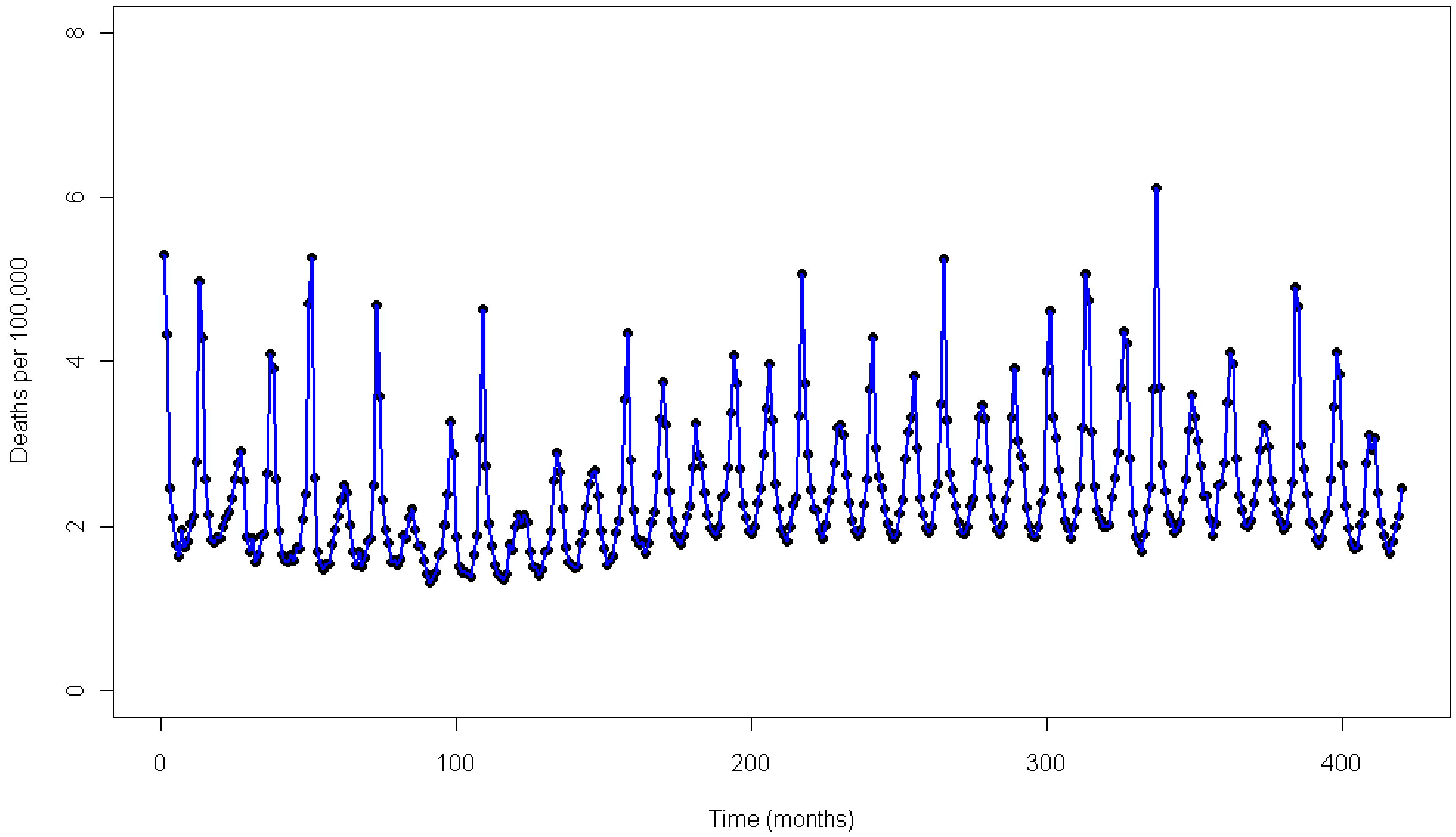
- Motivating question: How can we quantify the annual disease burden for diseases that show a strong seasonal pattern?
- Answer: Count excess above a “typical” seasonal baseline

Thanks to Cécile Viboud and Vivek Charu for contributing slides to this section

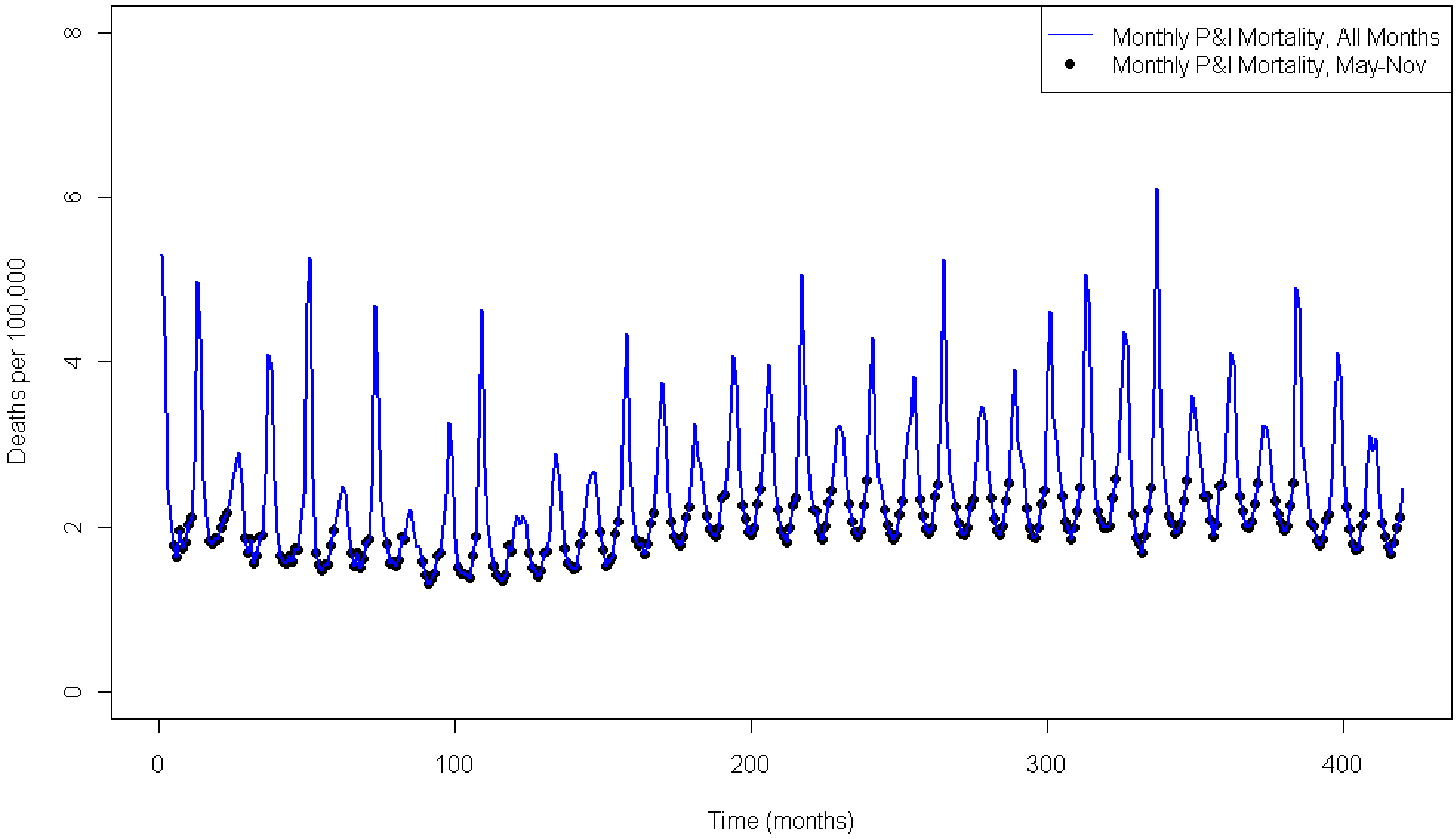
Serfling Regression

- Step 1: Define influenza, non-influenza period
- Step 2: Set a baseline and threshold (95% confidence interval) for pneumonia during non-influenza period
- Step 3: Calculate excess mortality for each year
  - Sum of observed mortality subtracted from the model baseline during “epidemic months” (when flu deaths cross threshold)

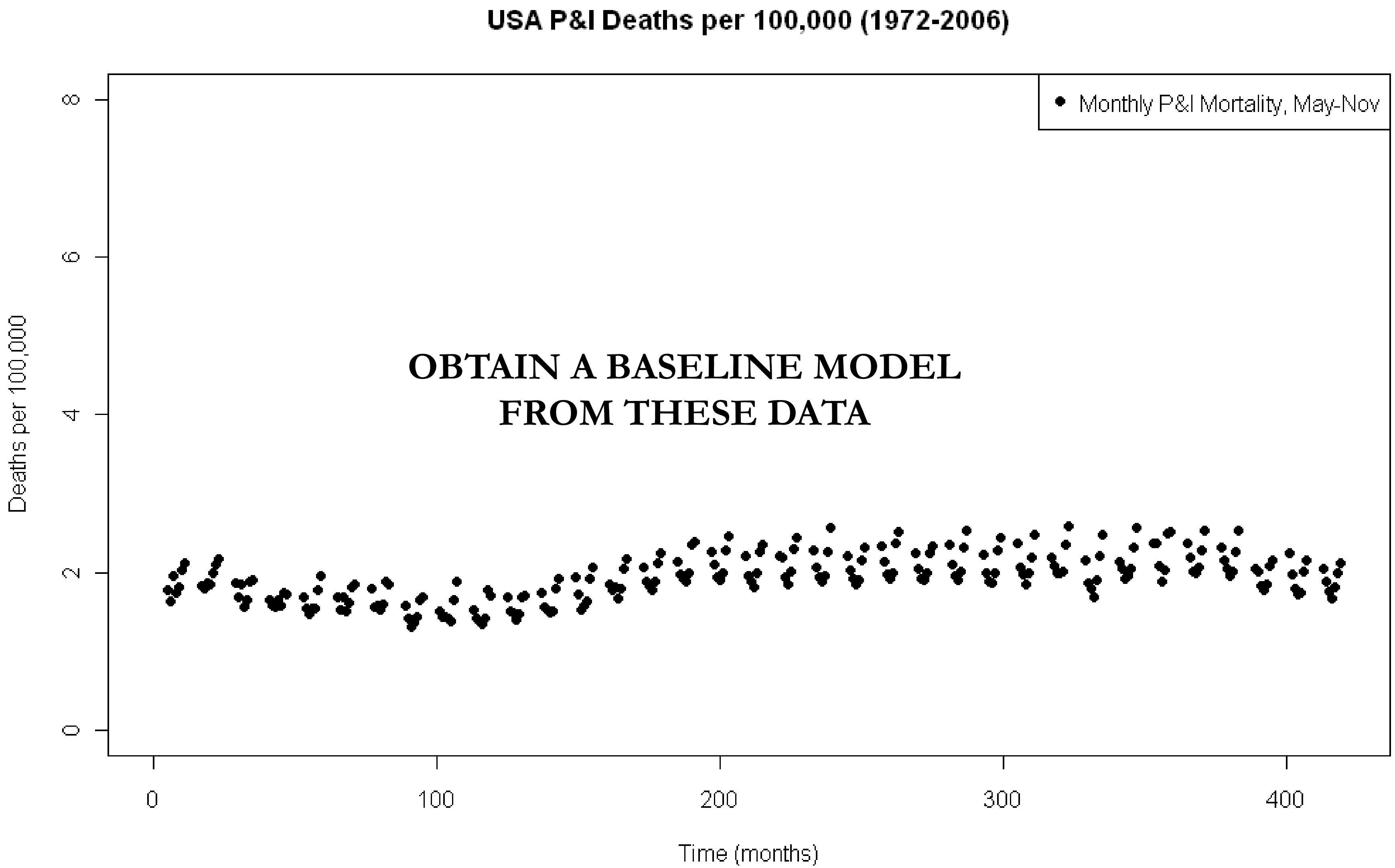
USA P&I Deaths per 100,000 (1972-2006), All Ages



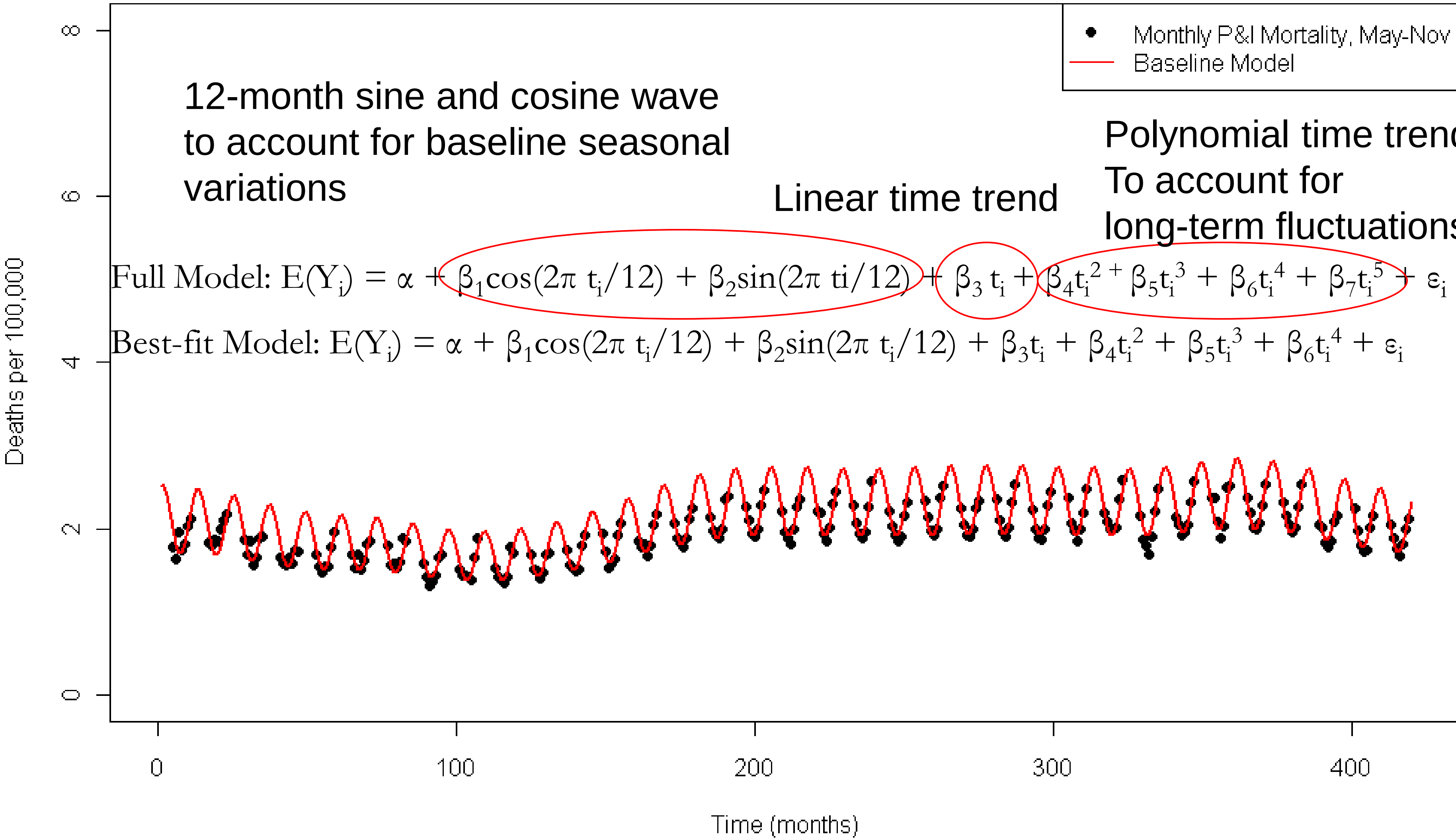
USA P&I Deaths per 100,000 (1972-2006), All Ages



USA May-Nov P&I Deaths per 100,000 (1972-2006), All Ages



USA May-Nov P&I Deaths per 100,000 (1972-2006), with Model Baseline, All Ages



12-month sine and cosine wave to account for baseline seasonal variations

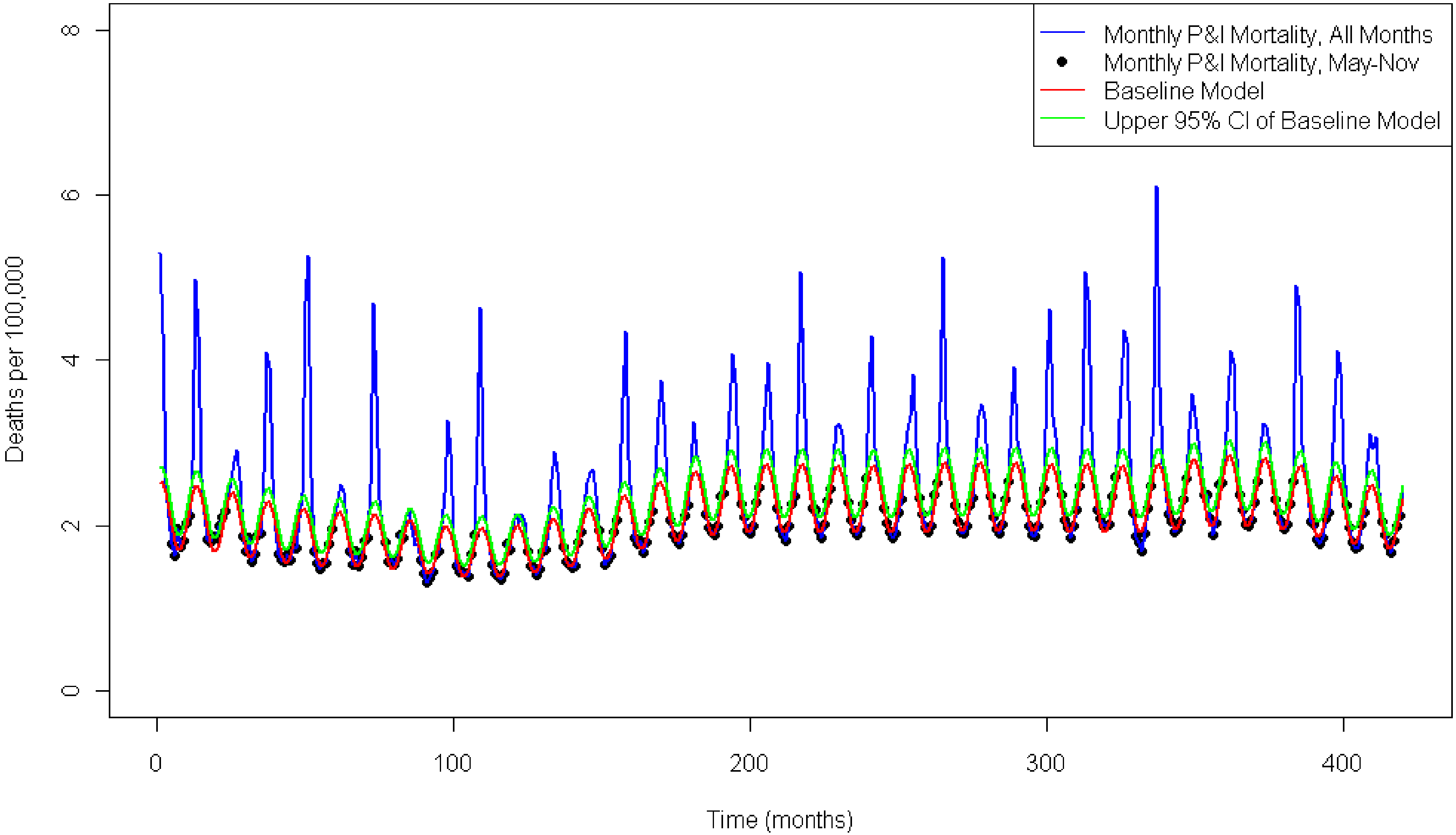
Linear time trend

Polynomial time trend To account for long-term fluctuations

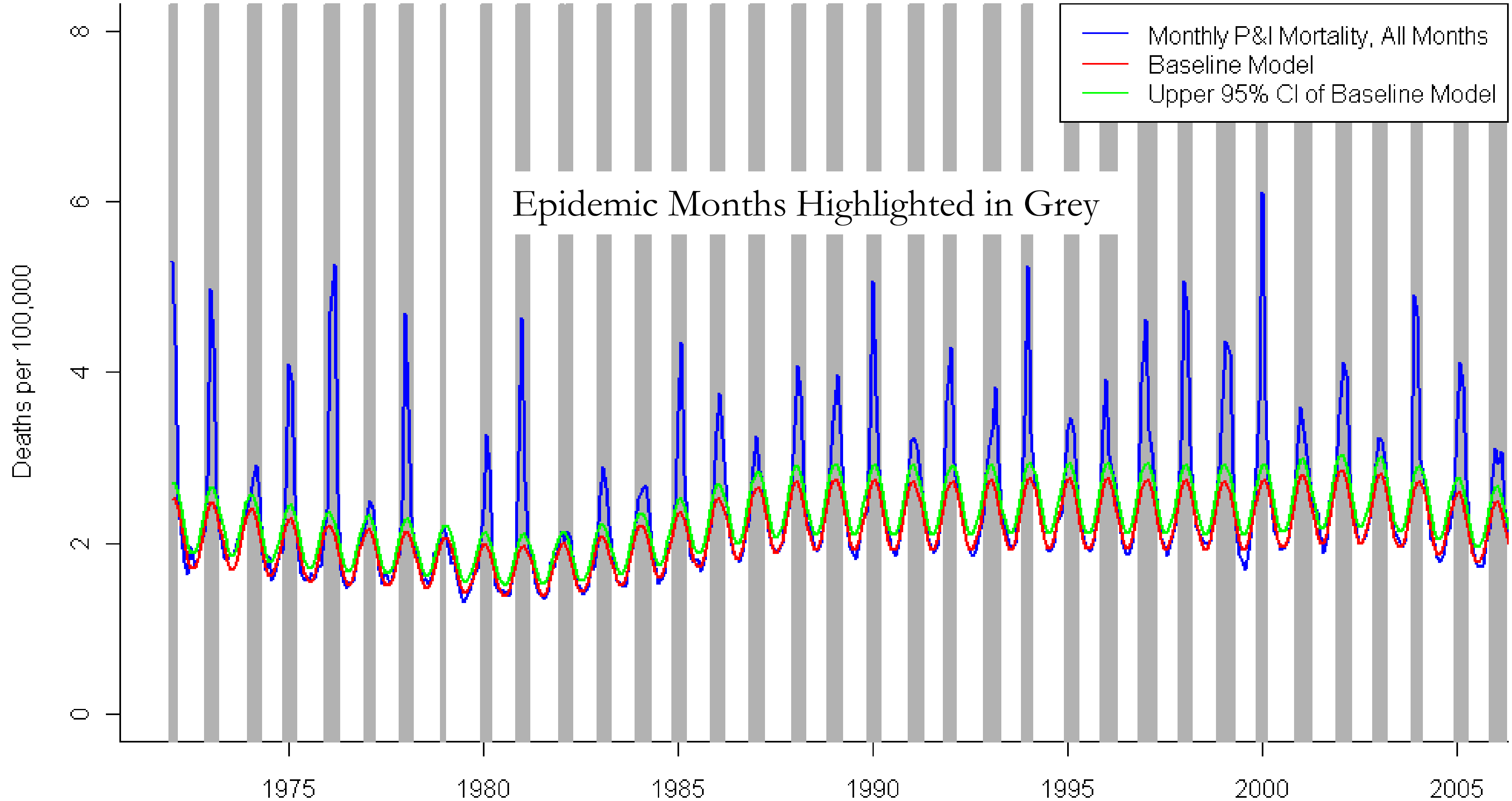
Full Model: $E(Y_i) = \alpha + \beta_1\cos(2\pi\ t_i/12) + \beta_2\sin(2\pi\ ti/12) + \beta_3\ t_i + \beta_4 t_i^{2} + \beta_5 t_i^{3} + \beta_6 t_i^{4} + \beta_7 t_i^{5} + \varepsilon_i$

Best-fit Model: $E(Y_i) = \alpha + \beta_1\cos(2\pi\ t_i/12) + \beta_2\sin(2\pi\ t_i/12) + \beta_3 t_i + \beta_4 t_i^{2} + \beta_5 t_i^{3} + \beta_6 t_i^{4} + \varepsilon_i$

USA P&I Deaths per 100,000 (1972-2006), Model Baseline and Upper 95% Confidence Band, All Ages

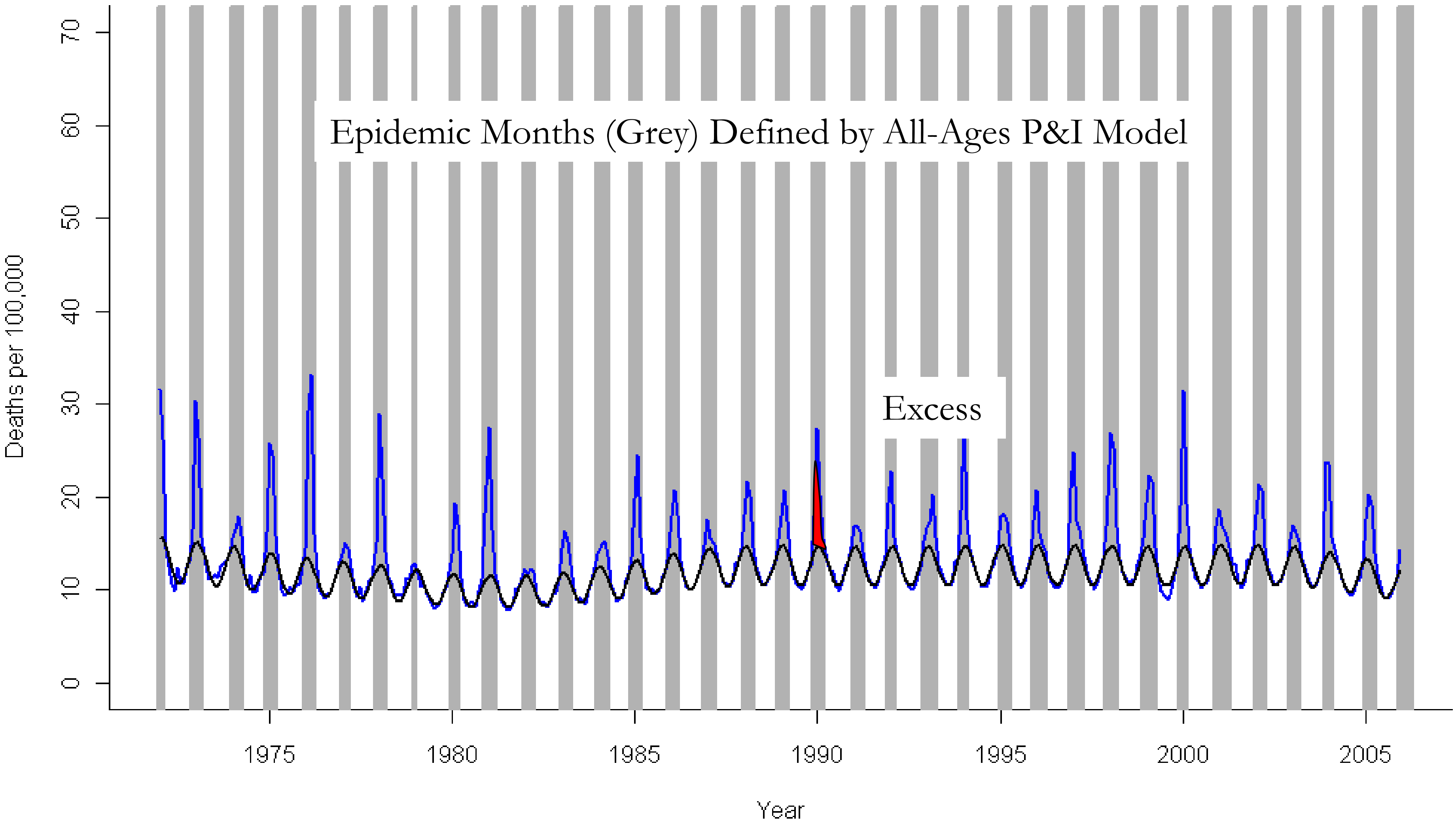


USA P&I Deaths per 100,000 (1972-2006), Model Baseline and Upper 95% Confidence Band, All Ages



USA P&I Deaths per 100,000 (1972-2006) and Model Baseline

65-89 Year Olds



Calculating Excess Mortality

- Monthly Excess Mortality:
  - For epidemic months (months in which the observed P&I mortality exceeds the upper 95% CI of the model baseline for all-ages):
    - ***Observed P&I mortality – model baseline predicted P&I mortality***
- Seasonal Excess Mortality:
  - For each influenza season (defined as Nov.-May in the US):
    - ***Σ Monthly excess mortality***

Seasonal US Excess Mortality Table

Season	Age Group	No. of Epidemic Months	Excess P&I Deaths per 100,000	Excess A-C Deaths per 100,000
1977/1978	65-89	4	30.41	137.05
1978/1979	65-89	1	0.75	7.40
1979/1980	65-89	3	15.68	66.85
1980/1981	65-89	3	28.45	136.71
1981/1982	65-89	3	3.67	27.35
1982/1983	65-89	4	12.68	47.83
1983/1984	65-89	4	9.51	33.90
1984/1985	65-89	4	21.38	122.61
AVERAGES				
-	-	3 (median = 3, range = 1-5)	41.34	78.37

Pros and Cons of the Serfling Approach

- Very flexible: can be used without virological data—especially useful for data on past pandemics
  – However, need at least three years of data
- Only works if disease is seasonal
  – Needs clear periods with no viral activity that can be used to create the baseline
  – Cannot be used as is for tropical countries that have year-round influenza circulation
    - There are techniques to adapt Serfling models for these purposes

An alternative/ complementary approach

- What proportion of “pneumonia and influenza” hospitalizations can be attributed to influenza?
- Use regression models with terms for seasonal variation, influenza, RSV (can be viral surveillance data, viral-specific hospitalization codes…)

A quick review of regression

- Linear regression:
  - $Y=\beta_1 x_1 + \beta_2 x_2 + a$
  - $\beta 1$: 1 unit increase in x1 results in β_1 increase in Y
- Poisson regression
  - Used when "Y" is a count variable/incidence rate rather than continuous
  - Usually has a skewed distribution
  - Multiplicative Poisson model: $Y=e^{(\beta_1 x1 + \beta_2 x2 + a)}$
  - If data are not Poisson distributed, use an alternative model, such as **negative binomial**

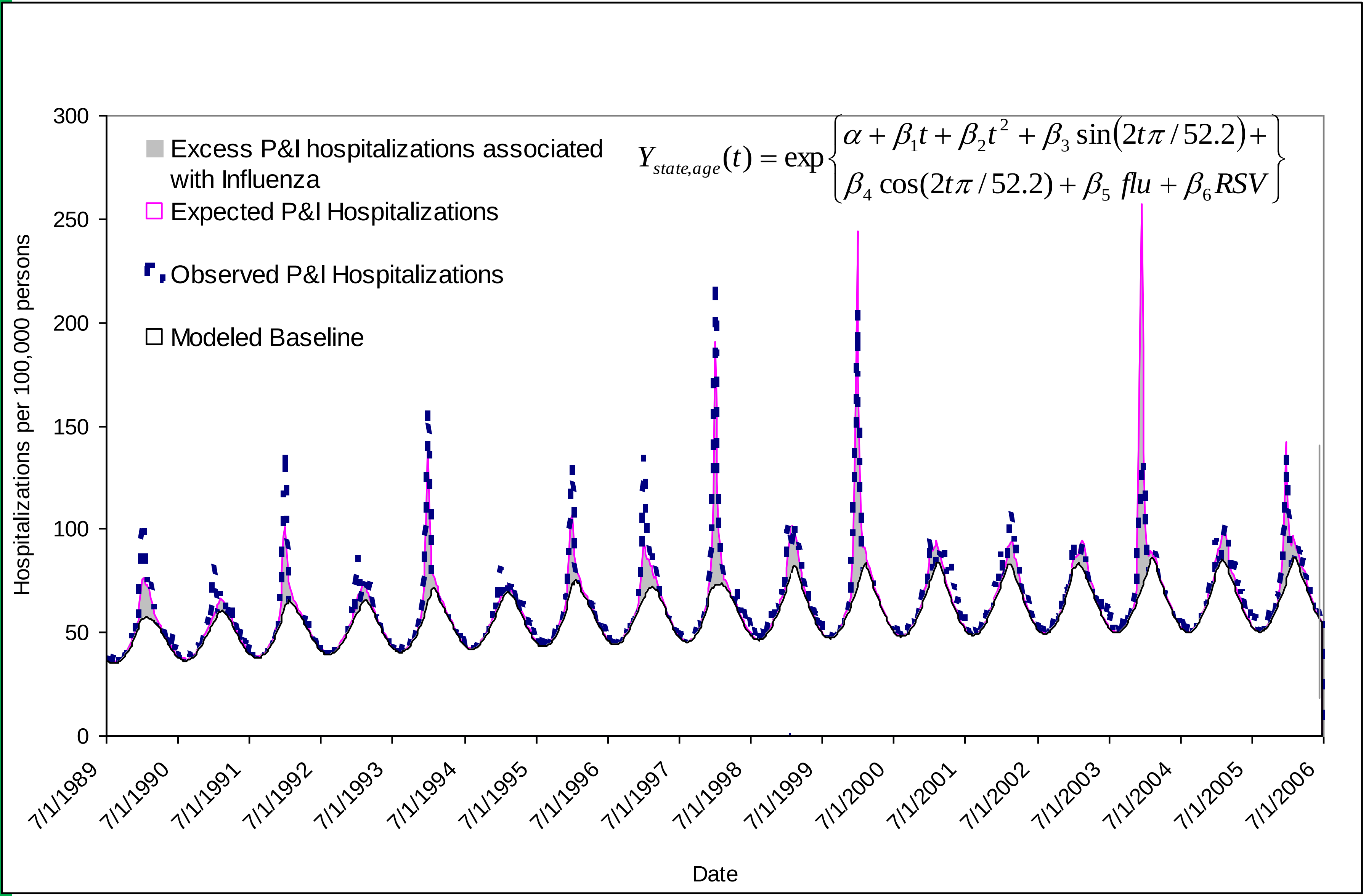
Estimation of influenza hospitalization burden

- Outcome = weekly pneumonia and influenza hospitalization rate

- Explanatory variables=influenza-specific and RSV-specific hospitalizations (proxies of viral activity),

- Seasonal estimates of influenza-related hospitalization rates obtained as sum of predicted rates minus baseline rates (influenza covariate set to 0)

Example: Estimation of influenza hospitalization burden in California in seniors



Comparison between Serfling and Poisson regression

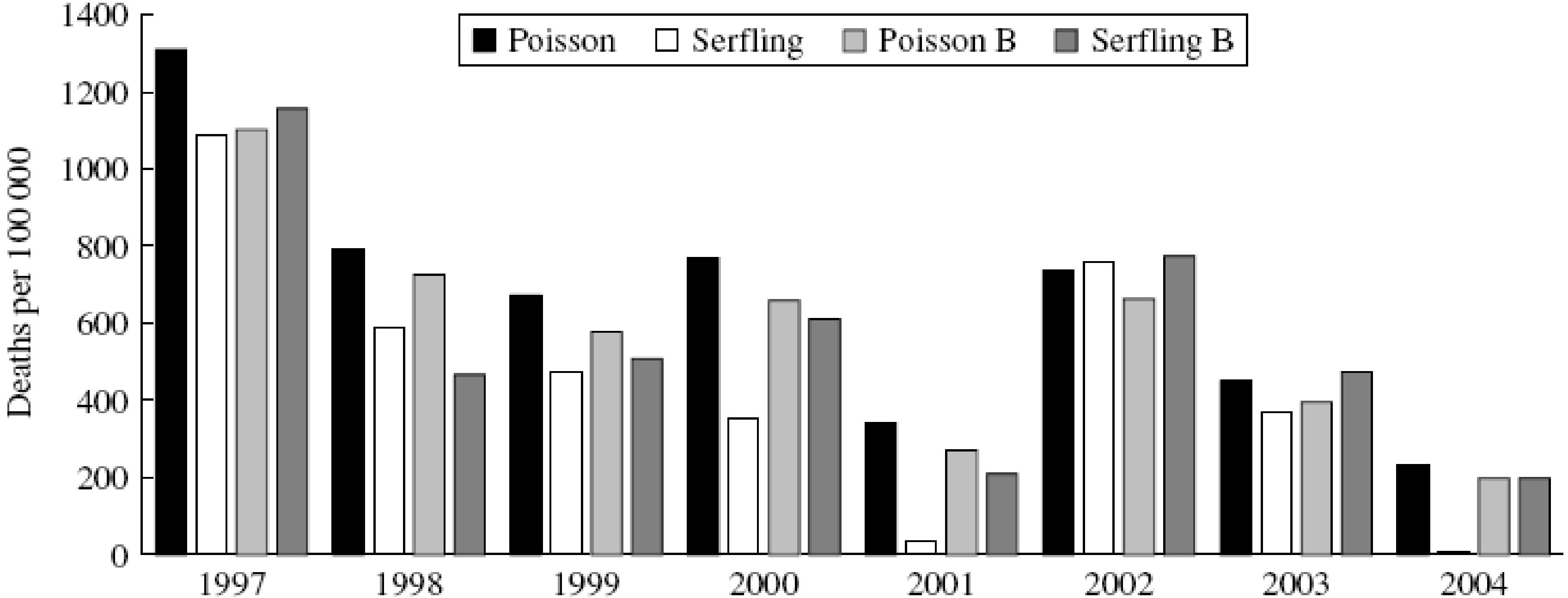


Fig. 4. Estimated all-cause influenza-attributable mortality by year in those aged ⩾85 years.

Newall, Viboud, and Wood, 2009 Epidemiolo Infect.

Alternative Models for Estimating Influenza Burden

- Peri-season models
  - Use months surrounding influenza epidemics as baseline
- ARIMA models
  - Estimate seasonal baseline by adjusting for serial autocorrelation.
- Serfling-Poisson combined models
  - Serfling seasonal excess mortality estimates are regressed against seasonal virus prevalence. Takes care of random variations in virus prevalence at small time scales
  - Iterative Serfling models (for non-seasonal data)

Validity tests for influenza disease burden models

- Regression diagnostics
- Checks based on the epidemiology of influenza
  - A/H3N2 vs A/H1N1/B dominant seasons (2-3 ratio)
  - RSV vs influenza (age!)
  - Higher rates in 65 yrs and over
  - Multiple years: seasons will little influenza circulation very precious!
- Difficult to estimate disease burden with precision
  - Mild seasons
  - Young children
  - Middle age groups

Acknowledgement

- Cecile Viboud for providing some R program samples
- Cécile Viboud and Vivek Charu for slides on Serfling regression
- http://www.ecs.syr.edu/faculty/lewalle/ (Jacques Lewalle) for some ideas on presentation