An Example of the Estimation and Display of a Smoothly Varying Function of Time and Space - The Incidence of the Disease Mumps*

William F. Eddy and Audris Mockus
Carnegie Mellon University
Pittsburgh, PA 15213-3890

June 30, 1993

Abstract

We consider any quantity which varies as a smooth function of space and time but is measured as the average value over a region in space-time. As a particular example, we consider the incidence rate of the disease mumps; the reported data are given on a state-by-month basis (for the period 1968-1988). By using animation techniques together with various smoothing methods we make a sequence of successively smoother dynamic maps which enable us to see the development of the disease in both space and time dimensions. We find that a smooth function (of the incidence rates) is the best way to convey the behavior of the spread of the disease. We consider various scaling techniques, ways to smooth the data, ways to estimate a smooth function of the incidence rates, and how to choose an appropriate color scale.

The paper also includes the description of algorithms to produce this kind of animation and a detailed description of the example concerning mumps disease in the United States.

*Partially supported by ONR Contract N00014-91-J-1024 and by Bureau of the Census under Joint Statistical Agreement 91-25. The equipment used to generate the videotape was acquired with partial support from NSF Grant DMS 88-05406.

1 Introduction

Most geographic data are reported as counts or averages over a region in space for an interval in time. We want to display the intensity function underlying the data using dynamic graphics. We will exemplify this problem with data on the disease mumps collected monthly in the United States from 1968 until 1988.

We created the dynamic graphic display from a sequence of images shown in a rapid sequence. We could not compute the images at the rate we wanted to see them, so we recorded them individually on a video recorder.

We begin in Section 2 with a detailed description of the problem we are trying to solve, namely estimation and display of a smoothly varying function of space and time. We also provide a detailed description of the data in the specific example we consider (mumps). In Section 3 we describe in some detail the animation methods we use. In Section 4 we consider the statistical problem of estimating a smooth function from averages over regions. Several different animations with varying degrees of smoothness in space and time are described in Section 5. We conclude with some unsolved problems in Section 6.

2 Problem Description

We are interested in the estimation and display of a smoothly varying scalar function f with a three dimensional argument (x, y, t). We think of x and y as indicating spatial coordinates and t as indicating a time coordinate. The data that we have available to estimate the function f consists of average values of the function over regions in space-time.

The natural way to display the function is by means of a dynamic graphical display. Because of the complexity of the estimation process the dynamics can only be done by "off-line" animation; that is, an interactive display cannot be computed in a reasonable amount of time. The animation consists of a sequence of images (frames) shown one after the other in a rapid sequence. Each image (frame) is a rectangular array of color values corresponding to every pixel of the display. A pixel (picture element) is the smallest element of a display that can be individually manipulated. Those color values indicate values of the function. Horizontal and vertical indexes of a pixel represent the two dimensions x and y and the sequence number of a frame represents the time dimension t of the function f.

2.1 Regional Averages

A great deal of geographic data is collected and reported as an average over a region in space and time. A list of examples includes:

- the weekly number of cases of a disease by city;
- the yearly total income by county;
- the monthly number of unemployed by state.

In each case one can imagine that the quantity of interest could be well-modeled by a smoothly varying function of time and space. The reported data is simply the average value of the function (with respect to some measure) over a region in space-time. For example, the monthly number of unemployed by state is the integral over the state of the unemployment rate (per unit population) for a month with respect to the measure of population density.

The problem of modelling a smoothly varying function of space has been considered by, for example, Tobler (1979). Tobler solves a discrete version of the Dirichlet integral subject to the boundary condition that the function is zero on the boundary of the region (or the condition that the derivative of the function is zero on the boundary of the region). Other related work concerning geographical interpolation is given in Tobler (1969), Tobler (1970), and Tobler and Kennedy (1985).

2.2 Mumps Data

The disease mumps causes severe morbidity and other side effects, particularly with increasing age. Mumps is preventable. The public health importance of the disease in the United States was recently highlighted by a large outbreak which occurred in 1986-1987, primarily among unvaccinated adolescents and young adults in states without requirements for mumps vaccination (see, e.g., Centers for Disease Control, 1989). Questions of public health importance concerning mumps include:

- Should specific populations be targeted with vaccine?
- How effective are interventions in the face of an outbreak?

Analytical questions which could affect decisions of public health importance include:

- Can annual or other periodicity be detected for mumps in the United States?
- Can geographic spread of mumps be demonstrated?

Disease maps (we are producing a dynamic disease map) are an important tool for answering such questions (see, e.g., Pyle (1979) or Cliff and Hagett (1992).

The Centers for Disease Control in partnership with the Council of State and Territorial Epidemiologists (CSTE) operates the National Notifiable Diseases Surveillance System (NNDSS) to provide weekly provisional information on the occurrence of diseases that are defined as "notifiable" by CSTE. The NNDSS also provides revised weekly, monthly, and annual totals. Further details concerning NNDSS can be found, for example, in Chorba et al. (1989). The NNDSS data are based on reports by state epidemiologists, who themselves receive reports from a variety of sources, such as individual practitioners, hospitals, laboratories, and health departments. Reports are received from all 50 States, Washington, D.C., New York City, and 5 United States territories.

The raw data we consider here consist of the revised total number of cases of mumps reported from each state for each month for the period 1968-1988. Data are not available for all states for all months during the period. Reasons a state may fail to report during a particular month include:

- increased workload of those who process disease surveillance reports,
- assumption that outbreak-related cases are known through other mechanisms,
- small numbers of cases may be combined with those for other reporting periods.

Also, the various states joined the NNDSS at different points in time.

The total data set contains 10,342 records. In 33 of the records the month is not specified; these records were removed from the file leaving 10,309 records. For comparison purposes, 48 states times 21 years times 12 months per year yields 12,096 possible state-month combinations. Thus there are 1787 missing (or unidentified) observations, approximately 15% of the possible combinations.

The reported counts were converted to reported incidence rates for each state (cases per 100,000 population per month) by dividing by the estimated population in units of 100,000 people. The state population estimates were obtained by linearly interpolating (or extrapolating) on a monthly basis from the 1970 and 1980 decennial census estimates of state population. Annual state-level population estimates were not readily available to us in computer-readable form. Given the qualitative nature of our maps we feel that using the linearly interpolated estimates has little or no effect on the results.

3 Animation Techniques

We are interested in the estimation and display of a smoothly varying scalar function f with a three dimensional argument (x, y, t) as described at the beginning of Section 2.

The following features of an animation should be noted:

1) The time dimension is substantially different from the space dimensions. For example, it would be very difficult to understand our animation if we had equated (x, t) in the data with (x, y) on the screen and equated y in the data with t on the screen.

2) The eye can not readily distinguish a single pixel from its neighbors.

3) Nonsmooth changes in time are more difficult to detect than similar changes in space.

The first and third properties suggest that interpolation in space could be done independently of interpolation in time.

The second property suggests that our computationally-intensive spatial interpolation method could be used to interpolate to some subset of the image pixels (preferably a regular grid) and the remaining pixels could be filled using simple bi-linear (see Section 4.3) interpolation (from this regular grid). This would save computation time without degrading the percepted smoothness of the animation.

In the next subsection we consider scaling the animated function (its values and its arguments) to fit into the range of available colors, pixels, and video frames. Then we consider temporal interpolation to animate the data and then consider smoothing the raw data before we animate it.

3.1 Scaling

Once the function of interest $f(x, y, t)$ is determined, an animation is like a generalization of plotting to three dimensions. To produce an animation one has to scale a four-dimensional object into the graphical device coordinates. Plotting a function of a one dimensional argument only requires scaling of a two-dimensional object (the values of a function and its argument) to fit on the screen or paper.

A function of three dimensions has to be scaled to be shown on a device with finite spatial and temporal resolution. We map the range of the function values into 256 color values. The region where the function is shown is mapped into the 512×512 pixels of our display and the time interval of one or two minutes (there are 1800 video frames per minute). These values are limitations of human perception and/or video equipment. According to experiments (see Levkowitz and Herman, 1992) people can clearly distinguish around 120 color values on the appropriately chosen color scale. The number of distinct pixels on a television screen is limited by the bandwidth of the display device screen imposing a range of possible values for two dimensions x, y. We find that more than one or two minutes of technical video is often boring to anyone other than a subject-matter specialist, limiting possible values in the remaining dimension t.

Mumps incidence decreases dramatically during from 1968 to 1988. In order to better use the color scale we made a nonlinear transformation of the incidence rates before transforming them linearly into colors. The transformation we chose was to use the rank of each incidence rate in the whole data set rather than the value. The resulting colors are then uniformly distributed over the color scale. We tried using the data values and the logarithm of the data values, but in both cases only hte large variations of incidence rates in the beginning of the period (1968-1978) were detectable, leaving the later period without much visible action. In the image processing literature such a transformation of the pixel intensities is called "equalization" (see, for example, Pavlidis, 1982). The actual color scale is displayed at the bottom of the frame with the corresponding incidence rates given just above the colors.

3.2 Temporal Interpolation

The entire mumps data set consists of 252 months. NTSC video is displayed at the rate of 30 frames per second (NTSC is the television signal that is used in the United States and Japan). After several experiments we decided that displaying the data at the rate of 20 frames per month was a reasonable compromise between the time required to look at the entire data set and the apparent speed with which changes take place. Thus each month is displayed for two-thirds of one second. If the recording were done so that twenty identical frames were recorded and then the switch were made to the next month's data, the viewer would be distracted by the jumpiness of the resulting images (see Section 5.1.1). Consequently, we chose to interpolate linearly between consecutive months. Precisely, the correctly colored maps for two consecutive months are calculated and then 19 intermediate maps are calculated by linear interpolation in the color scale. This results in substantially smoother appearance.

We considered other types of temporal interpolation (sinusoidal, trapezoidal, and quadratic) but decided that linear interpolation was preferable. Apparently, either we are sensitive to the maximum value of the derivative of the interpolating function (linear interpolation minimizes this maximum value) or we are sensitive to a non-zero second derivative of the interpolating function (linear interpolation has a zero second derivative).

3.3 Smoothing Raw Data

Observed data usually contains a substantial amount of noise which, if not removed, can produce a "jumpy" animation which, in turn, could hide interesting features of the animated process.

In time series analysis data are frequently smoothed using running averages or running medians to reduce noise. Given a series of observations z_i, the value of the running median at time i is

$$\hat{z}_{i,k} = \text{Median}\{z_j : |j - i| \le k\},$$

where k is called the size of the running median. Our definition of running median of size k is often referred to as "running median of $2k+1$", but the latter terminology does not extend to the multidimensional case. We prefer to use medians (as opposed to means) because means are not invariant under the transformation to ranks that we used for the color scale.

Tobler and Kennedy (1985) used an interpolation from spatial averages to fill in missing values. We use spatial (and space-time) medians to smooth the data and to fill in missing values.

In the mumps data there are both time and space components, so we could do running medians in time for every region, do moving medians in space for every time moment, or do moving medians in space and time together. To define moving medians in space we need to define adjacencies between the locations of observations because the simple (total) ordering by time is no longer present. In our case, regions A_i (states) form a partition of A (the continental US). We define two spatial regions to be adjacent (or 1-adjacent) if they share a common border consisting of more than one point. If there is a region to which they both are adjacent then we call them 2-adjacent. Note that a pair of regions that are 1-adjacent are automatically 2-adjacent. Similarly we can define k-adjacent regions. Given the values z_i for regions A_i we define a moving median of size k at the region A_i as $\hat{z}_{i,k} = \text{Median}\{z_j : A_j$ is k-adjacent to $A_i\}$. The time dimension can be thought of as just another space dimension and then we can apply the moving medians in space and time simultaneously.

We successfully used those techniques to improve the perceived smoothness of animations. We found that moving medians of size one (in space-time) produce a substantial amount of smoothing (see Section 5.1.3).

4 Estimation from Spatial Averages

4.1 The Problem

The problem of interest is to estimate a function $f(x, y, t)$ (incidence rates of the mumps disease at some location and time moment (x, y, t) given sets

$\{A_j\}$ and data $\{z_j\}$. Henceforth we will denote a space-time location as $\mathbf{x} = (x, y, t)$ to simplify the notation. The relationship between f and data is given by following equation

$$z_j = \int_{\mathbf{x} \in A_j} f(\mathbf{x}) dG(\mathbf{x}), \quad j = 1, \ldots, N$$

where $A_j \subset A \subset R^3$ and $G(\mathbf{x})$ is the population distribution.

One could assume that $f(\mathbf{x})$ is a fixed but unknown function. The problem then would be simply one of interpolation. Alternatively, one could assume that $f(\mathbf{x})$ is random. Then one would look for the predictor $\hat{f}$ that minimizes the mean square error (MSE),

$$MSE(\hat{f}(\mathbf{x})) = E((f(\mathbf{x}) - \hat{f}(\mathbf{x}))^2).$$

The function $f(\mathbf{x})$ has to be estimated over some set of $\mathbf{x} \in A \subset R^3$, so the integrated MSE

$$\int_A MSE(\hat{f}(\mathbf{x})) d\mathbf{x},$$

or maximal MSE

$$sup_{\mathbf{x} \in A} MSE(\hat{f}(\mathbf{x}))$$

could be of interest depending on the problem at hand.

Let $f(\cdot)$ be a zero mean stationary process and $c_i(\mathbf{x}) = E(f(\mathbf{x}) \cdot z_i)$, and let $C = (c_{ij})$ where $c_{ij} = E(z_i \cdot z_j)$. Then the minimum MSE predictor for $f(\mathbf{x})$ would be

$$\hat{f}(\mathbf{x}) = c(\mathbf{x}) C^{-1} \mathbf{z}, \tag{1}$$

where $c(\mathbf{x}) = (c_i(\mathbf{x})$ is a vector of length N, C^{-1} is the inverse of the $N \times N$ covariance matrix for the $\{z_i\}$, and $\mathbf{z} = (z_i)$ is the observation vector of length N. When the assumption that $f(\cdot)$ is a zero mean process is unreasonable the mean could be estimated taking global or local averages of the observations z_i.

Equation 1 has some drawbacks. The assumption that $f(\cdot)$ is stationary is probably unreasonable. Also, Equation 1 requires inversion of the matrix C as well as knowledge of the covariance function of the process to obtain $c_i(\mathbf{x})$ and c_{ij}. In the case of observations at a point ($z_i = f(\mathbf{x}_i)$) there are parametric and nonparametric ways to estimate the covariance function (see, e.g., Cressie, 1991). When data are aggregate, as is in our case, it is still possible to estimate the covariance function (see Mockus, 1994). Unfortunately, covariance function estimation is difficult as is the method of Tobler (1979). We considered alternative simpler solutions.

The estimation problem we are considering could be modified so that the interpolation would be done given the values (instead of integrals) of f at some points. This approach is described in next section.

4.2 Transforming the Problem

For the interpolation problem when data are values of the function at some points there exist a wide range of fast and simple-to-implement algorithms. A kernel estimator for $f(\mathbf{x})$ given $z_i = f(\mathbf{x}_i)$ is

$$\hat{f}(\mathbf{x}) = \frac{\sum_i K(\mathbf{x}, \mathbf{x}_i) z_i}{\sum_i K(\mathbf{x}, \mathbf{x}_i)},$$

where $K(\mathbf{x}, \mathbf{x}_i)$ is a kernel function. When data are aggregate

$$z_i = \frac{\int_{A_i} f(\mathbf{x}) dG(\mathbf{x})}{\int_{A_i} dG(\mathbf{x})}$$

one could define an estimator by analogy

$$\hat{f}(\mathbf{x}) = \frac{\sum_i z_i \int_{A_i} K(\mathbf{x}, \mathbf{x}_i) dG(\mathbf{x}_i)}{\sum_i \int_{A_i} K(\mathbf{x}, \mathbf{x}_i) dG(\mathbf{x}_i)}.$$

To approximate the integrals $\int_{A_i} K(\mathbf{x}, \mathbf{x}_i) dG(\mathbf{x}_i)$ we took a sample of points uniformly distributed within each state A_i. The number of points sampled in each state was taken proportional to the area of the state. We assigned the value to those points to be equal to the incidence rate for the particular state the points are in. In this way we take into account the differing areas and shapes of the states. This method cannot take account of the distribution of population within a state because we have only one estimate of the incidence rate for each state.

These sampled points are used to interpolate a function to every pixel on the map. We used a weighted combination of the function values at the sampled points to obtain the value at all pixels. The weight function $K(\mathbf{x}, \mathbf{x}_i)$ was chosen to be exponential in the squared distance between the sampled point $\mathbf{x}_i$ and the pixel $\mathbf{x}$ where the function was being interpolated.

The method we used to estimate $f(\mathbf{x})$ in the animations can be described as follows:

- Choose a set of points and a set of values for every A_j, $\mathbf{x}_{ij} \in A_j, z_{ij}$, $i = 1, \ldots, k_j$.
  - We took the number of points k_j in the region A_j to be proportional to the area of A_j.
  - The points $\mathbf{x}_{ij}$ are distributed in A_j so that they repel each other and the boundary of A_j. A point $\mathbf{x}_{i+1,j}$ is sampled uniformly from the set $A_j \setminus \cup_{k=1}^{i} r_k$, where r_k is a disk with a center at $\mathbf{x}_{kj}$. The radius of r_k depends on the total number of points to be sampled from A_j and on the size of A_j.
  - The values of f at $\mathbf{x}_{ij}$ are assumed constant for each A_j, $z_{ij} = z_j$.
- Use the estimator
$$\hat{f}(\mathbf{x}) = \frac{\sum_i K(\mathbf{x}, \mathbf{x}_{ij}) z_{ij}}{\sum_i K(\mathbf{x}, \mathbf{x}_{ij})}$$
  with $K(\mathbf{x}, \mathbf{x}_{ij}) = e^{-\lambda ||\mathbf{x}_{ij} - \mathbf{x}||^2}$.
- Choose the smoothing parameter λ to provide an acceptable degree of smoothness to the animation.

4.3 Two Levels of Interpolation

Estimation using exponential weights for each point as described in Section 4.2 can be very time consuming. The frame buffer (the device that generates the NTSC video signal) has more than 500×500 pixels. Assuming an average of 10 points in each state where the value of the function is assumed to be given, we have to perform approximately 10^8 distance and exponential weighting calculations for each frame. This is a substantial amount of time even on a fast workstation given that we want to record 252×20 of those frames.

Consequently the weighted estimation was performed only onto a regular grid over the United States. We chose the grid size to be 35 points by 25 points. We then used a bi-linear interpolant from the four values at the corners of each of the 34 times 24 rectangles of the regular grid to each pixel within a particular rectangle. The weights for each pair (regular grid point, sampled point) are computed only once and stored.

5 Videotapes

The animations were produced one after another, improving the result at each step. Despite those improvements the first steps are of interest by themselves.

The simplest possible animation is to display the raw data: a constant value of the incidence rates for each state during every month. The result is difficult to understand due to sharp changes between adjacent states, abrupt changes in time, and abundance of unreported cases. This animation creates a desire for a smoother picture in space and in time.

A smoother looking picture can be produced by smoothing the raw data and estimating missing values. In this case the smoothed data is displayed as being constant across each state and interpolated between months. The interpolation between months removes "jumps" in time. We used linear interpolation between months as described in Section 3.2. Various approaches are possible to smooth the data in space. From a practical point of view, running medians in space and time (see Section 3.3) is a simple method that also fills in the missing values. Although it is an improvement over the first step this animation still has jumps at state boundaries.

The last step was to produce a smooth animation both in time and space. This required use of techniques described in Section 4.2. The animation that is smooth in time and space turned out to be visually appealing and easier to understand.

A brief description of the equipment we used to produce those animations can be found in Eddy and Mockus (1993). A copy of a VHS videotape in NTSC format containing the displays generated to date is available from the authors at cost. The tape contains about twenty minutes of video.

5.1 Nonsmooth in Space

5.1.1 Nonsmooth in Time

We have used the background color to indicate missing data. The states which are missing seem to "disappear" into the background when there is no data. An initial version of the videotape switched instantaneously from a color to background when there was a missing observation and then back to a color from background when there was data. The abruptness of this scheme

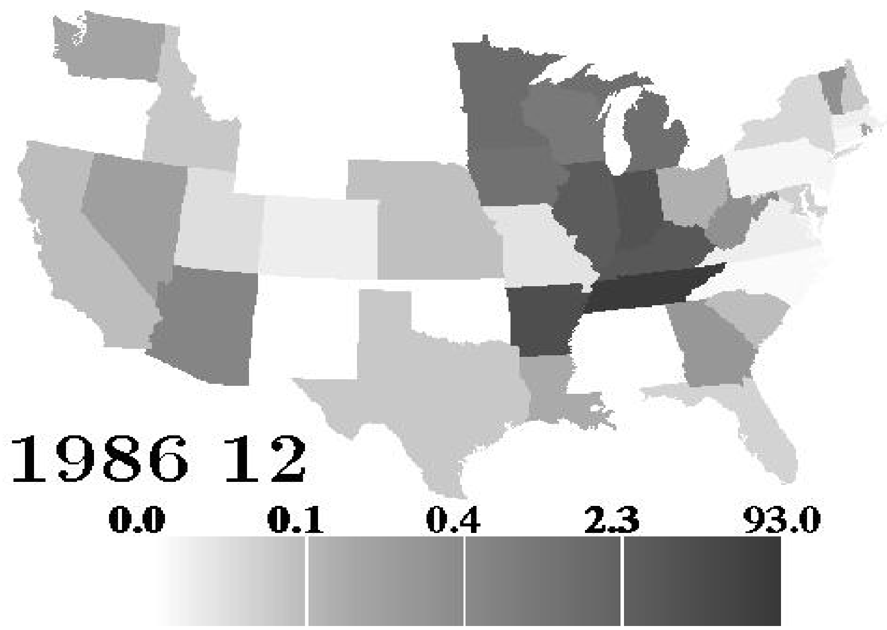


Figure 1: Raw Incidence Rates in December 1986

was sufficiently jarring that we modified the scheme to "fade" to background. This is actually done by linear interpolation between the particular color and the background color. One frame of this animation is displayed in Figure 1.

The time smoothing was performed as described in Section 3.2. Simple linear interpolation in time was the first method we used. We tried other interpolation methods but could not detect any improvement.

5.1.2 Filling in Missing Data

The number of states not reporting mumps cases increases in the later part of the data. This distracts the viewer from the overall pattern of the disease. We used methods described in Section 3.3 to fill in the missing values. To indicate the fact that the value was not reported we used a dotted fill pattern for the particular state. This way it was possible to show the overall predicted pattern of the disease together with information showing which part of the data was actually reported.

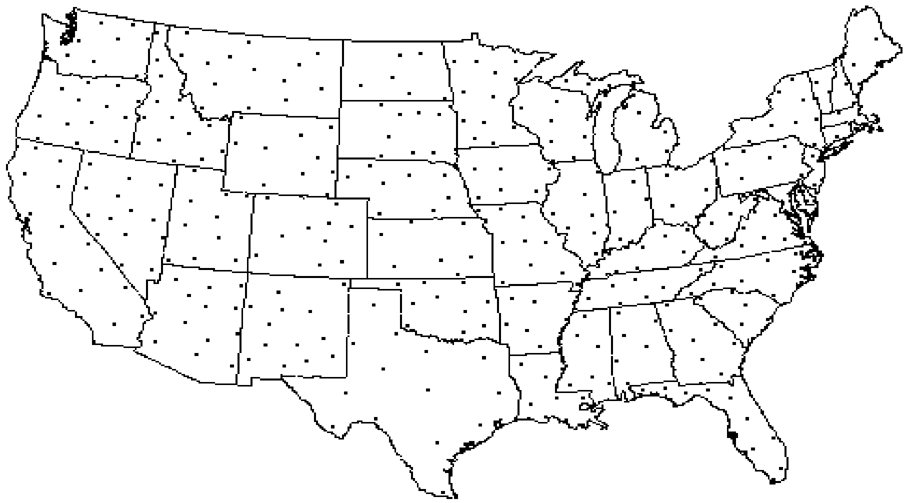


Figure 2: The locations of points used in smoothing

5.1.3 Smoothing in Missing Data

In an attempt to reveal the major patterns in the data we used moving medians as described in Section 3.3 not only to fill in the missing values but also to smooth the existing values. This resulted in large regions in space and time having roughly the same color.

5.2 Smooth in Space and Time

The smoothest animation was produced using independent time and space smoothing. The smoothing in space was done for every month. First we estimated the intensity on a regular 35×25 grid of points using the algorithm described in Section 4.2. The particular set of points $\mathbf{x}_{ij}$ (using the notation of Section 4.2) is shown in Figure 2.

To obtain estimates for the remaining pixels we used a simple bi-linear interpolation described in Section 4.3. As in previous animations we chose to interpolate linearly the 19 intermediate frames between the monthly smoothed maps. Thus the smoothing in space and in time are independent of each other. The single frame corresponding to December 1986 is displayed in Figure 3.

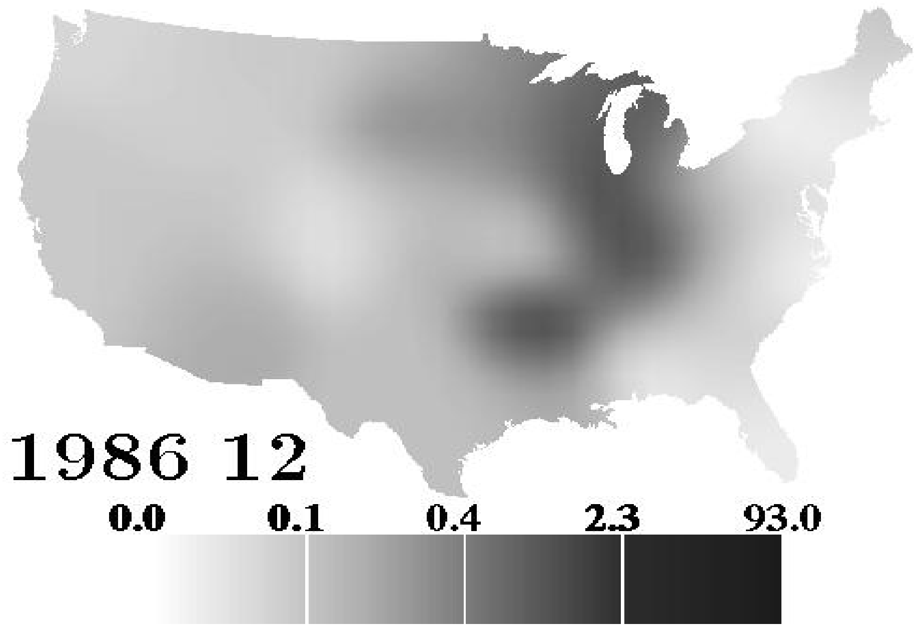


Figure 3: Smoothed Incidence Rates in December 1986

5.3 Detecting Disease Outbreaks

As an alternative to showing the incidence of the disease we considered inspection of the residuals from a simple statistical model. This approach was intended to emphasize outbreaks of the disease and mask normal patterns such as seasonal variations and different reporting practices across the states.

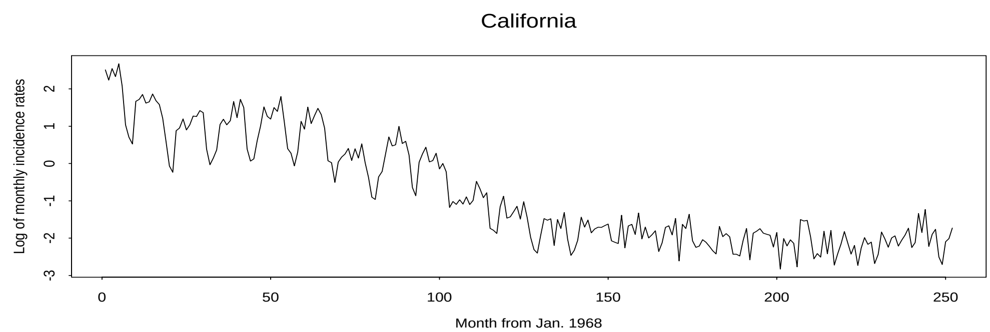
In this video we considered the later period of the disease (1980-1988) when the incidence rates have stabilized after the steep drop that was caused by the introduction of vaccination programs at the end of 1960's.

Let z_{ij} be the logarithm of the reported incidence rates in state i for month j (we added 1 before taking the logarithm to avoid problems with zero incidence rates). We used median polish (see, e.g., Siegel, 1983) to fit additive state effects s_i and time effects t_j. The residuals

$$\eta_{ij} = z_{ij} - s_i - t_j$$

for any particular state looked like a stationary time series except for one or two peaks caused by larger outbreaks.

To emphasize the outbreaks we smoothed out the "small" noise leaving only extreme peaks. We defined an η_{ij} to be unusual if it was in the upper



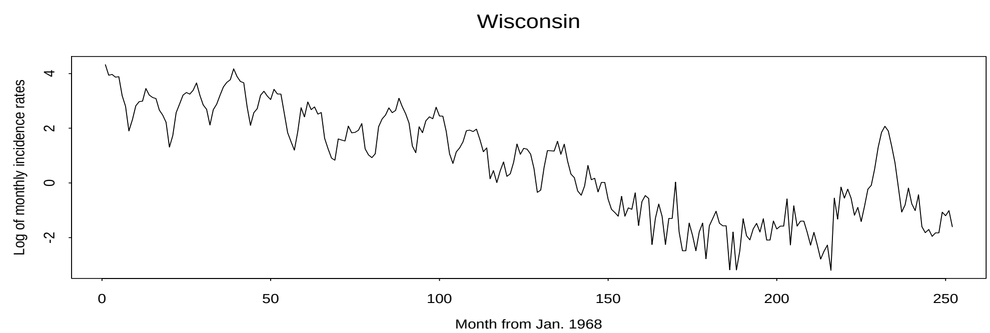


Figure 4: Log of the mumps monthly incidence rates versus months from Jan. 1968 to December 1988

0.95 quantile of the residuals. We then applied running medians of size 3 in time for every state to the residuals that were not considered unusual, i.e. we chose

$$\hat{\eta}_{ij} = \begin{cases} \text{Median}(\eta_{ij} : |j-k| \leq 3) \text{ if } \eta_{ij} \text{ was not unusual} \\ \eta_{ij} \text{ otherwise.} \end{cases}$$

The resulting animation identifies what one could define as an outbreak of the disease without confusing the scene with the seasonal and between state effects.

6 Discussion

Mumps in the US is a seasonal disease. The peak occurs in early spring, while the lowest incidence rates can be observed in autumn. As most of the cases are school age children, this can be in part explained by the school year. Over a longer period the mumps disease had a high incidence rate before the vaccination programs started around 1970. By 1980 these vaccination programs almost completely eradicated the disease, leaving only a few cases per state per month. Some states ceased mandatory vaccination programs at about that time and strong outbreaks of the disease occurred in 1986-1987, primarily among unvaccinated adolescents and young adults in those states. These statements are clearly supported by the graphs in Figure 4 of the logarithm of the incidence rates in California and Wisconsin. We can see seasonal periodicity (high in spring and low in autumn) and an outbreak in Wisconsin in the second half of eighties.

Annual periodicity in the incidence rate for mumps can be observed in both the raw data videos and the smoothed versions. The periodic effect is particularly striking in the early years of the data set, before the widespread use of the mumps vaccine reduced the typical monthly incidence rate below .1 (cases per 100,000 people). However, the effect can be discerned throughout the data set, especially in the smoothed version.

The geographic spread of mumps cannot be easily discerned in the raw data; however, repeated viewing eventually allows one to make such an interpretation. The effect is probably most noticeable in the winter of 1987-1988 in the states surrounding Illinois. In the smoothed data the geographic spread of the disease is readily apparent. This is particularly clearly visible during the late winter of 1986-87 when the disease spreads from Illinois to Arkansas and Tennessee and in the subsequent winter when the disease spreads to all the neighboring states.

References

Becker, R.A. and Wilks, A.R. (1991). Maps in S. Statistical Research Report No. 96, AT&T Bell Laboratories, Murray Hill, NJ.

Centers for Disease Control (1989). Mumps – United States, 1985-1988, *Morbidity and Mortality Weekly Report,* **38**(7): 101-5.

Chorba, T.L., Berkelman, R.L., Safford, S.K., Gibbs, N.P., and Hull, H.F. (1989). Mandatory Reporting of Infectious Diseases by Clinicians. *Journal of the American Medical Association*, **262,**: 3018-3019.

Cliff, A.,D., Haggett, P. (1992). *Atlas of Disease Distributions,* Blackwell Publishers.

Cressie, N.A.C. (1991). *Statistics for Spatial Data.* John Wiley and Sons, New York.

Eddy, W.F. and Mockus A. (1993). An Example of Noninteractive Dynamic Graphics for Manufacturing Process Data. In *International Statistical Review,* **61**(1): 81-95.

Levkowitz, H., Herman, G.T. (1992). Color scales in image data, in *IEEE Computer Graphics & Applications* January, 1992. 72-80.

Mockus A. (To appear). PhD Thesis. Department of Statistics, Carnegie Mellon University.

Pavlidis, T. (1982). *Algorithms for Graphics and Image Processing.* Computer Science Press, Rockville, MD.

Pyle, G.F. (1979). *Applied Medical Geography.* John Wiley, New York, NY.

Siegel, A.F. (1983). Low median and least absolute residual analysis of two-way tables. *Journal of the American Statistical Association.* **78**: 371-374.

Tobler, W.R., (1969). Geographic filters and their inverses, *Geographical Analysis*, **1**(3): 251-257

Tobler, W.R., (1970). A computer movie simulating urban growth in the Detroit region, *Economic Geography*, **26**(2):234-240

Tobler, W.R., (1979). Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association.* **74**(367): 519-536.

Tobler, W.R., Kennedy, S. (1985). Smooth multidimensional interpolation. *Geographical Analysis.* **17**(3): 251-257.