Noninteractive Display of Functions with Temporal and Spatial Variation (or Surreal-time Graphics) *

William F. Eddy Audris Mockus Department of Statistics Carnegie Mellon University Pittsburgh, PA 15213-3890

December 31, 1992

Abstract

For most statisticians the notions of dynamic graphics have been inextricably linked to the notions of interactive graphics. This paper begins by describing the reasons why not all dynamic graphics can be interactive. Then a very brief description of a system which can generate dynamic (but not interactive) graphics is given. Three specific example applications of the system will be described in some detail: (1) an earth sciences example in which we animate the daily sea-surface temperature; (2) an industrial example in which we animate the spatial-temporal distribution of temperature across a metal plate; and (3) an epidemiological example in which we animate the reported incidence rate of the disease mumps for the continental 48 states over a twenty-one year period.

^{*}Partially supported by ONR Contract N00014-91-J-1024 and by the 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

Graphical displays are widely utilized for the analysis of data and are invaluable for the presentation of results. Univariate and bivariate data are easily displayed. Effective display of higher dimensional data presents a challenge. Usually displays of higher dimensional data are presented by projecting into (or conditioning on) a two-dimensional subspace. Often, a sequence of subspaces are displayed creating a dynamic display. Generally, dynamic graphical displays provide an enormous increase in the amount of information conveyed to the viewer compared to static displays.

A basic problem with either dynamic or static displays of multidimensional data is the selection of which two-dimensional subspaces to use. One general technique which has been fairly widely used is the "Grand Tour." This method and other similar methods provide a smooth sequence of projections for viewing and thus lend themselves naturally to dynamic displays. A recent study of projection methods which specifically considers the interactive approach to dynamic graphics is Hurley and Buja (1990).

An alternative approach to high-dimensional graphics based on parallel coordinate plots has been pioneered by Inselberg (1985). The basic idea is quite simple: instead of picturing coordinate axes as being orthogonal, display them as parallel. This has the great advantage that a quite large number of coordinates can be displayed simultaneously and the great disadvantage that the resulting pictures depend strongly on the ordering of the coordinates. Furthermore, since few users have knowledge of projective geometry, interpretation of observed structure in the data is often difficult.

Techniques for the dynamic display of multidimensional data are of particular interest in statistical applications. The collection edited by Cleveland and McGill (1988) contains descriptions of several interactive dynamic graphical methods. Unfortunately, all of those methods are only suitable for relatively low dimensional problems (say, up to 9). With many tens of variables and hundreds of observations (or even larger data sets), a new approach will be needed to avoid throwing away the bulk of the information for the sake of a low dimensional display.

This paper considers the noninteractive display of dynamic graphics. In the next section we briefly describe the need for such graphics. In Section 3 we describe the technology we use to make the displays. We then consider three specific examples and briefly describe the methods we used to create

CPU TIME

		High	Low
	High		Real
USER		Surreal	Time
TIME	Low	Time	

Table 1: Real Time and Surreal Time

the displays. An important common feature of our three examples is that they can be viewed as space-time series.

2 Surrealism

Consider a two-way table in which the horizontal margin corresponds to the amount of computer time needed to generate a dynamic graphical display and the vertical margin corresponds to the amount of time a user is willing to wait for the computer to generate the display. Such a two-way arrangement is shown in Table 2. The upper right portion of the table corresponds to dynamic graphical displays which can be achieved interactively (that is, in real time) and the lower left portion of the table corresponds to displays which cannot be achieved interactively (thus, *surreal* time). An example of a display which requires a large amount of CPU time is one which shows a smooth function where the value of the function at each pixel is determined by some complicated calculation. If the function varies with time (or some other continuous variable) we have a dynamic graphical display. This cannot be achieved interactively and must therefore be accomplished in surreal time.

According to the American Heritage Dictionary of the English Language surrealism is

A literary and artistic movement launched in 1924 by the French poet André Breton (1896-1966), proclaiming the radical transformation of all existing social, scientific, and philosophical values through the total liberation of the unconscious.

In the oral presentation of this material a copy of Salvador Dali's oil

painting "The Persistence of Memory" was displayed. This painting has become the icon of surrealism and is often referred to in the popular press as "Limp Watches." (The original was created in 1931 and now hangs in the Museum of Modern Art in New York.) It is quite interesting to note that surrealism (the artistic movement) was concerned with control over time or perhaps more accurately how we are controlled by time. Since the time of the oral presentation we have noticed the painting "Time Transfixed" by René Magritte painted in 1938. This is the painting showing a steam locomotive emerging from a fireplace. Magritte's point was that everything is simultaneous. Apparently, the surrealists anticipated the importance of controlling time as it relates to the display of images.

3 Technology

The schematic of our equipment presented in Figure 1 may help in understanding the following discussion.

We have a DEC VAX 3200 color graphics workstation (labeled VAX in Figure 1), with an optional Peritek frame buffer (labeled VIDEO in Figure 1). The frame buffer generates an RGB (Red, Green, Blue) component video signal with NTSC compatible timing. NTSC composite video is the standard television signal in the United States, Canada, and Japan. Images are loaded into the frame buffer by a program running on the VAX.

Animation (the dynamic graphical display) is created by recording successive video frames, which are replayed at the standard rate of thirty frames per second. Currently, we record the generated images on a laser video disk recorder (LVR). Our model is a Sony LVR-5000A (labeled that in Figure 1). This recorder uses twelve-inch write-once disks. Each side of each video disk holds roughly 43,000 frames which is equivalent to about 23 minutes of dynamic graphics at 30 frames/sec. This laser video disk recorder has several special capabilities. First, because it is a disk and not a tape device, access to any individual frame is essentially instantaneous so recording speed is not limited by the LVR. Second, we are able to control the LVR from the workstation over a standard RS-232 serial line using a simple protocol. Third, it has a built-in encoder/decoder which can convert from component video (R,G,B) to composite video (used for broadcasting and cable transmission) and back again. An additional benefit during playback is the ability to es-



Figure 1: Schematic of the Graphics Equipment

sentially instantaneously jump between arbitrary frames (for comparison), an impossible task with tape. Also we can play back the recorded video at many different speeds both forward and backward (which is impossible with tape).

Before we acquired the LVR, recording was done on a Sony VO-5850 Umatic video tape recorder (VTR) which uses 3/4 in. videotapes (labeled UMATIC in Figure 1). The control signals for editing were sent to the VTR by a Diaquest controller (labeled CTRL in Figure 1) which resides in an IBM personal computer (labeled PC in Figure 1). The PC also contains a video sync generator (labeled SYNC in Figure 1). This signal is used to maintain synchronization of the video signal in the various separate components of the system. We still use the VTR for editing from the disks. We still use the sync generator in the PC. The video equipment is completed by a VHS video cassette recorder (labeled VHS in Figure 1), which is used to transfer animations onto conventional VHS videotapes, and a Trinitron color monitor (labeled TV MONI-TOR in Figure 1) that can display both the RGB component video signal and the composite video signal.

Recently, because the VAX workstation is several years old, we have acquired a much faster DECstation 5000/240 for doing the actual computations. In operation a program running on the DECstation produces an image which is then transmitted to the VAX station (over a private Ethernet) for loading into the frame buffer. When the VAX reports back to the DECstation that the image is loaded, the DECstation instructs the LVR to record the frame. Additionally, there is a 35mm film recorder (labeled SLIDE CAMERA) which allows accurate photographs of the DECstation monitor.

4 Examples

Three specific example applications of the system will now be described in some detail:

- 1. an earth sciences example in which we animate the daily sea-surface temperature;
- 2. an industrial example in which we animate the spatial-temporal distribution of temperature across a metal plate; and
- 3. an epidemiological example in which we animate the reported incidence rate of the disease mumps for the continental 48 states over a twentyone year period.

4.1 Sea-surface Temperatures

From an oceanographer we obtained data which purported to give the sea surface temperature in degrees Kelvin once each day for five years. The temperatures were given on a square 11×11 grid. Thus the total data set contained $11 \times 11 \times 5 \times 365 = 220,825$ observations.

We have made several different animations of these data. The simplest and most natural simply colors a square region around each grid point with



Figure 2: Sea Surface Temperature - Raw Data

a color selected to correspond to the observed temperature. We chose to use an inverse spectral scale so that reds corresponded to high temperatures and blues corresponded to low. We have included a monochrome version of one frame from this simple dynamic graphic in Figure 2. It was apparent even from this very coarse display that there were interesting patterns in the data. After some experimentation, we chose to display the data at the rate of 5 frames per day so there are 6 days per second. This seemed (to us) to give a speed appropriate to the activity in the graphical display.

Because the frame buffer consists of 512×512 pixels any smoothing technique which estimates each pixel for each day separately as a function of the data is going to be very computationally intensive. Exactly what surrealtime graphics is for! Because the temperatures were recorded on a regular grid we decided to use a bilinear interpolant within each grid square and to ignore the data in other squares. Each grid square has roughly 1600 pixels in it. Because this is much smaller than the whole screen, we can precompute the weights associated with each corner of the square for each pixel within the square and store them. If we had chosen to use a method which de-



Temperature on Day 0

Figure 3: Sea Surface Temperature - Smoothed

pended on all the data we would not have had sufficient storage to keep the weights and would have needed to recompute the weights for each frame, a huge computational burden. The result of our bilinear spline interpolant was sufficiently smooth that we did not attempt more complex interpolation or smoothing schemes. A monochrome version of one frame of the result is given in Figure 3. It is striking how smooth the result is, especially when one realizes that the bilinear interpolant that we used makes the fitted function only continuous. The interpolated function has a derivative which is discontinuous on the boundary of every grid square. This apparent over-smoothness led us to examine the pedigree of the data more carefully. We discovered that the word "data" means something different to an oceanographer than it does to a statistician. These "data" turned out to be what is called "Level 3" data in a NASA-defined hierarchy: "geophysical parameters resampled onto a regularly spaced spatial, temporal or space-time grid by some sort of averaging or interpolation. The "data" were already smoothed onto this regular grid and were not data as we think of it, at all.



Figure 4: Schematic of the Metal Plate Showing Locations of Thermocouples and Electrical Conductors

4.2 Temperature Distribution Across a Metal Plate

This example is described in more detail in Eddy and Mockus (1993). A particular industrial process generates product on a continuous basis. The crucial aspect of the process that we are focusing on here concerns the spatial and temporal distribution of temperature over a thin rectangular metal plate. The product flows through a large number (thousands) of small holes in the plate. A diagram of the plate as seen from the direction of product flow is given in Figure 4. The dots in the diagram indicate the locations of thermocouples used for gathering test data and will be described further below. The holes which pass the product are not shown but are distributed fairly uniformly over the plate.

Process engineers believe that it is essential to have the temperature maintained constant (in time) and uniform (in space) over the surface of the plate. The mass flow rate of the product is proportional to temperature (over a narrow range of operating temperatures). Consequently low variability of the temperature (in time) increases uniformity of the product. Furthermore, uniformity (in space) of the temperature improves process efficiency. Of course, there is a temperature gradient over the plate because of the detailed physical structure that is a necessary aspect of the production process and because of boundary (of the plate) conditions. These boundary conditions are not only at the four edges but also the two faces of the plate. Additionally, the completion of a production unit and the start of a new empty unit introduces transient temperature effects on the metal plate. For temperatures outside the operating range, the process simply fails to produce the desired product.

The temperature of the plate is affected by three main factors:

- 1. the flow of product, itself of varying temperature, through holes in the plate;
- 2. the (uncontrolled) directed flow of coolant across the discharge surface of the plate from front to back;
- 3. the (controlled) flow of electric current from right to left within the plate.

To achieve an approximately constant mass flow rate (and hence uniformity of the product) the present control system uses two thermocouples measuring the temperature at points near the front edge of the plate. The two locations are indicated by large dots in Figure 4. The use of two thermocouples is an improvement over the previous use of only one control thermocouple. The reason these locations are on the front of the plate is partly historical (that is where the single thermocouple had been located) and partly for convenience. Also, since the coolant flow is from front to back, changes in that flow will be noticed more rapidly by the control system if the thermocouples are near the front. The left to right locations were chosen for geometric symmetry. The outputs of the two thermocouples are added by a process controller which raises the amperage of an electric current through the plate (and, because of electrical resistance, the temperature) when the measured temperature falls (and lowers the amperage when the temperature rises). In Figure 4 the vertical bars at the right end of the plate indicate where the electric current enters the plate (identified with the letter I); the vertical bars at the left end indicate where it leaves the plate (identified with the letter O). The gap between the two bars at each end was an earlier attempt to make the temperature more uniform over the plate. When a bar extended across the entire end of the plate there was a considerably hotter region through the middle of the plate.

The process engineers believed that the control system was able to maintain the control points within specified limits about the desired temperature. Examination of the temperature measurements from those two thermocouples confirm this.

Process engineers were interested in developing a more detailed understanding of the temporal and spatial distribution of the temperature for several reasons. If the temperature distribution were better understood it should be possible:

- 1. to develop flow control means to direct coolant at the "hot spots" on the plate;
- 2. to change the detailed structure of the plate to compensate for variations in the resistivity (and hence temperature) of the plate; and
- 3. to improve the electric control system to maintain a more uniform temperature.

To gather that information the process engineers designed a special metal plate with 72 thermocouples in addition to the control thermocouples (bringing the total to 74) at various locations on the plate. The particular locations were dictated by the structure of the plate and the conditions of the production process. Subject to these binding constraints, the locations were chosen so that the distance from any point on the plate to the nearest thermocouple was small. The result was, unfortunately, not a regular pattern. Figure 4 uses small dots to indicate the locations on the plate of the 72 additional thermocouples.

Our focus was on learning about the time period during which the temperature of the plate would complete its change in response to a step change in the current (the *time constant* of the plate in engineering jargon). A related interest was in determining the behavior of the control system in response to these changes.

There are five short dark bars on the edges of the plate in Figure 4, two in the back identified with the letters A and B and three in the front identified with the letters C, D, and E. These bars indicate the locations of auxiliary electric current taps which were installed on this special plate for the specific purpose of altering the current flow through the plate. Various test conditions were created by connecting pairs of these taps together. This provided an alternate flow path for the electric current. Approximately 1% of the current would follow one of these alternate paths when available. This small change in the current was sufficient to cause profound changes in the heat distribution over the plate.

Data were gathered automatically during continuous operation of the process under various test conditions from the 74 thermocouples (which included



Figure 5: Heat Distribution on Metal Plate

the two which were used to control the process). The temperature at each thermocouple was determined by averaging the output of a digital voltmeter over 100 powerline cycles (60 cycles per second). The result was stored in two bytes of binary data. The time interval between observations was determined by the overall capabilities of the data gathering equipment and remained constant at approximately 6 seconds throughout. Data were gathered for 300 time intervals. This time period corresponds to the production of a single unit of product. Note that 30 minutes \times 60 seconds/minute \doteq 300 intervals \times 6 seconds/interval. This data gathering process was repeated a number of times under varying conditions resulting in several data sets.

The graphical display we used was built on analysis of the data and discussions with the process engineers. The spatial variation in the data was roughly ten times the temporal variation at any one point in space. The bottom rectangle contains a display of the "normal" temperature distribution when the system is not perturbed. The top rectangle contains a display of the temperature deviations from normal at the particular time. In between there is a time series plot showing the median temperature across the plate as a function of time; the vertical bar through the plot indicates the current time. The color bar at the bottom gives the scale for the two rectangular plots.

4.3 Incidence of the Disease Mumps

The disease mumps is of current public health interest in the United States because of a large outbreak which occurred in 1986-1987, primarily among unvaccinated adolescents and young adults in states without requirements for mumps vaccination. Questions of public health importance concerning mumps include:

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

4.3.1 The Data

The Centers for Disease Control in partnership with the Council of State and Territorial Epidemiologists (CSTE) operates the NNDSS to provide weekly provisional information on the occurrence of diseases that are defined as "notifiable" by CSTE. Further details concerning NNDSS can be found 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 consist of the 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 the entire 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.

We presume there are other reasons. 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 statemonth combinations. Thus there are 1787 missing observations, approximately 15% of those possible.

The raw counts were converted to incidence rates for each state (cases per 100,000 population) 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.

In order to better utilize the color scale we made a nonlinear transformation of the incidence rates before transforming them linearly into colors. Our transformation used the empirical distribution function of the data so that the resulting colors would be approximately uniformly distributed over the color scale. In the image processing literature such a transformation of the pixel intensities is called "equalization." The actual color scale is displayed at the bottom of the frame with the corresponding incidence rates given just above the colors.

We have used the background color to indicate missing data. Thus the states which are missing seem to "disappear" when there is no data. An initial version of the videotape switched instantaneously from a color to the background when there was a missing observation and then back to a color from the background when there was data. The abruptness of this scheme was sufficiently jarring that we modified the scheme to "fade" to the background. This is actually done by linear interpolation between the particular color and the background color.

The entire 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



Figure 6: Raw Incidence Rates in December 1986

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. Consequently, we choose 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. The year and the month are displayed in the lower left corner of the image. Figure 6 displays a single frame of the raw data for the month of December 1986.

It seems natural to assume that the incidence rate for a disease is actually given by a smooth function over the entire United States. The incidence rates computed for each state are the integral of this unknown function with respect to the distribution of population over the respective state. Our problem then is to estimate the smooth function given its integral over the different states. We first considered smoothing the step function given by the raw incidence rates.

The simplest smoothing would be a spatial moving average. There are a

number of serious difficulties with such an approach. The most important of the difficulties concerns the amount of computer time required. Also, it is not clear what value to assign to the integral of the function in those statemonth combinations with missing data values. Another question concerns the size and shape of the smoothing window. For smaller states a large window could effectively wipe out the entire state, while for larger states it would not provide sufficient (visually appealing) smoothing. Varying the size (and shape) of the window in a systematic way related to the areas of the states would be difficult because of the irregular shapes of the states.

An alternative to smoothing is to interpolate a smooth function. In order to interpolate we must know the values of the function at some fixed points. Unfortunately we don't know the values at any points. If we assume that the raw data is observed at, say, the centroid of each state and perform interpolation we can observe the strong influence of a smaller state deep into the territory of a larger neighboring state. Varying the interpolating function to account for the varying areas (and shapes) of the states would be extremely difficult.

To get around this problem we took a sample of points uniformly distributed within each state. 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.

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 weights were chosen to be exponential in the squared distance between the sampled point and the pixel where the function was being interpolated. We also adjusted the weights so that each state has the same influence. The weight for each sampled point in the state is made proportional to the reciprocal of the number of sampled points in the state. It might be also reasonable to adjust the weights according to population in the state but we have not done this.

Estimation using exponential weights for each point can be very time consuming. The frame buffer has more than 500×500 pixels. Assuming an average of 10 points in each state where the value of function is assumed to be given, we have to perform on the order of 10^8 distance and exponential weighting calculations. This requires substantially more than 30 seconds of



Figure 7: Smoothed Incidence Rates in December 1986

CPU time for each frame.

Consequently the weighted estimation was performed only onto a regular grid over the United States. We chose the grid size to be 40 points by 30 points. We then used a bilinear interpolant from the four values at the corners of each of the 39 times 29 rectangles of the regular grid to each pixel within a particular rectangle. The weights for each pair (regular grid point, pixel within grid rectangle) are computed only once and stored.

As with the "raw" data, 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 7.

Mumps in the US is a seasonal disease. The peak occurs in early spring, while 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 at the end of 1960's. In the 1970's these vaccination programs almost completely eradicated the disease, leaving only



Figure 8: Log of the mumps monthly incidence rates versus months from Jan. 1968 to December 1988

a few cases per state per month. Some states ceased mandatory vaccination programs in the early eighties and strong outbreaks of the disease occurred in 1986-1987, primarily among unvaccinated adolescents and young adults in states without requirements for mumps vaccination. This is well supported by the graphs in Figure 8 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 and the smoothed version. 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.

5 Discussion

We have described the reasons why not all dynamic graphics can be interactive and we have provided a description of the system which we use for making surreal-time dynamic graphics. Finally, we have described three specific examples. In each example, the computational burden imposed by the dynamic display of a smooth function of time and space prevented the more usual interactive approach. In the sea-surface temperature example the dynamic display led us to the discovery that the "data" were not really data in the usual sense. In the manufacturing example the dynamic display led the engineers to develop a new control system for the temperature of the plate and helped convince management of the need for the new system. In the epidemiological example we were able to observe the spatial-temporal spread of the disease, particularly during the epidemic in 1986-1987; we know of no other way to observe this spatial-temporal effect.

6 References

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.

Cleveland, W. S. and McGill, M. E. (1988). *Dynamic Graphics for Statistics*, Wadsworth, Inc., Belmont, CA.

Eddy, W. F. and Mockus, A. (1993). An Example of Dynamic Graphics for Manufacturing Process Data, *International Statistical Review* (in press).

Hurley, C. and Buja, A. (1990). Analyzing High-dimensional Data with Motion Graphics, *SIAM J. Sci. Stat. Comput.*, **11**, 1193-1211.

Inselberg, A. (1985). The Plane with Parallel Coordinates, *The Visual Computer*, Special Issue on Computational Geometry, **1**, 69-91.