VISUAL EXPLORATION OF SATELLITE IMAGES

Jürgen Symanzik*, **Louise Griffiths, Robert R. Gillies, Utah State University.**
*Department of Mathematics and Statistics, Logan, UT 84322–3900,
e–mail: symanzik@sunfs.math.usu.edu

Key Words: Atlanta, ArcView/XGobi, Dynamic Statistical Graphics, NDVI, Remote Sensing, Surface Radiant Temperature, Urban Heat Island, Visual Data Mining.

Abstract

In this paper, we focus on visual exploration methods that allow us to explore satellite images that consist of multiple spectral bands that have been remotely sensed by earth observation satellites. The area of interest is the greater Atlanta, GA, region. 18 satellite images from January 1997 through December 1997 form the basis of our analysis. Visual data mining techniques that are available in the ArcView/XGobi link provide us with valuable insights into this spatial/temporal data set.

1. Introduction

As cities increase in size and number around the globe, the effects of the related land cover changes have raised concerns for scientists, planners, policy–makers, and people who live and work in urban areas. As land cover is transformed from porous soils and vegetation to non–porous materials such as asphalt, brick, and stone, the average surface and air temperatures of the area increase. This effect is known as the urban heat island and has been studied extensively, e.g., in Kim (1992) and Lo, Quattrocchi & Luvall (1997).

Remotely sensed data are being widely utilized in the study of the detection of land cover change related to urbanization and the subsequent changes in vegetation levels and surface moisture. Satellite–derived data maps provide scientists, urban planners, utility companies, and others with useful tools to model and monitor the impact of the seasonal energy fluxes in urban areas. Researchers involved in the National Aeronautic and Space Administration's (NASA) Earth Observation Satellite (EOS) Interdisciplinary Science (IDS) project entitled ATLANTA (**AT**lanta **L**and–use **AN**alysis: **T**emperature and **A**ir quality) have been collecting data to study the effects of the rapid growth of the Atlanta, GA, metropolitan area. The primary goal of the study is to improve scientific understanding of the effects through time of the land cover change from vegetated areas to urban structures on the local and regional climate, surface energy fluxes, and air quality. The information gained from this project could help in planning to make urban areas more comfortable, pleasing, and healthful environments for their inhabitants.

The city of Atlanta is located in north central Georgia in the foothills of the Blue Ridge Mountains. It sits at an elevation of 320 meters (1050 feet) and encounters four climatically distinct seasons during the course of a year. Figure 1 (a) shows a highway map and Figure 1 (b) a landcover map of the study region for this paper. The Atlanta region has experienced rapid growth over the past quarter century and is now a thriving transportation, commercial, and financial center for the southeastern United States. The land–atmosphere energy balance of the area has changed as the land cover has been altered over the past several decades from vegetation (primarily forests and agricultural fields) to urban structures, affecting climate, air quality, and the overall environment (Quattrochi & Luvall 1997).

The satellite data explored in this paper were collected from January through December 1997 by the Advanced Very High Resolution Radiometer (AVHRR) sensor aboard a NOAA–14 polar orbiting satellite from the National Oceanic and Atmospheric Administration (NOAA). 18 scenes were chosen based primarily on the clearest (cloud free) views of the Atlanta region and proximity of the sensor viewing angle to nadir. The AVHRR is a cross–track scanning system which collects data in five spectral channels. The swath width is 2700 kilometers (km) and the resolution for each channel is 1.1 km for a nominal altitude of 833 km. The spectral band widths (in micrometers) of the NOAA–14 AVHRR channels are: Band 1 (ρ_1), 0.58–0.68 (Visible red); Band 2 (ρ_2), 0.725–1.0 (Near–infrared); Band 3 (ρ_3), 3.55–3.93 (Thermal–infrared); Band 4 (ρ_4), 10.3–11.3 (Thermal–infrared); and Band 5 (ρ_5), 11.5–12.5 (Thermal–infrared) (Kidwell 1997).

Atmospheric corrections were accomplished using a radiative transfer model (Berk, Bernstein & Robertson 1989). The appropriate sun angles for each date due to seasonal variations were incorpo-

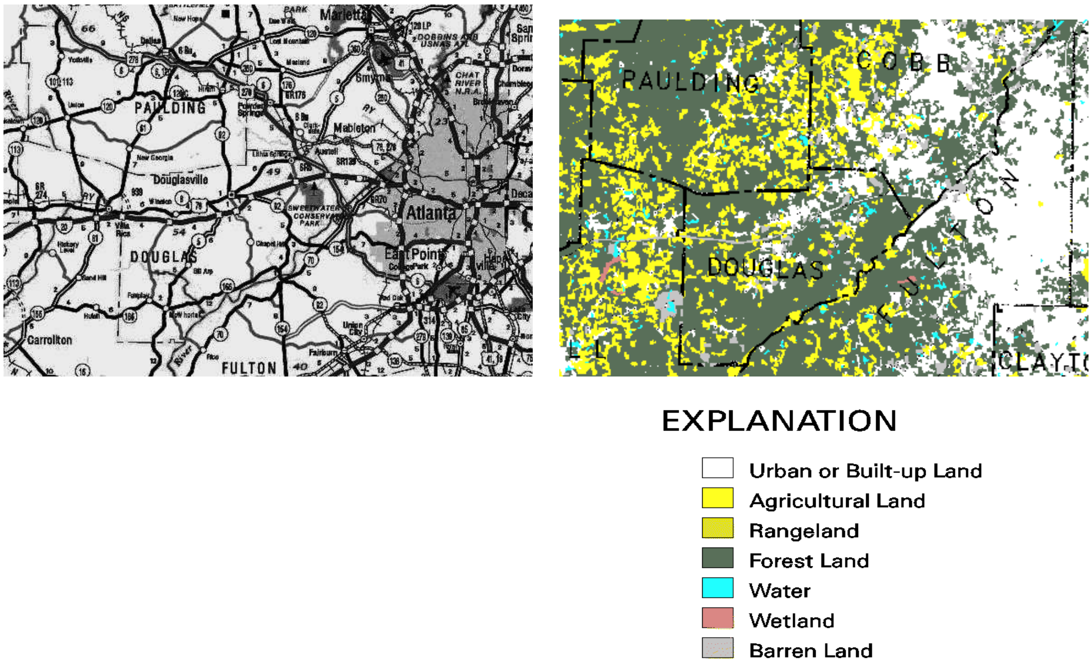
(a)  (b)

Figure 1: Highway map (a) and landcover map (b) of greater Atlanta. While the main urban area stretches in the east, the northwestern corner is densely forested.

rated in the radiance models. Each scene was collected by the sensor at approximately the same time of day (mid afternoon, between 2:00 and 3:00 p.m., local time). Moreover, the normalization and scaling of the satellite data that occurred in the course of this research facilitates the direct comparison of images collected under differing solar and atmospheric conditions (Gillies & Carlson 1995).

After the adjustment process of the raw data, we calculate the Normalized Difference Vegetation Index (NDVI) for each 1.1 km $\times$ 1.1 km pixel as $NDVI = \frac{\rho_2 - \rho_1}{\rho_2 + \rho_1}$, where $-1 \leq NDVI \leq 1$. Highly vegetated surfaces have NDVI values around 0.8 while bare soil surfaces have NDVI values around 0.1. The adjusted band 4 represents the surface radiant temperature (in degrees Kelvin, where $0°C =$ 273.15 Kelvin and $100°C = 373.15$ Kelvin), our second variable of interest.

The main focus of this paper is to demonstrate how visual data mining techniques can be applied to multiple satellite images of the same region. We want to explore structures and changes in the NDVI/surface radiant temperature relationship over the year for different surface types such as forest, sparse vegetation, buildings, etc. Our graphical approach also allows us to find unusual observations, and to handle and reclassify missing data.

In Section 2 of this paper, we give a brief overview on visual data mining techniques (often also called dynamic statistical graphics — DSG) and introduce the software environment we use for our work, i.e., the ArcView/XGobi link. In Section 3, we look at previous efforts that combine DSG and remote sensing. In Section 4, we learn more about the relationship between NDVI and surface radiant temperature. The visual exploration of the greater Atlanta area remote sensing data will be described in Section 5. A conclusion follows in Section 6.

2. Visual Data Mining/Dynamic Statistical Graphics

Visual data mining (often also called dynamic statistical graphics — DSG) uses interactive dynamic graphical methods that allow the user to explore and find unexpected features in (statistical) data sets. Basic DSG techniques such as scatterplot matrices, grand tour, parallel coordinate plots, and linked brushing have been well described in the statistical literature. Visual data mining can be applied whenever we have a complex data set and want to extract as much information as possible. Often, we do not know in advance what to expect from our data.

Fields where visual data mining techniques have been recently applied are telephone calling fraud detection (Cox, Eick, Wills & Brachman 1997), VLSI chip production (Inselberg 1998), archaeological sites (Wilhelm, Wegman & Symanzik 1999), and brain cell simulations (Symanzik, Ascoli, Washington & Krichmar 1999). Obviously, DSG can be very useful in the exploration of satellite images, as discussed in more depth in the next section.

The work in this paper is based on the ArcView/XGobi software link (Symanzik, Cook, Lewin-Koh, Majure & Megretskaia 2000), which is part of the ArcView/XGobi/XploRe environment (Symanzik, Kötter, Schmelzer, Klinke, Cook & Swayne 1998, Symanzik, Klinke, Schmelzer, Cook & Lewin 1997, Symanzik, Cook, Klinke & Lewin 1998). This environment consists of three closely coupled components: the Geographic Information System (GIS) ArcView, the dynamic statistical graphics program XGobi (Swayne, Cook & Buja 1998), and the statistical computing environment XploRe (Härdle, Klinke & Turlach 1995).

3. DSG and Remote Sensing

Various statistical methods have been developed for the classification of remote sensing imagery. Many of these methods are based on principal components and are widely used in the remote sensing community for non–supervised classification purposes. Other statistical approaches that are also well–acknowledged in the remote sensing community have been focused on the optimum band selection for supervised classification. However, most of these solutions ignore the visual capabilities of human beings, highly depend on probably incorrect ground truth, or omit potentially valuable information at an early stage of the analysis.

Exploration of satellite images via DSG is mostly based on human interaction and visual capabilities. This approach helps to detect incorrect ground truth and it does not ignore any of the information that is available. Unfortunately, there exist only experimental software solutions within the statistical community so far that support this approach. Some examples of the exploration of satellite images through DSG follow.

In Buja, McDonald, Michalak & Stuetzle (1991), a Landsat 2 multispectral scanner (MSS) image of the confluence of the Rio Solimões and the Rio Negro near Manaus, Brazil, has been considered. The authors use the Data Viewer (Buja, Hurley & McDonald 1986, Hurley 1988), a program for the graphical data analysis, to explore the four–dimensional spectral data. The authors are able to graphically confirm (through linked brushing and the grand tour) that the measurements on a transect across the confluence (mixing) of the two rivers can be represented as a convex combination of the corresponding two water spectra. An earlier video (McDonald & Willis 1987) provides more details of this analysis.

In Klein & Moreira (1994), the simple image program MTID and XGobi have been linked together. The authors analyze Landsat Thematic Mapper (TM) images of the agricultural Paraná State region. Based on their graphical analysis, a discrepancy between the images and the ground truth becomes immediately obvious. Probably a shift of the recorded ground truth in relation to the images has occurred.

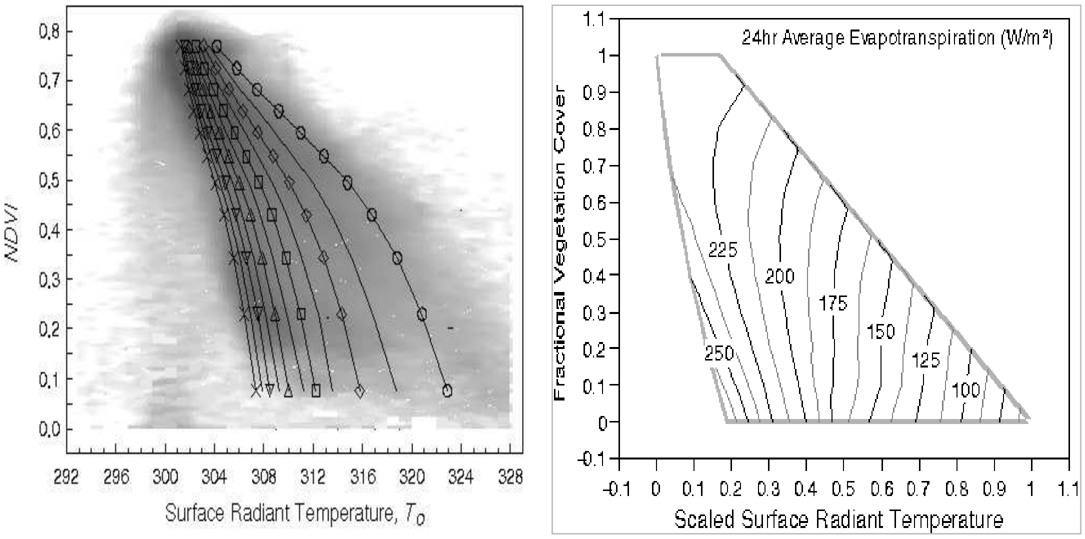
Scott & Jee (1984) and Scott (1986) consider four–channel MSS data from NASA's Landsat IV satellite, taken from an agricultural scene in North Dakota on 5 different days in 1977. They display bivariate and trivariate scatterplots and density contours of the transformed 20–dimensional data and they find clusters that clearly discriminate between sunflowers and small grains. In a second example, Scott & Jee (1984) also work with seven–channel spectral measurements of aspen and spruce forests. After reducing the dimensionality from 7 to 4 by using the major 4 principal components, they create sequences of contour shells (based on the remaining 4 dimensions) for visual inspection.

Salch & Scott (1997) demonstrate a 2–dimensional density grand tour of a Landsat satellite image based on three groups of farm crops. The density grand tour reveals that at least one of the crops contributes to the multimodality of the densities. This is a feature that is also detected in other agricultural settings if one (or more) crop(s) has been planted at different points in time, thus yielding a wide variety of the underlying spectral measurements.

Finally, Carr (1991) presents a hexagon bin scatterplot matrix based on the seven spectral bands of a Landsat image at a Nevada test site. This binned scatterplot matrix shows areas of high and low concentrations of numbers of pairs of the seven spectral bands, almost perfect linear relationships between several pairs of bands, a cutoff of intensity values for the thermal band, and bivariate outliers.

Unfortunately, in the last two examples the statistical plots are not directly linked to the original satellite image. Therefore, no immediate geographic conclusion can be drawn from the graphical statistical exploration of the data.

The ArcView/XGobi/XploRe environment described in the previous section has been used previously to explore remote sensing data (Cook, Majure, Symanzik & Cressie 1996, Symanzik, Majure & Cook 1997, Symanzik, Cook, Klinke & Lewin 1998). However, so far, DSG techniques usually have been applied to only one satellite image of the region of interest. This holds for our previous work with ArcView/XGobi/XploRe as well as for most other DSG approaches towards remote sensing data. Scott



(a) (b)

Figure 2: (a) Airborne–derived scatterplot of NDVI vs T_0 (gray spectral scaling) at a 5 meter spatial resolution for a 7×3 km area of the Mahantango Watershed, Pennsylvania. 18 July 1990, 1145 LST. Isopleths representing moisture availability index, M_0, are overlaid with the legend, $\circ = 0.0$ ("warm" edge), $\diamond = 0.2$, $\square = 0.4$, $\triangle = 0.6$, $\nabla = 0.8$, and $\times = 1.0$ ("cold" edge). (b) Isopleths of 24hr evapotranspiration (W/m^2) for the 18 July 1990 Mahantango Airborne image overlaid on the triangle domain. In this plot, the abscissa is scaled using observed maximum and minimum temperatures (Gillies et al. 1997).

(1986) is one of the few exceptions since he uses remote sensing images from the same area from 5 different days. In this paper, we focus on DSG methods that have been proven useful for the comparison of multiple images rather than on DSG methods useful for the classification of individual pixels.

4. About NDVI and Surface Radiant Temperature

The triangle method is a procedure that combines satellite measurements of surface radiant temperature, T_0, and surface reflectance in the red and the near–infrared. The reflectance measurements are then used to calculate the Normalized Difference Vegetation Index (NDVI) as descibed in Section 1. NDVI is subsequently plotted as a function of T_0 and the observed association between the two variables is illustrated in Figure 2 (a). Biophysical fields (i.e., surface moisture availability (M_0), fractional vegetation cover (Fr), sensible heat flux (H) and evapotranspiration (ET)) are subsequently generated though the coupling of the satellite variables to a Soil Vegetation Atmosphere Transfer Model (SVAT). The SVAT fit for M_0 is overlaid in Figure 2 (a).

The basis of the technique lies in defining the physical limits that constrain the solution for the biophysical fields. It relies theoretically upon observing the complete spectrum of Fr and M_0 at the ground surface. By definition then, the observations must be bounded by their physical limits and will manifest themselves as boundaries — visually characterized in the observations by distinct edges.

Such edges for Fr are straightforward, where a low NDVI extreme and a narrow vertex at the high NDVI extreme represent zero (bare soil) to 100% (complete canopy) vegetation cover respectively. The edges associated with the limits of M_0 are somewhat more difficult to visualize and describe. This is due to, (a) the nature of the spectral plot, as shown in Figure 2 (a), which makes it difficult to isolate extraneous relatively few pixels that mask the edge and, (b) the physics specifying the temperature soil water content (M_0) relationship. Careful observation (with the aid of the model output of M_0 as overlaid) helps delineate (a). For (b), the work of Idso, Schmugge, Jackson & Reginato (1975) who describes the physics, is translated by Goward, Cruickshanks & Hope (1985), Price (1990), and Carlson, Gillies & Perry (1994) who apply the theory to NDVI / T_0 space. The relevant point to come from their interpretation of the physics is that the warmest pixels, despite variations in Fr correspond to the driest surfaces and define an edge — the so–called "warm edge" that represents a minimum limit of M_0. At the other extreme, the opposite applies where a "cold edge" represents a maximum limit of M_0.

The output of the process is a matrix of data that assigns the modeled biophysical variable (M_0) to

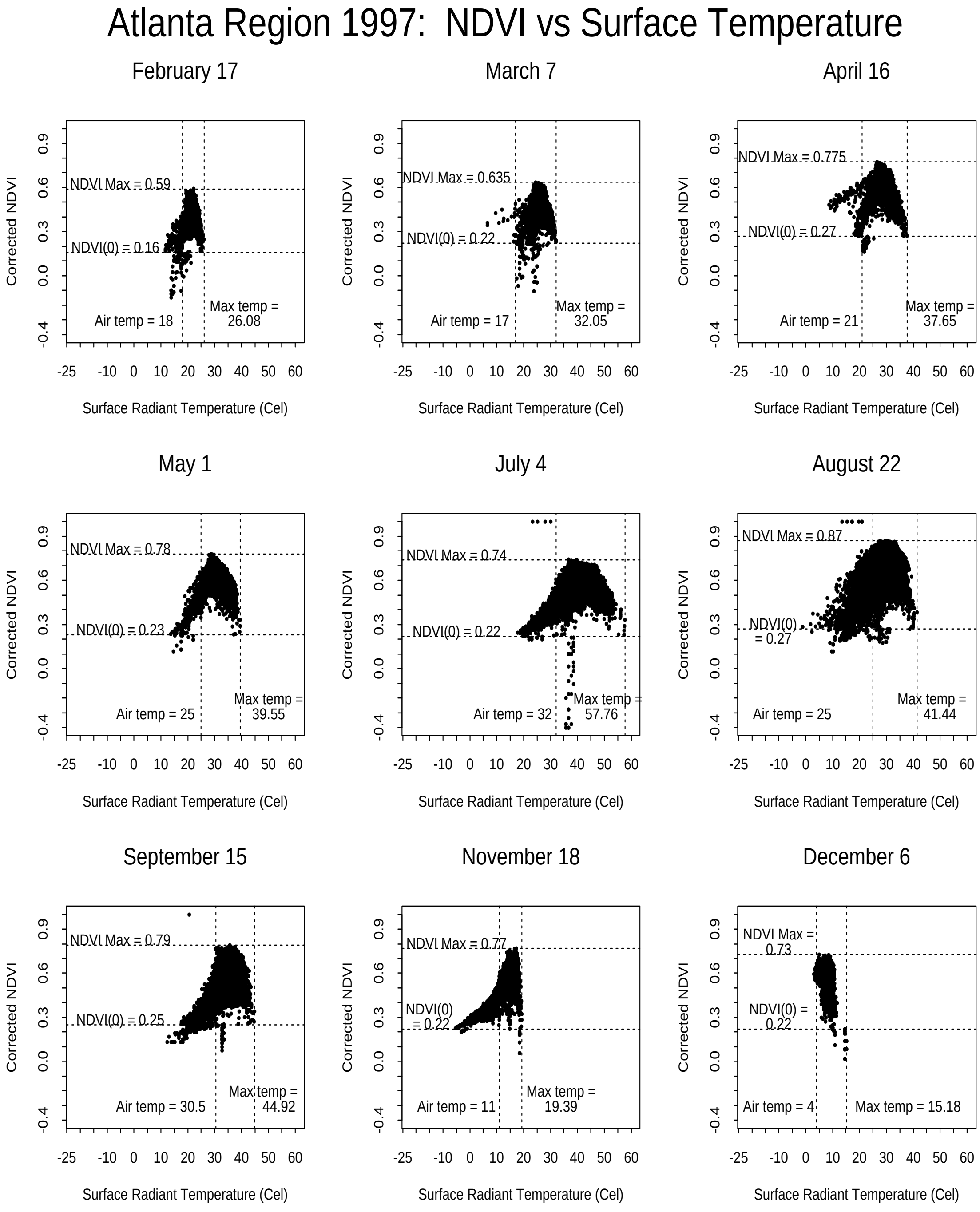


Figure 3: Scatterplots of NDVI vs Surface Radiant Temperature for 9 days in 1997. Different shapes are due to cloud coverage while different scales and ranges can be explained by different climatic conditions over the year. Valid data usually falls into the rectangle enclosed by the pairs of parallel lines in each scatterplot.

the remote measurements (T_0) where the biophysical variable is constrained by the physical limits implied by the remote observations as evidenced by the M_0 overlay in Figure 2 (a). The is represented mathematically by a polynomial of the form:

$$M_0 = \sum_{i=0}^{3}\sum_{j=0}^{3} a_{ij}\hat{T}^i Fr^j$$

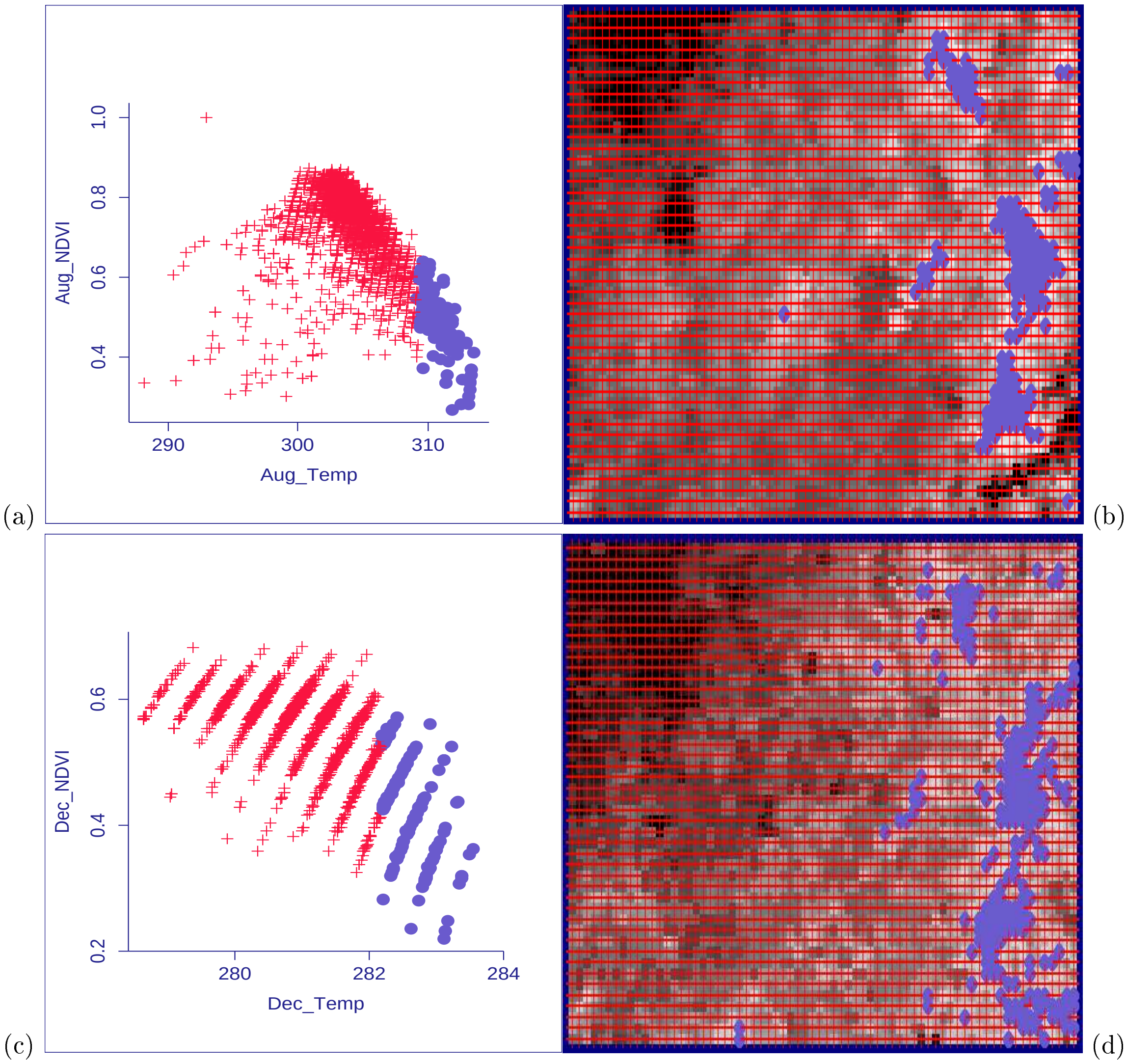
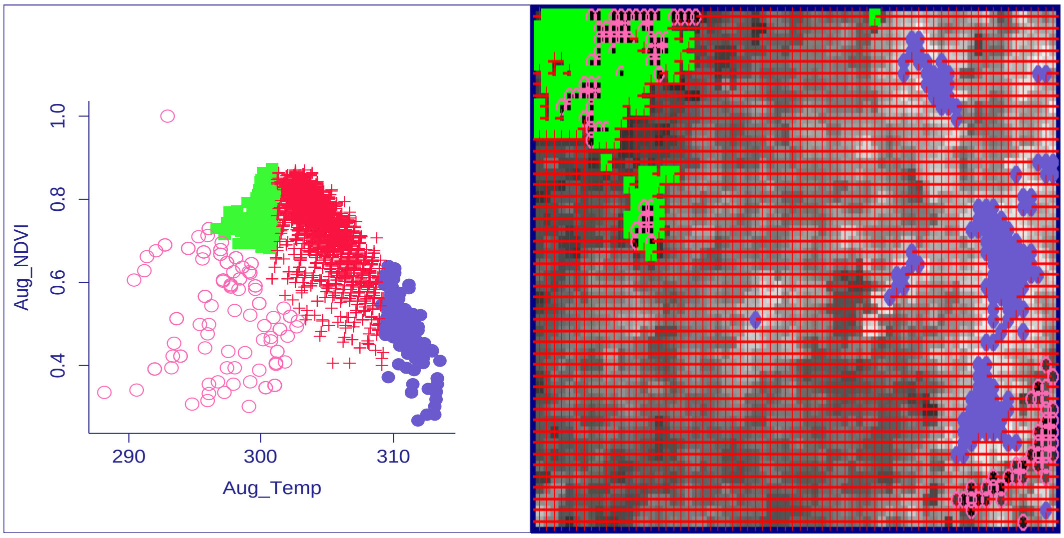


Figure 4: Brushing the high temperature/low to medium NDVI segment with a "•" in XGobi in August (a) and December (c) results in highlighting the corresponding geographical regions, i.e., Atlanta and Marietta, in the ArcView views for August (b) and December (d), respectively. Note that what appears to be a grid in ArcView in fact are "+" symbols representing the unbrushed pixels.

This matrix can be expanded to include model derived sensible heat (H) and evapotranspiration (ET), which are model output variables. For example, $ET = f(Fr, M_0)$ as shown in Figure 2 (b). Each pixel in the image is therefore assigned its analogous biophysical value. A complete description is available in Gillies et al. (1997)

5. Analysis of Atlanta, GA

As desribed earlier, in this paper we will use DSG techniques mostly to compare and explore differences in satellite images taken from the same area, but on different days. Overall, we have data from 18 days from January 1997 through December 1997. The scatterplots in Figure 3 created under S–Plus (Becker, Chambers & Wilks 1988) show NDVI vs Surface Radiant Temperature for 9 of these 18 days. Note that the data displayed in this figure relates to a region that is larger than our region of interest. However, the region studied in this paper is completely enclosed in this larger region. Different shapes of the point clouds and different scales and ranges of both variables are immediately visible. Different shapes can be directly contributed to cloud coverage — the satellite is unable to read an entire pixel or parts of a pixel since the corresponding area is (partially) covered by clouds, resulting in a mixture of ground readings and cloud readings. Different scales and ranges of NDVI and temperature are easily explainable: there is more vegetation

(a) 

(b)

Figure 5: Urban pixels ("•"), forest pixels ("■"), clouds ("∘"), and "other" pixels ("+") in the August NDVI vs temperature scatterplot in XGobi (a) and ArcView (b).

in the early summer than in winter and temperatures are higher (and show more variation) in summer than in winter. The pairs of parallel lines in each scatterplot enclose the part of the data cloud that contains mostly valid readings. Points outside this rectangle are usually missing pixel readings or readings from pixels obscured by clouds. For the remainder of this paper, we will concentrate on two extreme situations: the August 22 data with a large cloud coverage and the December 6 data without any cloud in our region of interest. Figure 4 (a) and (c) shows that the scatterplots for this smaller region resemble those for the larger region.

From Section 4, we know that urban regions have a high surface radiant temperature and low to medium NDVI values. In the August and December scatterplots of NDVI vs temperature, we brushed the high temperature/low to medium NDVI segment in XGobi (Figure 4 (a) and (c)) with a "•". Through linked brushing, the corresponding regions are highlighted in ArcView (Figure 4 (b) and (d)). Though the shapes are slightly different for those two months, it is obvious after comparison with the regional map in Figure 1 (a) that the brushed regions relate to Atlanta and Marietta.

When brushing the low temperature/high NDVI segment with a "■" in the August XGobi scatterplot (Figure 5 (a)), we see in the upper left corner of the ArcView view (Figure 5 (b)) that this relates to the forested region as expected. Brushing the pixels that are outside the main point cloud in XGobi with a "∘" (also Figure 5 (a)), we notice that these relate to pixels in the lower right corner and to pixels near the forested region in the upper left corner of the ArcView view (also Figure 5 (b)). From our theoretical considerations we know that these pixels outside the main point cloud are clouds, which is confirmed by the dark appearance of these pixels in the ArcView view.

In Figure 6 (a), we use the brushing information from the August NDVI vs temperature scatterplot, but now look at the December NDVI vs temperature scatterplot in XGobi. While pixels representing urban regions ("•") and forested regions ("■") tend to appear in the lower right corner and upper left corner of the scatterplot respectively, we also notice some cluttering of the boundaries of these regions. The cloud pixels ("∘") appear everywhere in this scatterplot. Using our theoretical knowledge that urban regions have high temperatures and low to medium NDVI, while forested regions have low temperatures and high NDVI, we reclassify all pixels accordingly based on the December values (Figure 6 (b)). Figure 6 (c) shows in the ArcView view how clouds in different regions get reclassified. Figure 6 (d) shows the August NDVI vs temperature scatterplot, using the brushing information from December. In fact, clouds have been reclassified as urban ("•"), forested ("■"), and "other" ("+"). Note that the urban pixels from December are mixed with "other" pixels in the August NDVI vs temperature scatterplot.

In Figure 7 (a) we brushed the two pixels with the lowest NDVI and near to the highest temperature with a dark "■" in the August NDVI vs temperature scatterplot in XGobi, expecting that these pixels would relate to a highly urbanized region such as downtown Atlanta. Naturally, one would expect these pixels to appear in the lower right corner of

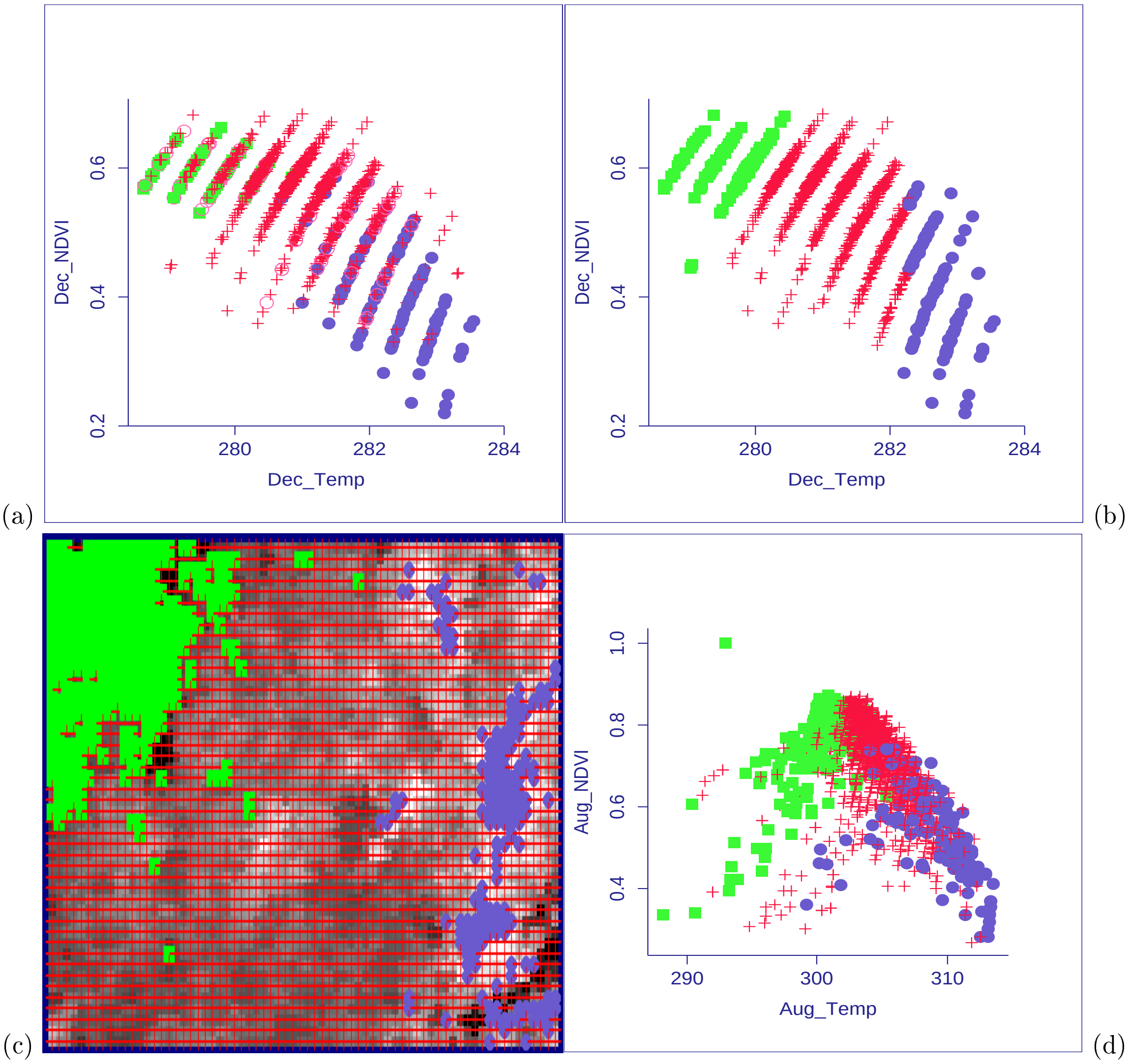
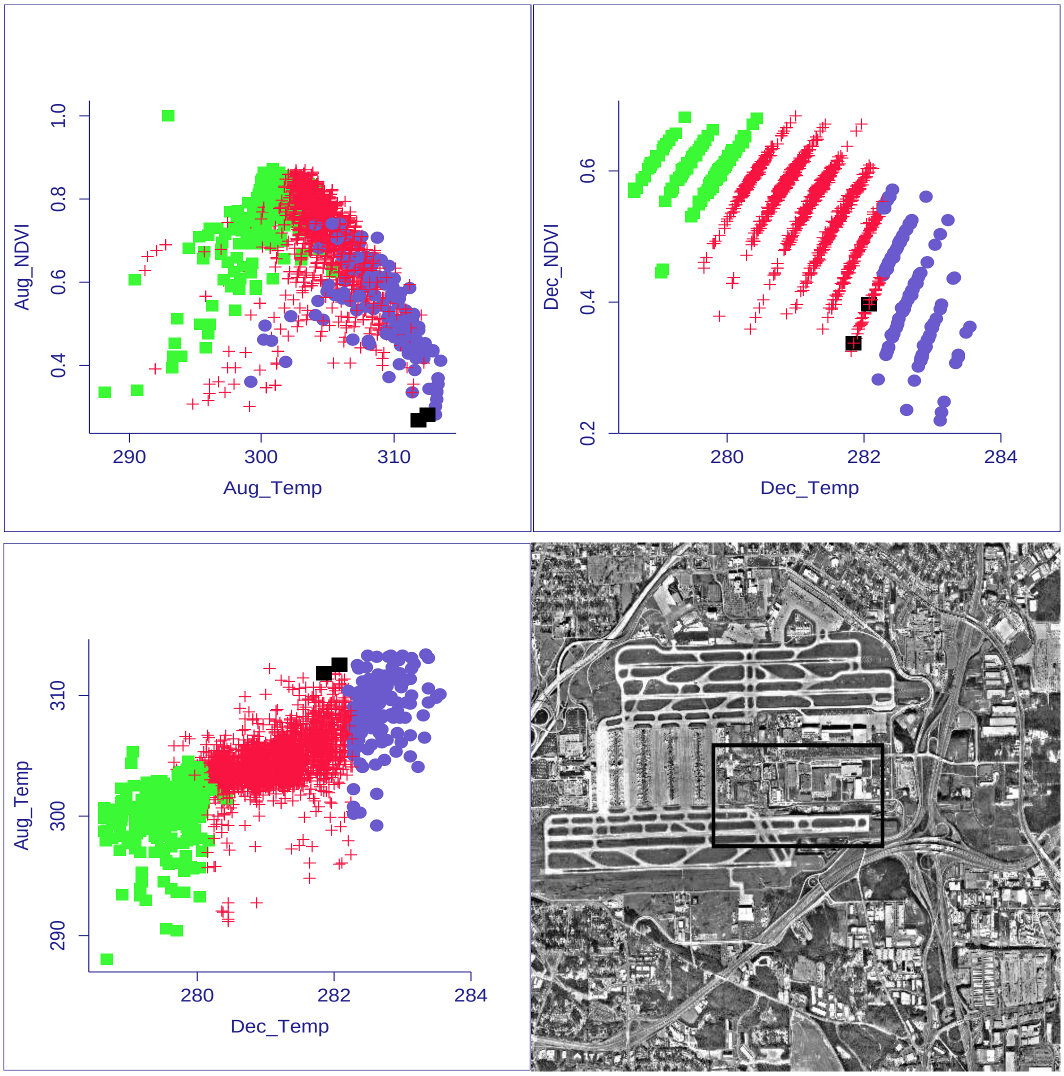


Figure 6: Strating with the brushing information from August, now displayed in the December NDVI vs temperature scatterplot in XGobi (a), we reclassify pixels according to the December values (b). The result can be seen in the ArcView map view of August (c) and the effect on clouds is best visible in the August NDVI vs temperature scatterplot in XGobi (d).

any NDVI vs temperature scatterplot over the year. But, surprisingly, in the December NDVI vs temperature scatterplot (Figure 7 (b)), these pixels have neither the lowest NDVI nor fall in the high temperature range classified as urban in Figure 6 (b). A scatterplot of August vs December temperatures (Figure 7 (c)) shows the overall high correlation between temperatures for these two months as well as the somewhat unusual behavior of these two brushed pixels.

Finally, a look at the map view and background information on remote sensing technology solves this mystery. First of all, the two brushed pixels do not relate to downtown Atlanta but to the Hartsfield Atlanta International Airport a few miles south of downtown. When zooming into the airport region, we see in Figure 7 (d) that the two brushed pixels should mostly relate to buildings on the airport. However, the remote sensing technology and following adjustments that have been used to obtain the 18 images from the Atlanta region is not very precise. We have to consider the possibility of a pixel shift up to one pixel in each direction. Thus, the pixels that are supposed to relate to buildings could be shifted to the air field with intermediate ground vegetation on one day and the interstate intersection on another day. Certainly, we would get different readings for these pixels.

An additional explanation is the precision in the temperature measurements. The obtained surface



(a) (b) (c) (d)

Figure 7: Two pixels with lowest NDVI and close to highest temperature (brushed with a dark "■") in the August scatterplot in XGobi (a) reappear at unexpected locations in the December NDVI vs temperature scatterplot (b) and in a scatterplot of August vs December temperatures (c). A closer look at the map view (d) reveals that these pixels theoretically should relate to buildings on the Hartsfield Atlanta International Airport. However, due to measurement inprecision, these pixels could have been easily shifted up to one pixel in either direction during each measurement, thus covering a completely different surface each time.

radiant temperature can shift by about ± 1 Kelvin. Thus, the 282 Kelvin reading for the airport pixels in December basically relates to a $(281, 283)$ Kelvin confidence interval for the true temperature.

Note that the December temperature range is only about 5 Kelvin while the August temperature range is more than 20 Kelvin. A possible inprecision in the obtained temperature also explains why our reclassification of urban areas in Figure 6 based on the December temperatures results in a mixture of urban pixels with "other" pixels in the August scatterplot. December temperatures are simply not accurate enough to serve as a good predictor for other months.

6. Conclusion

Even though we only concentrated on two different days for the same region within this paper, it should be obvious how to use visual data mining techniques for the full data set, i.e., all 18 days. Linked brushing in multiple views, combined with zooming into the underlying geography is a powerful approach to identify common structures in the data, reclassify missing pixels, or find pixels and regions that behave unusually over a particular time span.

It should be noted that the grand tour was not very helpful for this data set since clouds cover different parts of the region on different days, thus randomly distort individual pixels. While theory tells us where to expect pixels related to clouds in a bivariate scatterplot of NDVI vs temperature, there is no such generalization to projections in the grand tour. However, the grand tour might be very valuable when comparing cloud–free days only.

Overall, visual data mining techniques available in software packages such as the ArcView/XGobi link provide us with valuable insights into complex spatial/temporal data sets such as the 18 satellite images from the greater Atlanta region.

Acknowledgements

Figure 1 (b) originates from Musser, Jonathan W., Digital Atlas of State of Georgia, Level–I Landuse, US Geological Survey, Digital Data Series, Atlanta, GA 1996. Produced by the US Geological Survey in cooperation with the Georgia Department of Natural Resources, Geologic Survey. Data and metadata available from US Geological Survey, Georgia District, Center for Spatial Analysis Technologies, Atlanta, GA 30332. Internet Address: `http://csat.gatech.edu/csat.html`

References

Becker, R., Chambers, J. & Wilks, A. (1988), *The New S Language — A Programming Environment for Data Analysis and Graphics*, Wadsworth and Brooks/Cole, Pacific Grove, CA.

Berk, A., Bernstein, L. S. & Robertson, B. D. (1989), MODTRAN: A Moderate Resolution Model for LOWTRAN 7, Report GL–TR–89–0122, Geophysics Laboratory, Bedford, MA.

Buja, A., Hurley, C. & McDonald, J. A. (1986), A Data Viewer for Multivariate Data, *in* T. J. Boardman & I. M. Stefanski, eds, 'Proceedings of the 18th Symposium on the Interface between Computer Science and Statistics, Fort Collins, CO', American Statistical Association, Washington, D.C., pp. 171–174.

Buja, A., McDonald, J. A., Michalak, J. & Stuetzle, W. (1991), Interactive Data Visualization Using Focusing and Linking, *in* G. M. Nielson & L. Rosenblum, eds, 'Proceedings of Visualization '91, Los Alamitos, CA', IEEE Computer Society Press, pp. 156–163.

Carlson, T. N., Gillies, R. R. & Perry, E. M. (1994), 'A Method to Make Use of Thermal Infrared Temperature and NDVI Measurements to Infer Surface Soil Water Content and Fractional Vegetation Cover', *Remote Sensing Reviews* **9**, 161–173.

Carr, D. B. (1991), Looking at Large Data Sets Using Binned Data Plots, *in* A. Buja & P. T. Tukey, eds, 'Computing and Graphics in Statistics', Springer, New York, NY, pp. 7–39.

Cook, D., Majure, J. J., Symanzik, J. & Cressie, N. (1996), 'Dynamic Graphics in a GIS: Exploring and Analyzing Multivariate Spatial Data Using Linked Software', *Computational Statistics: Special Issue on Computeraided Analysis of Spatial Data* **11**(4), 467–480.

Cox, K. C., Eick, S. G., Wills, G. J. & Brachman, R. J. (1997), 'Visual Data Mining: Recognizing Telephone Calling Fraud', *Data Mining and Knowledge Discovery* **1**, 225–231.

Gillies, R. R. & Carlson, T. N. (1995), 'Thermal Remote Sensing of Surface Soil Water Content with Partial Vegetation Cover for Incorporation into Climatic Models', *Journal of Applied Meteorology* **34**, 745–756.

Gillies, R. R., Cui, J., Carlson, T. N., Kustas, W. P. & Humes, K. S. (1997), 'Verification of a Method for Obtaining Surface Energy Fluxes of Surface Soil Moisture Availability from Remote Measurements of NDVI and Surface Radiometric Temperature', *International Journal of Remote Sensing* **18**(15), 3145–3166.

Goward, S., Cruickshanks, G. D. & Hope, A. (1985), 'Observed Relation between Thermal Emission and Reflected Spectral Radiance of a Complex Vegetated Landscape', *Remote Sensing of Environment* **18**, 137–146.

Härdle, W., Klinke, S. & Turlach, B. A. (1995), *XploRe: An Interactive Statistical Computing Environment*, Springer, New York, NY.

Hurley, C. (1988), A Demonstration of the Data Viewer, *in* E. J. Wegman, D. T. Gantz & J. J. Miller, eds, 'Proceedings of the 20th Symposium on the Interface between Computing Science and Statistics, Fairfax, VA', American Statistical Association, Alexandria, VA, pp. 108–114.

Idso, S. B., Schmugge, T. J., Jackson, R. D. & Reginato, R. J. (1975), 'The Utility of Surface Temperature Measurements for the Remote Sensing of Surface Soil Water Status', *Journal of Geophysical Research* **80**(21), 3044–3049.

Inselberg, A. (1998), 'Visual Data Mining with Parallel Coordinates', *Computational Statistics* **13**(1), 47–63.

Kidwell, K. B., ed. (1997), *NOAA POD Guide (NOAA Polar Orbiter Data User's Guide*, National Oceanic and Atmospheric Administration, U.S. Department of Commerce.

Kim, H. H. (1992), 'Urban Heat Island', *International Journal of Remote Sensing* **13**, 2319–2336.

Klein, R. & Moreira, R. I. (1994), Exploratory Analysis of Agricultural Images via Dynamic Graphics, Technical Report 9/94, Laboratório Nacional de Computação Científica, Rio de Janeiro, Brazil.

Lo, C. P., Quattrocchi, D. A. & Luvall, J. C. (1997), 'Application of High–Resolution Thermal Infrared Remote Sensing and GIS to Assess the Urban Heat Island Effect', *International Journal of Remote Sensing* **18**, 287–304.

McDonald, J. A. & Willis, S. (1987), 'Use of the Grand Tour in Remote Sensing', ASA Statistical Graphics Video Lending Library (contact: dfs@research.att.com).

Price, J. C. (1990), 'Using Spatial Context in Satellite Data to Infer Regional Scale Evapotranspiration', *IEEE Transactions on Geoscience and Remote Sensing* **GE–28**, 940–948.

Quattrochi, D. A. & Luvall, J. C. (1997), 'High Spatial Resolution Airborne Multispectral Thermal Infrared Data to Support Analysis and Modeling Tasks in EOS IDS Project ATLANTA', `http://wwwghcc.msfc.nasa.gov/atlanta/`.

Salch, J. D. & Scott, D. W. (1997), 'Data Exploration with the Density Grand Tour', *Statistical Computing and Statistical Graphics Newsletter* **8**(1), 7–11.

Scott, D. W. (1986), Data Analysis in Three and Four Dimensions with Nonparametric Density Estimation, *in* E. J. Wegman & D. J. DePriest, eds, 'Statistical Image Processing and Graphics', Marcel Dekker, New York, NY, pp. 291–305.

Scott, D. W. & Jee, R. (1984), Nonparametric Analysis of Minnesota Spruce and Aspen Tree Data and Landsat Data, *in* 'Proceedings of the Second Annual Symposium on Mathematical Pattern Recognition and Image Analysis', NASA Contractor Report, Houston, TX, pp. 27–49.

Swayne, D. F., Cook, D. & Buja, A. (1998), 'XGobi: Interactive Dynamic Graphics in the X Window System', *Journal of Computational and Graphical Statistics* **7**(1), 113–130.

Symanzik, J., Ascoli, G. A., Washington, S. S. & Krichmar, J. L. (1999), 'Visual Data Mining of Brain Cells', *Computing Science and Statistics* **31**, 445–449.

Symanzik, J., Cook, D., Klinke, S. & Lewin, N. (1998), Exploration of Satellite Images in the Dynamically Linked ArcView/XGobi/XploRe Environment, *in* B. A. Bodt, ed., 'Proceedings of the Third Annual U.S. Army Conference on Applied Statistics, 22–24 October 1997', Army Research Laboratory ARL–SR–74, Aberdeen Proving Ground, MD, pp. 23–33.

Symanzik, J., Cook, D., Lewin-Koh, N., Majure, J. J. & Megretskaia, I. (2000), 'Linking ArcView and XGobi: Insight Behind the Front End', *Journal of Computational and Graphical Statistics* **9**(3), 470–490.

Symanzik, J., Klinke, S., Schmelzer, S., Cook, D. & Lewin, N. (1997), The ArcView/XGobi/XploRe Environment: Technical Details and Applications for Spatial Data Analysis, *in* '1997 Proceedings of the Section on Statistical Graphics', American Statistical Association, Alexandria, VA, pp. 73–78.

Symanzik, J., Kötter, T., Schmelzer, S., Klinke, S., Cook, D. & Swayne, D. (1998), 'Spatial Data Analysis in the Dynamically Linked ArcView/XGobi/XploRe Environment', *Computing Science and Statistics* **29**(1), 561–569.

Symanzik, J., Majure, J. J. & Cook, D. (1997), 'Dynamic Graphics in a GIS: A Bidirectional Link between ArcView 2.1 and XGobi — An Update', *Computing Science and Statistics* **29**(2), 35–40.

Wilhelm, A. F. X., Wegman, E. J. & Symanzik, J. (1999), 'Visual Clustering and Classification: The Oronsay Particle Size Data Set Revisited', *Computational Statistics: Special Issue on Interactive Graphical Data Analysis* **14**(2), 109–146.