Statistical Analysis of Remote Sensing Data using Systat

Introduction

Since the beginning of the space age, remarkable progress has been made in utilizing remote sensing data to describe, study, monitor and model the earth's surface and interior. Improvements in sensor technology, especially in the spatial, spectral, radiometric and temporal resolution, have enabled the scientific community to operationalize the methodology. The trend of development of remote sensing is being from panchromatic, multi-spectral, hyper-spectral to ultra-spectral with the increase in spectral resolution. On the other hand, spatial resolution is reaching its highest side of one meter resolution. The operational remote sensing satellites LANDSAT, Indian Remote Sensing Satellites (IRS series of satellites), IKONOS etc. are providing earth data in different improved spatial and spectral resolutions.

Data Types

Usually, remote sensing is the acquisition of analog (photo) and digital (image) data from platforms ranging from hand-held devices to space-borne systems. That is, remote sensing data are satellite images (typically LANDSAT or other visible imagery), which are used in the geosciences for a multitude of applications, including land-use, urbanization, sea-ice expansion, to name only a few.

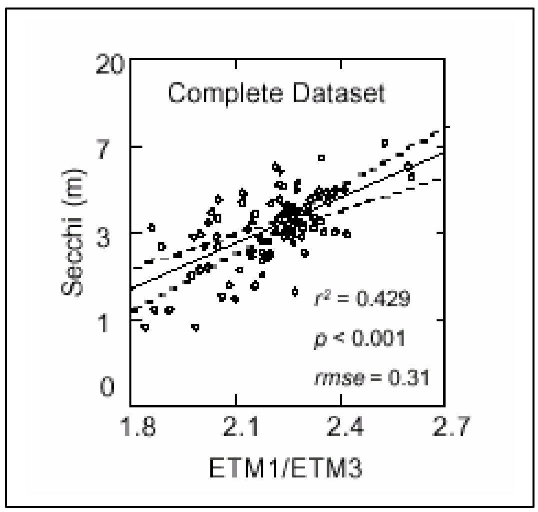
Or remote sensing data is any data collected by an instrument that uses a transmission/receiving technology, such as acoustic, seismic, radio, or other electromagnetic waves. Typical examples are acoustics for the survey of bathymetry, seismics for observation of stratigraphic layers, observations of land surfaces from space (radar waves). The variable or location sampled using remote sensing is commonly difficult to access, or is more effectively sampled in large quantities with this technology.

A wealth of technologies has been developed for data collection in oceanography and marine geophysics, where direct observation is the exception (for example, seafloor drilling, biologic sampling using submarines). Another set of remote-sensing data consists of technologies applied in terrestrial geophysics, for example in resource exploration (seismics, electromagnetics, radar imaging).

But the real guns come from the ability of statistical analysis through **Systat** to understand the data that provides major benefits to scientists studying and understanding human impacts on the global environment, managing the earth's natural resources, and planning and conducting many other activities of scientific and social importance.

Applications

Remote sensing has been effectively used to measure water clarity in several single-lake studies. However, there are still many challenges to applying these approaches to



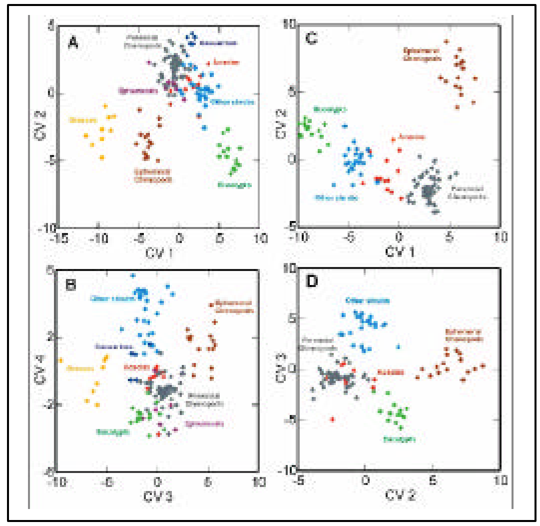
large numbers of lakes within a large geographic region. Nelson et al. (2002) set objectives to: (1) develop a model to predict lake water clarity from Landsat data based on 93 calibration lakes across the state of Michigan, and (2) examine how the distribution of water clarity across the 93 lakes, as measured by Secchi disk transparency (SDT), influenced the model. They hypothesized that the distribution of

SDT in the calibration dataset would influence model calibration, and that it is necessary to include a complete range of calibration SDT values to achieve a better regional model. Their **regression** model of field-collected SDT data and Landsat-7 ETM+ data for the 93 lakes resulted in an r^2 of 0.43 (p < 0.001), which was substantially lower than many previous single-lake studies. Also they simulated a calibration dataset with a different SDT distribution similar to previous studies to examine the role of lake SDT distribution in the model fit. The percent of lakes that had a SDT value less than 1.5 m was increased from 8.6% (original dataset) to 47% (subsampled dataset).

The **regression** model for the subsampled dataset resulted in an r^2 of 0.82 (p < 0.001). Their results show that Landsat can be used to measure water clarity across a large number of lakes with a wide range of SDT values. However, they found the **regression** model to be sensitive to the distribution of SDT values used in the calibrated dataset, and conclude this distribution must be taken into account when developing regional models to predict lake water clarity using remote sensing.

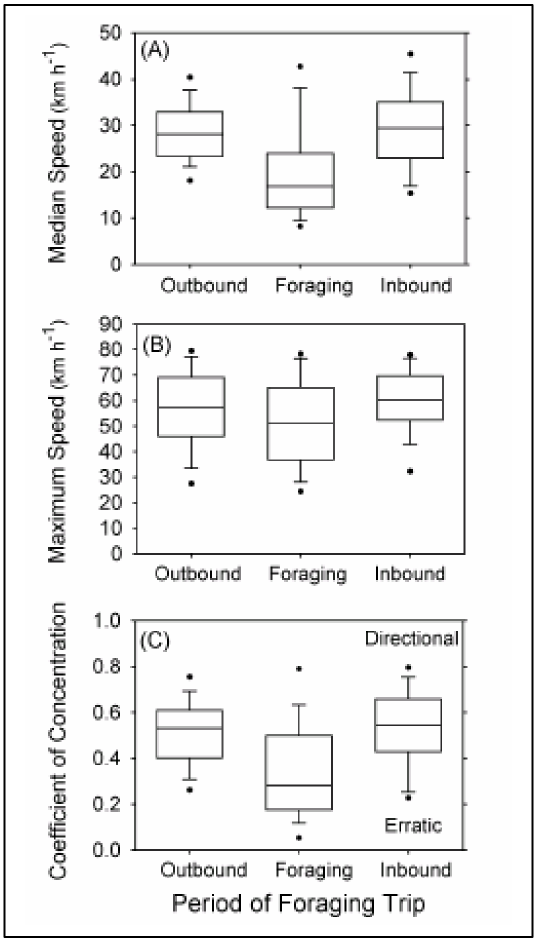
All the analysis was carried out using **Systat**.

Visible – near-infrared reflectance spectra were measured for a selection of common trees, shrubs, ephemerals, and grasses in the rangelands of the southeastern Australian arid zone (Lewis (2002)). The plants varied in leaf structure, leaf coatings, woodiness, succulence, chlorophyll and other pigment content, and physiological status, and the spectra showed considerable variation within and between species. Factors contributing to their spectral variation, identified by principal components analysis, were albedo over the entire spectral range, followed by contrast between near-infrared and visible reflectance, and blue and red contrast. Visible – near-infrared regions most important in variation within the spectral collection were the red edge, visible green, near infrared beyond 760 nm, blue, and red. Between 1000 and 1800 nm variation could be explained by albedo and the water absorption region near 1420–1450 nm. Functional and taxonomic groups of plants were classified using discriminant analysis of the narrow-band visible – near-infrared spectra. Spectral regions most influential in the discrimination were the chlorophyll absorption near 680 nm, the infrared beyond 720 nm, and the blue–green and other visible wavelengths. Fair discrimination (k = 0.48) was achieved with 70 wavebands spanning the visible – near infrared but was slightly reduced with 16 selected narrow bands (k = 0.44). Better discrimination was achieved with five rather than eight plant groups (k = 0.53).



Both **principal components analysis** and **discriminant analysis** using **Systat** pointed to broad spectral regions that were important in the arid plant spectral variation, rather than specific narrow bands.

Hyrenbach et al. (2002) characterized the movements and oceanographic habitats of black-footed *(Phoebastria nigripes*) and Laysan *(P. immutabilis*) albatrosses during the brooding and the rearing periods of the breeding cycle. Analyses of satellite telemetry data in conjunction with remotely sensed sea surface temperature and chlorophyll concentrations revealed substantial differences in habitat use between these 2 sympatrically breeding species. During the brooding period, black-footed albatross restricted their foraging to tropical waters (>20°C), while Laysan albatross ventured into the colder waters of the Transition Domain (15 to 12°C) and the Subarctic Frontal Zone (12 to 10°C). This



pelagic segregation became more apparent with the expansion of the foraging ranges later in the breeding season. During the chick-rearing period, black-footed albatross commuted to the California Current (15 to 12°C) and Laysan albatross foraged in subarctic (<12°C) and Transition Domain (15 to 12°C) waters. The foraging behavior of albatrosses was scale-dependent. Over macro-mega scales of (1000 to 3000 km) albatross dispersion was influenced by large-scale ocean productivity patterns and water mass distributions. Over smaller coarse-meso scales of (10 to 100 km) albatrosses focused their foraging activities along oceanic habitats characterized by elevated ocean productivity and prey aggregation. The foraging birds traveled more slowly in the vicinity of highly productive continental shelves (central California to Washington State, Aleutian Islands), and hydrographic fronts (Transition Domain, North Pacific Transition Zone Chlorophyll Front). Conversely, the satellite tracked albatrosses commuted rapidly over tropical and subtropical waters between these foraging areas and the breeding colony. These results highlight the significance of macro-mega scale of (1000 to 3000 km) water mass distributions and coarse-meso scale (10 to 100 km) hydrographic features to far-ranging marine predators, and underscore the need to understand how physical-biological processes sustain predictable regions of elevated ocean productivity and prey aggregation in marine systems.

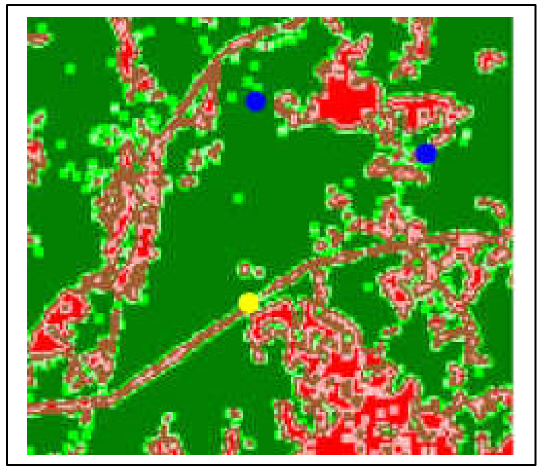
Statistical analysis was carried out using **repeated measures ANOVA through GLM** and **ANCOVA** with **Systat**.

Conclusions

The description above just gave a bird's eye view of **Systat's** capabilities. But **Systat** provides a powerful statistical and graphical analysis system in a graphical environment using descriptive menus and simple dialog boxes. **Systat's** command language provides functionality not available in the dialog box interface in addition to complete coverage of menu-based functionality. Robust algorithms from leading statisticians give meaningful results even with extreme data. Create missing value estimates using regression-based point estimation or an EM algorithm. Obtain complete distributions and standard errors using **Systat's** bootstrapping capability implemented globally across 21 statistical procedures even when normality assumptions are violated and no model is available. Matrix procedure allows you to use matrix algebra to specify statistical analyses and perform data management tasks.

Systat offers more scientific and technical graphing options than any other desktop statistics package. Compare subgroups, overlay charts, and transform coordinates, change colors, symbols and more to create insightful presentations. Speed up your analysis by rotating your 3-D graphs to visually determine the perfect power or log transformation to normalize your data using the Dynamic Explorer to speed up your analysis.

Create compelling reports by combining formatted statistical output with publication-quality graphs in **Systat's** rich text output window.



References (in order of appearance)

Stacy A. C. Nelson, Patricia A. Soranno, Kendra Spence Cheruvelil, Sam A. Batzli and David L. Skole (2002). Assessing regional lake water clarity using Landsat and the role of inter-lake variability. *To be submitted to the journal: Remote Sensing of Environment* (as mentioned in the paper).

Megan Lewis (2002). Spectral characterization of Australian arid zone plants. *Can. J. Remote Sensin*g, Vol. 28, No. 2, pp. 219–230.

K. David Hyrenbach, Patricia Fernández and David J. Anderson (2002). Oceanographic habitats of two sympatric North Pacific albatrosses during the breeding season. MARINE ECOLOGY PROGRESS SERIES, Vol. 233, pp. 283–301.

Appendix - Spatial statistics using Systat

Spatial statistics involve a variety of methods for analyzing spatially distributed data. **Systat** Spatial Statistics covers two principal areas: fixed-point methods (kriging and Gaussian simulation) and random-point methods (nearest-neighbor distances, polygon area/volumes, quadrat procedures).

Spatial statistics compute a variety of statistics on a 2-D or 3-D spatially oriented data set. Variograms assist in the identification of spatial models. Kriging offers 2-D or 3-D kriging methods for spatial prediction. Simulation realizes a spatial model using Monte Carlo methods. Finally, a variety of point-based statistics are produced, including areas (volumes) or Voronoi polygons, nearest-neighbor distances, counts of polygon facets, and quadrat counts. Graphs are automatically plotted and summary statistics are printed for many of these statistics.