MATH 498/598 Spatial Statistics
Final Project Poster Session

Tuesday, May 10th 3:15-5:15 pm
GRL Conference Room
(across from Geology Museum)

Titles, abstracts, and authors of each project are listed in this booklet, sorted by student classification (undergraduate vs. graduate) and by major.

The project number assigned to each project corresponds to the number on the poster score sheet and also to the order in which posters will be displayed around the room.

Contents

1 Undergraduate Math/Computer Science 2
2 Graduate Math/Computer Science 3
3 Graduate Mining & Earth Sys Engineering 6
4 Graduate Hydrology 6
5 Graduate Geophysics 8
1 Undergraduate Math/Computer Science

PROJECT 1

Author(s): Angela Bauer, Regeena Martinez, and Andrea Mccarty-Daniels

Title: Inhomogeneous Point Process Childhood Leukemia and Lymphoma

Abstract: This project analyzes marked spatial point pattern data for leukemia and lymphoma versus control cases in North Humberside, England. Our analysis will use this dataset to answer the question: how much of the clustering in the data can be attributed to a geographical pattern of cases of leukemia and lymphoma and how much is simply due to clustering in residences? We will use the control cases along with constant risk hypothesis analysis to answer this question. Our analysis leads us to fail to reject the null hypothesis of constant risk and conclude that there is not significant evidence that the cases of childhood leukemia and lymphoma are more clustered than we would expect with constant risk. Additionally, through the Cuzick and Edwards NN test, we find that there is only significant clustering on the very local scale, indicating that there is not regional clustering. We therefore conclude that when considering North Humberside as a whole, there is no strong clustering among locations of cases of childhood leukemia and lymphoma. Clustering is thus most likely due to population areas such as cities and not due to geographical characteristics.

PROJECT 2

Author(s): Jake Rezac and Sarah Valovcin

Title: Earthquake Modeling: A Groundbreaking Statistical Analysis

Abstract: We analyze a dataset of earthquake aftershocks caused by the 6.9 magnitude Loma Prieta earthquake which struck California in 1989. Two approaches are taken to analyze the data: first, the aftershocks are considered as an unmarked point process. Various inhomogeneous Poisson processes are fit to the data and we validate the models with goodness-of-fit and residual analyses. Second, the aftershocks are considered as marked point processes with the mark being a time variable. Using Mantel’s test for clustering, we examine correlation between aftershock location and time. Finally, we briefly consider using these two approaches to generate a risk assessment for future earthquakes in the San Francisco Bay area.
PROJECT 3

Author(s): Jamie Johnson

Title: Analysis of Earthquakes off the Coast of Fiji

Abstract: This report examines the trend between magnitude, depth, and location of earthquakes off of Fiji since 1964. By performing point process spatial analysis on this dataset, I discovered that magnitude and depth are not correlated and that the earthquakes of the greatest magnitude usually occur near major tectonic plate junctions. Although this research can not tell us where an earthquake will hit, it can tell us where the earthquakes with the greatest magnitude are likely to occur. With this information we can further examine other tectonic plate junctions to see if the conclusion holds true and hopefully someday be able to implement safety procedures in those areas that are prone to earthquakes with highly dangerous magnitudes.

2 Graduate Math/Computer Science

PROJECT 4

Author(s): Chris Bukowski

Title: Investigating spatial correlation estimation when nonparametrically smoothing the mean trend

Abstract: Francisco-Fernandez & Opsomer (2005) introduce data-driven methods for bandwidth selection in a locally linear estimator of the mean trend for spatially correlated data. With spatial data, standard bandwidth selection methods for nonparametric estimation used for independent data are not appropriate. Under positive spatial correlation, the bandwidth selected will be too small, resulting in an overfit mean function and a spatial correlation that is too weak. The authors assess estimation of the mean function but not the correlation parameters. In some applications, it is of interest to correctly separate the mean trend from the spatial correlation. In simulated data, we show that these methods still do not completely differentiate between the mean function and the spatial correlation.
PROJECT 5
Author(s): Anna Forssén
Title: Unemployment Rates in Colorado in 2008: An Investigative Analysis
Abstract: This study examined the relationship between unemployment rates, percentage of minorities and high school graduation rates per county in Colorado in 2008. The spatial dependence among unemployment rates per county was incorporated into a simultaneous autoregressive (SAR) model. A transformation of the SAR model was used to assess whether or not the model sufficiently removed spatial dependence in the data. Two different neighborhood structures were considered in this study: a first-order weight matrix and a second-order weight matrix. Results indicate that percentage of minorities within each county, but not high school graduation rates, had a significant influence on unemployment levels in each county. Differences in results using the first-order and second-order weight matrices are also discussed.

PROJECT 6
Author(s): Tonya Lauriski-Karriker
Title: Can Respiratory Deaths in Colorado Counties be Linked to Air Pollution Sources?
Abstract: According to the U.S. Environmental Protection Agency, air pollution particles have been linked to various respiratory health problems. Particulate matter particles with diameter less than 10 micrometers posing the greatest problems. This project will determine if county particulate matter particles less than 10 micrometers and sources known to cause air pollution can be linked to respiratory deaths in Colorado during 2009. Respiratory deaths by county in Colorado are significantly spatially clustered. The goal of this project is to determine if the spatial dependence of respiratory deaths can be explained with particulate matter levels and common sources of air pollution including the presence/absence of interstates and coal power plants. Kriging was used to predict particulate matter levels at the centroid of each county. An iterative method that took the variability of these estimates into account was used to determine that mean particulate matter levels are not significant predictors of respiratory related deaths in Colorado during 2009. Further investigation also determined that the presence/absence of interstates and coal power plants are also not significant predictors of respiratory deaths. This project determined that the variability of respiratory related deaths can be best explained by the location of the centroid for each county.
PROJECT 7

Author(s): Chris Lorenzini

Title: Clustering Analysis of Las Vegas Criminals

Abstract: Police Officers are always looking for technology to aid in the understanding of criminal activity. By observing locations of crimes, Police Departments can potentially target areas of high risk and educate Law Enforcement Officers as to the inner working of common criminals. By analyzing criminal groups from 1999-2002 in Las Vegas, NV with the use of spatial statistics can help identify potential clustering. By using Ripley’s K-function under a homogeneous and inhomogeneous intensity, a set of data can be simulated to test whether or not we see a cluster in the data. Through the use of the K-function, several criminal groups showed a tendency to commit crimes in a clustered fashion, while other groups committed crimes in a random fashion. Results indicate spatial data may not be the only variables of interest, thus leading one to believe other predictors may indicate a better basis of criminal activity. Through further analysis of other variables, Police Departments will be better equipped to handle and prevent criminal activity.

PROJECT 8

Author(s): Megan Yoder

Title: Short Term Forecasting of Categorical Changes in Wind Power

Abstract: Wind power is becoming a much more common source of energy, so being able to accurately predict how much power will be generated by the wind is very important. Not only is it important to utility companies and wind power generators but also to consumers who need uninterrupted power. There are many models currently forecasting wind speed and power that use numerical models or statistical models or both. The model presented in this paper is unique in that the forecast is for a categorical change in wind power. This kind of forecast, as part of a possible suite of forecasts, can help utility managers when deciding how to allocate resources. If a manager has prediction and uncertainty information about whether the wind power will decrease or increase in the next hour or two, preparations can be made to bring in more or less power from other traditional sources such as coal or gas. Competing models are presented as benchmarks used to evaluate the performance of a proportional odds model. This new model uses ordinal logistic regression to predict a decrease, no change or increase in wind power over a short term horizon. Comparisons of the forecasts from the new model are made to the benchmarks.
3 Graduate Mining & Earth Sys Engineering

PROJECT 9
Author(s): Tracy Barnes

Title: Geostatistical Classification of Resources

Abstract: Classification of mining resources is often left to the subjective judgment of the persons producing a resource estimate. The goal of this project is to outline a method where resources can be classified in a consistent and statistically valid manner that does not rely on subjective judgment. Statistical methods currently in use in the mining industry classify large volumes of material in mass. The second goal of this project is to develop a method allowing blocks to be classified on an individual block by block basis fully accounting for any spatial correlation between blocks. Through the use of sequential simulation, the method presented here meets these goals and would be generally applicable for any mining resource estimation.

4 Graduate Hydrology

PROJECT 10
Author(s): Laura Condon

Title: Constructing gridded subsurface data from point measurements using spatial statistics

Abstract: The upper Klamath Basin in Southeastern Oregon and Northern California has experienced conflicts in recent years regarding allocation of its limited surface water supply. A physical hydrology model is currently being developed to assess optimal management strategies to deliver adequate water to agriculture while maintaining environmental standards. A vital input for this model is gridded subsurface aquifer parameters. This project develops gridded inputs from well measurements. Transmissivity measurements from well logs are analyzed to characterize the spatial structure of hydrogeologic units. Exponential semivariogram models are fit to the data and confidence intervals are estimated using the empirical semivariogram. Comparison of models shows that variance is isotropic but correlation lengths vary significantly with direction. Furthermore, tertiary volcanic rock and tertiary sedimentary rock have sufficiently different characteristics to warrant separate modeling. Overall, there is clear evidence of spatial correlation between measurements within distances of one to five kilometers. These finding are used to generate correlated Gaussian random fields with the turning bands method. Two example fields are presented at the end of this work. Future work will incorporate these simulations into the regional hydrology model and assess the impact that different random field generation approaches can have on regional water balance calculations.
PROJECT 11

Author(s): Erica Siirila

Title: An analysis of potential spatial correlations between the PCE partition coefficient ($K_d$) and permeability ($k$): revisiting the Borden field site

Abstract: Solute transport of reactive contaminants is a complex process involving the interaction between a contaminant and the surrounding porous media, and is further complicated by the effect of both chemical and physical heterogeneity. Spatial dependence of the retardation process sorption is investigated with respect to the perchloroethene (PCE) partition coefficient ($K_d$). Sorption, a process affecting subsurface contaminant transport, describes the immobilization of contaminant mass onto the solid phase of the surrounding aquifer material. The focus of this analysis is the investigation of spatial autocorrelation of the attributes $K_d$ and the permeability of the porous media ($k$). The semivariogram is used to test for autocorrelation, and is also used to perform a parametric sensitivity analysis for differences in, e.g., estimator of the empirical semivariogram used, bin width, fitted semivariogram model, and the effect of anisotropy. Spatial dependence between the attributes is also investigated via cross-semivariograms and a linear regression technique. Results show both $\ln(k)$ and $\ln(K_d)$ are sensitive to the empirical semivariogram estimator in conjunction with the selection of a bin width, where the appropriate fitted semivariograms all utilize the Cressie Hawkins empirical estimator at a bin width of 0.25 [m]. The effect of anisotropy in both attribute datasets is apparent by the lack of observed spatial dependence in the omnidirectional analysis, but apparent spatial correlation in the vertical direction. This result is also valid for the cross-semivariogram of the two attributes, where cross-correlation in the omnidirectional and horizontal directions fluctuate positive and negative cross-semivariance values, whereas a positive correlation is observed in the vertical direction.

PROJECT 12

Author(s): John Williams

Title: Propagation of spatial correlation structure through correlated land surface and atmospheric variables

Abstract: In order to optimize data collection strategies related to weather forecasts, an understanding of spatial relationships not only within individual variables, but among several variables could prove to be invaluable. For example, if the effects of soil moisture on wind speeds are found to extend over distances larger than the correlation lengths in heterogeneous soil moisture fields, observations of soil moisture in one location could be used to improve forecasts of wind speed in another. This study uses cross-semivariograms to perform a first-pass exploration of the spatial relationships that exist between the subsurface, land surface and atmosphere. This analysis shows qualitatively that there is a spatial
correlation across interfaces between soil moisture and latent heat flux, and between latent heat flux and wind speed, as long as these variables are coupled. The relationships break down when the atmosphere is decoupled from the land surface due to external forcing, such as a sea breeze, dominating the atmospheric system.

5  Graduate Geophysics

PROJECT 13

Author(s): Leon Foks

Title: Differences between raw airborne magnetic data and gridded data using minimum curvature

Abstract: Raw and gridded airborne magnetic measurements are compared with each other in order to understand the differences between the two. The first method is the multi-metric multi-response permutation procedure which compares two or more groups of data. The null hypothesis states that there are no differences across the groups and in this paper it is shown that there are no significant difference between the measured total field and the gridded total field using minimum curvature. Similar conclusions are reached for the two horizontal derivatives and the vertical derivative. Directional semivariograms of both raw and gridded data show that gridding does indeed have a smoothing effect and is pronounced in the calculation of the semivariogram.

PROJECT 14

Author(s): Leslie Godfrey

Title: New Strategies for Developing Vs30 Maps

Abstract: A limiting aspect of existing state-of-the-art strategies for generating estimated Vs30 maps from geologic (e.g., Wills and others) and topographic base maps (e.g., Wald and others) is that, while initially derived from and constrained by observed Vs30 values, both approaches fail to incorporate the original Vs30 measurements back into the map that has been created. Furthermore, while both predictive methods bring unique benefits to the problem, they are rarely used in combination with each other and with the Vs30 measurements. Because many seismic-hazard mapping studies and ground-motion prediction equations fundamentally rely on Vs30 as the site response explanatory variable, these problems require further attention. In this study, we examine alternative strategies to map Vs30 with estimated and observed Vs30 combinations, recognizing that the weighting at any location should be made with consideration of the spatial uncertainties of each of the contributing inputs. We examine kriging using Wald & Allen and Lee et al., cokriging using
Wald & Allen and Lee et al., and kriging using Iteratively Reweighted Generalized Least Squares (IRWGLS) methodologies. In the process of developing an optimal strategy for Vs30 map development, we are exploring: 1) the incorporation of measured Vs30 back into the map, 2) the employment of on-the-fly slope-based Vs30 correlations to better fit regional trends of Vs30 datasets, and 3) the addition of many Vs30 data not available in earlier studies. Ideal candidates for calibrating this methodology include Taiwan; Salt Lake City, Utah; and California.

Upon examining this study, one major find is that cokriging is not much better at prediction Vs30 than ordinary kriging is. Additionally, the kriged maps of Vs30 from topographic slope, of cokriging, the original Vs30 data, and the IRWGLS geology-based approach are surprisingly similar. The farthest from those was the IRWGLS slope-based method, surprisingly enough (since the regular slope-based method worked great). Overall, further investigation is necessary to determine which method is best, whether fitting the Vs30 vs. Topographic slope plot with a curve is better than binning, and whether the resolution of topographic slope plays a role in Vs30 estimation.

PROJECT 15

Author(s): Cericia Martinez

Title: Estimating terrain correction density value for gravity gradient data using spatial inference

Abstract: In geophysics, it is necessary to reduce measured potential field data for interpretation purposes. The gravity method requires a few corrections to the observed data in order to obtain the residual, or anomalous response due to geologic features. One of the corrections that must be applied to the observed data is called a terrain correction. Just as the name suggests, the terrain correction removes the response of the terrain from the observed data. In order to remove the contribution of the terrain (or terrain effect) from the overall measured value, a representative density value must be selected in order to calculate the magnitude of the terrain response to be removed. The purpose of this work is to explore methods to estimate the representative or average density that sufficiently removes the terrain effect from observed gravity gradient data. The observed gravity or gravity gradient data can be treated as geostatistical data. The hope is to exploit the fact that their is a spatial relationship in the observed data. The spatial correlation of the data is expected to vary according to the density value.