B61A MCC: Hall C Saturday 0830h

Automated Data Analysis: Research and Results Posters *(joint with OS, P)*

Presiding: **J Coughlan,** NASA Ames Research Center**; D Thompson,** NASA Ames Research Center**; C Potter,** NASA Ames Research Center

B61A-0700 0830h POSTER

Image Segmentation and Classification Using Local Fractal Dimension and Spatial Autocorrelation

Charles W Emerson[1] (616-387-3430; charles.emerson@wmich.edu)

Dale A Quattrochi[2] (256-961-9887; dale.quattrochi@msfc.nasa.gov)

Nina Siu-Ngan Lam[3] (225-388-6197; nlam@lsu.edu)

[1]Western Michigan University, Department of Geography, Kalamazoo, MI 49008, United States

[2]Louisiana State University, Dept of Geography and Anthropology, Baton Rouge, LA 70803, United States

[3]National Aeronautics and Space Admistration, Earth Science Dept. Global Hydrology and Climate Center SD60, Huntsville, AL 35812, United States

Fractal dimension is a measure of image complexity, which ranges from 2.0 for a perfectly uniform image to approximately 3.0 for a rough texture with adjacent bright and dark areas. The Morans I index of spatial autocorrelation is a measure of the degree to which a spatial pattern is clumped, random, or dispersed, and ranges from negative 1.0, to zero, to positive 1.0, respectively. When measured in overlapping and non-overlapping subsets of a remotely sensed image, these metrics provide both a means for content-based image segmentation and edge detection.

The problem of characterizing the spatial extent of urban development was used as an example of the utility of local measures of Morans I and fractal dimension in both a single-band image segmentation context and as an additional layer for spectral/spatial image classification. Landsat Thematic Mapper imagery from 1990 and Enhanced Thematic Mapper images from 2000 were combined with data from the last two decennial Censuses to benchmark these indices. In this example, Morans I, and to a lesser extent, local fractal dimension measures were significantly correlated to data on new housing starts, population growth, and density of the transportation network. Outliers from a regression of Morans I or fractal dimension values obtained for quarter-quadrangle (approximately 3.75 degrees latitude and longitude) segments of the original Landsat scenes obtained in 1990 and 2000 in most cases highlighted areas undergoing development or construction activities that were much greater than the regional average.

B61A-0701 0830h INVITED POSTER

Automated Analysis of Earth Sciences Data for Prediction of Land Surface: Ocean Climate Couplings

Vipin Kumar[1] (kumar@cs.umn.edu); Michael Steinbach[1] (steinbac@cs.umn.edu); Pang-Ning Tan[1] (ptan@cs.umn.edu); Steven Klooster[2] (sklooster@mail.arc.nasa.gov); Christopher Potter[3] (cpotter@mail.arc.nasa.gov); Shashi Shekhar[1] (shekhar@cs.umn.edu); Pusheng Zhang[1] (pusheng@cs.umn.edu); Vanessa Brooks Genovese[2] (vbrooks@mail.arc.nasa.gov)

[1]University of Minnesota, Dept. Computer Science and Engineering, Minneapolis, MN 55455, United States

[2]California State University Monterey Bay, Earth System Science and Policy, Seaside, CA 93955, United States

[3]NASA Ames Research Center, Ecosystem Science and Technology Branch, Moffett Field, CA 94035, United States

Our approach to automated computation analysis of Earth Sciences data is aimed at improving predictions of global biosphere properties such as net primary production (NPP) and surface hydrologic fluxes. Our main research objective has been to identify the best ocean climate indices (OCIs) to use for land biosphere predictions over the past 20 to 40 years, and presumably into the future. Well-known OCIs, such as the Southern Oscillation Index (SOI) or the Arctic Oscillation (AO) index, are single time series that represent the dynamics of surface temperature or pressure over large areas of the Earth's oceans. Results from our automated analysis suggest that, on a global level, OCIs can be highly correlated to NPP and predicted ecosystem carbon fluxes over more than 58 percent of the non-desert/ice covered land surface, commonly with a lead period of 2-6 months. SOI and AO equally explain the largest portions of these highly correlated area fluxes. To improve future biosphere predictions, we have developed methods based on cluster analysis for efficient automated discovery of previously undetected OCI areas. The centroids of these ocean clusters have the potential to summarize the global temporal dynamics of large ocean-land teleconnections for biosphere surface climate better than any known OCIs, including SOI and AO.

URL: http://www.ahpcrc.umn.edu/nasa-umn/

B61A-0702 0830h POSTER

On the Uncertainty of Satellite-Derived Biospherical Parameters

Robin D Morris[1] (650 604 0158; rdm@email.arc.nasa.gov)

Vadim N Smelyanskiy[1]

[1]NASA Ames Research Center , MS 269-2, Moffett Field, CA 94035-1000, United States

Satellite remote sensing typically measures reflected radiation in a number of spectral bands. From these measurements a large number of data products are produced. For example, the MODIS web site lists over 30 data products that are derived from the MODIS observations and ancillary data. These data products are then used as inputs to various ecological modeling systems.

It is germane to study the uncertainties in the estimates derived from the satellite observations, and their sources, as these uncertainties will affect the precision of the outputs of any ecological model using them as inputs. Knowledge of the sources of uncertainty may point to ways to reduce them.

We look in detail at the derivation of Leaf Area Index (LAI) and Available Water Capacity (AWC) from AVIRIS Normalized Difference Vegetation Index (NDVI) data. We attempt to trace back the origins of the relationships used to estimate LAI from NDVI, and to determine the uncertainty in each step. These uncertainties can be combined to give an estimate of the uncertainty in the final data product.

B61A-0703 0830h POSTER

Information Theoretic Approaches to Rapid Discovery of Relationships in Large Climate Data Sets

Kevin H. Knuth[1] ((212) 678-5568; knuth@email.arc.nasa.gov)

William B. Rossow[2] ((212) 678-5567; wrossow@giss.nasa.gov)

[1]NASA Ames Research Center, Computational Sciences Division MS 269-3, Moffett Field, CA 94035-1000, United States

[2]NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025, United States

Mutual information as the asymptotic Bayesian measure of independence is an excellent starting point for investigating the existence of possible relationships among climate-relevant variables in large data sets. As mutual information is a nonlinear function of its arguments, it is not beholden to the assumption of a linear relationship between the variables in question and can reveal features missed in linear correlation analyses. However, as mutual information is symmetric in its arguments, it only has the ability to reveal the probability that two variables are related. It provides no information as to how they are related; specifically, causal interactions or a relation based on a common cause cannot be detected. For this reason we also investigate the utility of a related quantity called the transfer entropy. The transfer entropy can be written as a difference between mutual informations and has the capability to reveal whether and how the variables are causally related. The application of these information theoretic measures is tested on some familiar examples using data from the International Satellite Cloud Climatology Project (ISCCP) to identify relations between global cloud cover and other variables, including equatorial pacific sea surface temperature (SST), over seasonal and El Nino Southern Oscillation (ENSO) cycles.

B61A-0704 0830h POSTER

Visualization of 2D Distributions from Models with Uncertainty

David L. Kao[1] (1-650-604-4375; davidkao@nas.nasa.gov)

Jennifer L Dungan[2] (1-650-604-3618; jdungan@mail.arc.nasa.gov)

Alison Luo[3] (1-831-459-2712; alison@cse.ucsc.edu)

Alex T Pang[3] (1-831-459-2712; pang@cse.ucsc.edu)

[1]NASA Ames Research Center, M/S T27A-2, Moffett Field, CA 94035

[2]NASA Ames Research Center, M/S N242-4, Moffett Field, CA 94035

[3]University of California, Santa Cruz, Computer Science Department, Santa Cruz, CA 95064

Maps representing geo- and bio-physical variables result from many Earth science models. The models may generate probability distributions at each location describing the full range of possible values. Alternatively, the models may generate multiple possible realizations from which probability densities may be estimated. These distributions are a rich description of uncertainty about the variables predicted by a model. Much in the same way confidence intervals are reported for non-spatial variables, there is a need to report distributions to completely represent model output.

Visualization methods exist for handling scalar, vector, tensor, or multivariate fields. However, if the data consist of distributions, the available options are quite limited. One way to handle this problem is to use scalar fields based on meaningful summaries of the distributions on a per grid-cell (or per-pixel) basis. For distributions that are parametric, summaries such as the mean and variance may provide complete description. For nonparametric distributions, statistics such as the median, inter-quartile range, kurtosis and skewness provide a useful, but not complete, description. Other useful summary information for nonparametric distributions include the number of peaks or modes. We present a tool that presents several summaries simultaneously in the form of colored maps, deformed surfaces, and line glyphs. We have also experimented with providing a probe for selecting pixels or profiles (row or column) for a graphical or shaded-surface display of the complete distribution(s).

In addition to pixel-wise distributions, information on uncertainty may be needed on a feature-wise basis. A feature, or multiple-pixel region, is often of greater interest than a single pixel. The area of the feature may change across realizations, implying an area distribution. Distributions for the area of each feature may also be summarized and/or more completely rendered.

Our exploratory tools for both pixel-wise and feature-wise uncertainty visualization have been developed on two small data sets, one representing spatial prediction (retrieval) of a variable using remotely sensed spectral information and conditional simulation, the other representing an ocean shelfbreak front simulation using error subspace statistical estimation (ESSE).

URL: http://www.cse.ucsc.edu/research/avis/nasa_is

B61A-0705 0830h POSTER

Machine Learning Techniques for Decision Support in Intelligent Data Management

Christopher Lynnes[1] (301-614-5185; lynnes@daac.gsfc.nasa.gov)

Jerome Miller[2] (jmiller@pop500.gsfc.nasa.gov)

Hampapuram Ramapriyan[3] (rama@rattler.gsfc.nasa.gov)

David Isaac[4] (david.isaac@teambps.com)

Robert Harberts[5] (harberts@gst.com)

[1]NASA/GSFC, Code 902 NASA/GSFC, Greenbelt, MD 20771, United States

[2]NASA/GSFC, Code 586 NASA/GSFC, Greenbelt, MD 20771, United States

[3]NASA/GSFC, Code 423 NASA/GSFC, Greenbelt, MD 20771, United States

[4]Business Performance Systems, 7364 Clifton Road, Clifton, VA 20124, United States

[5]Global Science Technology, Code 423 NASA/GSFC, Greenbelt, MD 20771, United States

NASA's growth in remote sensing data volumes has kept pace with Moore's Law, i.e., doubling every 18 months, with future growth likely from new instruments. Also, advances in instrumental design (e.g., hyperspectral scanners) and science algorithms are enabling more near-real-time applications of the data. The confluence of low-latency requirements with high data volumes and numbers of files poses major challenges for archive data management. In order to make the right data available at the right time, an archive will need to apply knowledge of the data content in its data management decisions. This decision support domain includes aspects such as automatic quality assessment, feature detection to support caching decisions, and content-based metadata to support efficient data selection. In this study, we evaluate a variety of machine learning algorithms for use in several decision support roles in intelligent data management.

Machine learning algorithms such as neural networks and clustering have been used for decision support in business and policy domains. These techniques have found some use in remote sensing, e.g., for cloud and land cover classification. Yet most research on remote sensing data rests on science-based algorithms, such as those based on radiative transfer equations. Machine learning for scientific applications faces challenges such as discretization constraints, non-physical basis, and the difficulty of assembling training sets. However, these difficulties may be less significant in the decision support role. For instance, it is often enough to know whether a data attribute exceeds a certain threshold when selecting it for an application, without knowing the exact value. The training data problem can be surmounted by using products output by the science-based algorithms. On the other hand, an advantage of machine learning algorithms for decision support is their speed once they have been trained. Data management decisions must be made while the "fresh" data are still on disk, and in time to service near-real-time applications, i.e., within a few hours or even minutes. The difficulties and advantages of machine learning algorithms are examined for their utility in decisions regarding data quality assessment, feature-based caching strategies and content-based data selection.

B61A-0706 0830h INVITED POSTER

Terrestrial Observation and Prediction System: Integration of satellite and surface weather observations with ecosystem models

Ramakrishna R Nemani[1] (nemani@ntsg.umt.edu); Petr Votava[1] (votava@ntsg.umt.edu); John Roads[2] (roads@ucsd.edu); Mike White[3] (mikew@cc.usu.edu); Peter Thornton[4] (thornton@ucar.edu); Joseph Coughlan[5] (jcoughlan@mail.arc.nasa.gov)

[1]University of Montana, NTSG, Missoula, MT 59812, United States

[2]Scripps Inst. of Oceanography, University of California, La Jolla, CA 92093, United States

[3]Utah State University, Department of Geography, Logan, UT 84322, United States

[4]NCAR, Climate Modeling Group, Boulder, CO 80305, United States

[5]NASA/Ames Research Center, Ecosystem Science Technology, Moffett Field, CA 94035, United States

Satellite data are widely used in land surface models to compute carbon and water exchange processes. However, much of this work is retrospective in nature. To better represent current land surface conditions in weather/climate models or to provide timely information on ecosystem conditions for natural resource management, one must move from retrospective to real-time analysis. A number of recent advances allow us to develop a system that would allow such real-time assimilation. These include consistent and timely availability of land surface products from EOS/MODIS, and on-line availability of weather data from a number of surface weather stations. We have developed a data assimilation system, terrestrial observation and prediction system, that integrates satellite data, surface weather observations and weather/climate forecasts with a terrestrial ecosystem model. TOPS produces daily 1km estimates of carbon and water fluxes using MODIS derived LAI, land cover and gridded meteorological data created using over 2000 surface weather stations over the conterminous U.S. Daily outputs are expressed as anomalies from historical normals that were computed using 20 years (1982-2001) of satellite and surface weather data. TOPS is also capable of using short/mid-term weather/climate forecasts to produce forecasts of land surface conditions (snow pack, run off, soil moisture and primary production) that are useful in resource management.

B61A-0707 0830h POSTER

Analysis of Salinity Intrusion in the San Francisco Bay-Delta Using a GA-Optimized Neural Net, and Application of the Model to Prediction in the Elkhorn Slough Habitat

David E. Thompson[1] (1-650-604-4759; dethompson@mail.arc.nasa.gov)
Thirumalainambi Rajkumar[2] (1-650-604-1148; rajkumar@mail.arc.nasa.gov)

[1]NASA Ames Research Center, M.S. 269-1; Bldg 269, Rm 290, Moffett Field, CA 94035-1000, United States

[2]SAIC, NASA Ames, M.S. 269-1, Moffett Field, CA 94035-1000, United States

The San Francisco Bay Delta is a large hydrodynamic complex that incorporates the Sacramento and San Joaquin Estuaries, the Suisan Marsh, and the San Francisco Bay proper. Competition exists for the use of this extensive water system both from the fisheries industry, the agricultural industry, and from the marine and estuarine animal species within the Delta. As tidal fluctuations occur, more saline water pushes upstream allowing fish to migrate beyond the Suisan Marsh for breeding and habitat occupation. However, the agriculture industry does not want extensive salinity intrusion to impact water quality for human and plant consumption. The balance is regulated by pumping stations located along the estuaries and reservoirs whereby flushing of fresh water keeps the saline intrusion at bay. The pumping schedule is driven by data collected at various locations within the Bay Delta and by numerical models that predict the salinity intrusion as part of a larger model of the system. The Interagency Ecological Program (IEP) for the San Francisco Bay / Sacramento-San Joaquin Estuary collects, monitors, and archives the data, and the Department of Water Resources provides a numerical model simulation (DSM2) from which predictions are made that drive the pumping schedule. A problem with DSM2 is that the numerical simulation takes roughly 16 hours to complete a prediction. We have created a neural net, optimized with a genetic algorithm, that takes as input the archived data from multiple gauging stations and predicts stage, salinity, and flow at the Carquinez Straits (at the downstream end of the Suisan Marsh). This model seems to be robust in its predictions and operates much faster than the current numerical DSM2 model. Because the Bay-Delta is strongly tidally driven, we used both Principal Component Analysis and Fast Fourier Transforms to discover dominant features within the IEP data. We then filtered out the dominant tidal forcing to discover non-primary tidal effects, and used this to enhance the neural network by mapping input-output relationships in a more efficient manner. Furthermore, the neural network implicitly incorporates both the hydrodynamic and water quality models into a single predictive system. Although our model has not yet been enhanced to demonstrate improve pumping schedules, it has the possibility to support better decision-making procedures that may then be implemented by State agencies if desired. Our intention is now to use our calibrated Bay-Delta neural model in the smaller Elkhorn Slough complex near Monterey Bay where no such hydrodynamic model currently exists. At the Elkhorn Slough, we are fusing the neural net model of tidally-driven flow with in situ flow data and airborne and satellite remote sensing data. These further constrain the behavior of the model in predicting the longer-term health and future of this vital estuary. In particular, we are using visible data to explore the effects of the sediment plume that wastes into Monterey Bay, and infrared data and thermal emissivities to characterize the plant habitat along the margins of the Slough as salinity intrusion and sediment removal change the boundary of the estuary. The details of the Bay-Delta neural net model and its application to the Elkhorn Slough are presented in this paper.

B61A-0708 0830h INVITED POSTER

Deductive Coordination of Multiple Geospatial Knowledge Sources

Richard Waldinger[1] (1-650-859-2216; waldinger@ai.sri.com); Martin Reddy[2] (1-510-752-4093; reddy@pixar.com); Christopher Culy[1] (1-650-859-5082; culy@ai.sri.com); Jerry Hobbs[3] (1-310-488-9140; hobbs@isi.com); Peter Jarvis[1] (1-650-859-3209; jarvis@ai.sri.com); Jennifer L Dungan[4] (1-650-604-3618; jdungan@mail.arc.nasa.gov)

[1]Artificial Intelligence Center, SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025, United States

[2]Studio Tools, Pixar Animation Studio, 1200 Park Avenue, Emeryville, CA 94608, United States

[3]Information Sciences Institute University of Southern California, 4676 Admiralty Way, Marina del Rey, CA 90292, United States

[4]NASA Ames Research Center, MS 242-4, Moffett Field, CA 94035-1000, United States

Deductive inference is applied to choreograph the cooperation of multiple knowledge sources to respond to geospatial queries. When no one source can provide an answer, the response may be deduced from pieces of the answer provided by many sources. Examples of sources include (1) The Alexandria Digital Library Gazetteer, a repository that gives the locations for almost six million place names, (2) The Cia World Factbook, an online almanac with basic information about more than 200 countries. (3) The SRI TerraVision 3D Terrain Visualization System, which displays a flight-simulator-like interactive display of geographic data held in a database, (4) The NASA GDACC WebGIS client for searching satellite and other geographic data available through OpenGIS Consortium (OGC) Web Map Servers, and (5) The Northern Arizona University Latitude/Longitude Distance Calculator.

Queries are phrased in English and are translated into logical theorems by the Gemini Natural Language Parser. The theorems are proved by SNARK, a first-order-logic theorem prover, in the context of an axiomatic geospatial theory. The theory embodies a representational scheme that takes into account the fact that the same place may have many names, and the same name may refer to many places. SNARK has built-in procedures (RCC8 and the Allen calculus, respectively) for reasoning about spatial and temporal concepts. External knowledge sources may be consulted by SNARK as the proof is in progress, so that most knowledge need not be stored axiomatically. The Open Agent Architecture (OAA) facilitates communication between sources that may be implemented on different machines in different computer languages. An answer to the query, in the form of text or an image, is extracted from the proof. Currently, three-dimensional images are displayed by TerraVision but other displays are possible. The combined system is called Geo-Logica. Some example queries that can be handled by Geo-Logica include: (1) show the petrified forests in Oregon north of Portland, (2) show the lake in Argentina with the highest elevation, and (3) Show the IGPB land cover classification, derived using MODIS, of Montana for July, 2000.

Use of a theorem prover allows sources to cooperate even if they adapt different notational conventions and representation schemes and have never been designed to work together. New sources can be added without reprogramming the system, by providing axioms that advertise their capabilities. Future directions include entering into a dialogue with the user to clarify ambiguities, elaborate on previous questions, or provide new information necessary to answer the question. In addition, of particular interest is to deal with temporally varying data, with answers displayed as animated images.

URL: http://www.ai.sri.com/project/ComposeMultipleSources

B

B61A-0709 0830h POSTER

Distributed Datamining for NASA/NOAA databases

Rong Chen[1] (1-509-335-4969; rchen@eecs.wsu.edu); Byung H Park[3]; Krishnamoorthy Sivakumar[1] (1-509-335-4969; siva@eecs.wsu.edu); Hillol Kargupta[2] (hillol@csee.umbc.edu); Jianjie Ma[1]; Meng Da[1]

[1]Washington State University, School of EECS, Pullman, WA 99164-2752, United States

[2]University of Maryland Baltimore County, Department of CSEE 1000 Hilltop Circle, Baltimore, MD 21250, United States

[3]Oak Ridge National Laboratory, PO BOX 2008, Oak Ridge, TN 37831, United States

Raw data is useful only when it is transformed into knowledge or useful information. This involves data analysis and transformation to extract interesting patterns and correlations among the problem variables. The advent of large distributed environments in both scientific and commercial domains introduces a new dimension to this process — a large number of distributed sources of data that can be used for discovering knowledge. Cost of data communication between the distributed databases is a significant factor in an increasingly mobile and connected world with a large number of distributed data sources. This cost consists of several components like (a) Limited network bandwidth, (b) data security, and (c) existing organizational structure of the applications environment. The field of Distributed Knowledge Discovery and Data Mining (DDM) studies algorithms, systems, and human-computer interaction issues for knowledge discovery applications in distributed environments for minimizing this cost.

In this work, we consider a Bayesian network (BN) model to represent uncertain (probabilistic) knowledge. We consider a collective approach to learning a Bayesian network from distributed heterogenous data. In this approach, we first learn a local Bayesian network at each site using the local data. Then each site identifies the observations that are most likely to be evidence of coupling between local and non-local variables and transmits a subset of these observations to a central site. Another Bayesian network is learnt at the central site using the data transmitted from the local site. The local and central Bayesian networks are combined to obtain a collective Bayesian network, that models the entire data. Theoretical justification that demonstrate the feasibility of our approach have presented elsewhere.

In the present work, we have applied our distributed BN learning algorithm to some earth science data. Specifically, our experiments use data from two distributed sources: NASA DAO data and NOAA SAA data. The NASA DAO data is a subset of the Data Assimilation Office's (DAO) monthly mean data set. It has global spatial coverage and a temporal coverage ranging from March 1980 to November 1993. The NOAA SAA data is a product of NOAA and US department of defense (DOD) US Polar-orbiting environment satellites (POES). Seventeen features from NASA DAO and eight features from NOAA SAA data was used in

our experiments. A Bayesian network (BN) model was first contructed from the two datasets combined. This BN, referred to as the centralized BN, served as the ground truth for comparing the performance of our collective BN learning algorithm. Our preliminary experiments reveal a number of interesting trends. Correlations between specific DAO and NOAA data features are evident. Specific features are consistently observed as root nodes in the BN, suggesting that these features could possibly be the "cause" for certain phenomenon. Seasonal trends in the data reflect appropriate seasonal changes in the BN model.

URL: http://www.csee.umbc.edu/∼hillol

B61A-0710 0830h POSTER

Knowledge Discovery from Remotely Sensed Vegetation Indices

Amanda B. White[1] (217-333-0107; abwhite@uiuc.edu)

Praveen Kumar[1] (217-333-4688; kumar1@uiuc.edu)

[1]University of Illinois, Hydrosystems Laboratory 205 N. Mathews Ave., Urbana, IL 61801, United States

The objective of this research is to develop KDD (knowledge discovery in databases) techniques for spatio-temporal geo-data, and use these techniques to examine seasonal and inter-annual vegetation health signals. The underlying hypothesis of the research is that the signatures of inter-annual variability of climate on vegetation dynamics as represented by the statistical descriptors of vegetation index variations depend upon a variety of attributes related to the climate, physiography, topography, and hydrology. Several scientific questions related to the identification and characterization of the inter-annual variability ensue as a consequence of this hypothesis. Various vegetation indices will be enlisted to represent vegetation health, such as NDVI (normalized difference vegetation index), EVI (enhanced vegetation index), LAI (leaf area index), FPAR (fraction of photosynthetically active radiation), PSN (photosynthesis), and NPP (net primary product). Relationships between these indices and topography and its derivatives (slope, aspect, etc.), nearness to water bodies, precipitation, temperature, etc. will be analyzed.

Preliminary investigations were performed using 13 years of 1 km resolution NDVI data from the AVHRR instrument on NOAA's POES (polar-orbiting operational environmental satellite). Deviations from the 13-year average were utilized in order to identify anomalous behavior. The pilot KDD technique includes distance-based clustering algorithms and Apriori association rule algorithms adapted for spatial-temporal data. Future work will incorporate more complex algorithms such as density-based clustering and constraint-based association mining algorithms.

B61A-0711 0830h POSTER

Self-Organizing Maps Applied to Mineral Spectra: Pure and Mixed

Robert C. Hogan[1,2] (650 604 0780; hogan@cosmic.arc.nasa.gov)

Ted L. Roush[2] (650 604 3526; troush@mail.arc.nasa.gov)

[1]Bay Area Environmental Research Institute, 645 Anthony Court, Sonoma, CA 95476, United States

[2]NASA Ames Research Center, MS 245-3, Moffett Field, CA 94035-1000, United States

Meaningful classification of spectral data is an important first step in mineralogical analysis of terrestrial and planetary surfaces. The large data sets being gathered (Tbytes) can not be manually evaluated in a timely, efficient manner. This suggests automated classification and interpretative processing must be developed and applied.

Spectral libraries, databases of simulated or experimental spectra, play a crucial role in the identification of remotely obtained spectral data. At infrared wavelengths, linear mixing models are widely used to ascertain the principle mineral components contained in an unknown sample. An exhaustive search for spectral library candidates epresenting these components for a large database of samples is very CPU intensive, especially as the number of required components grows. A method to reduce the set of candidates in a linear mixing analysis would greatly improve the efficiency of the identification process. Classifications related to chemical and structural properties which are manifested in spectral features can be useful in this regard.

We have developed a scheme based upon self-organizing maps (SOM) that organizes spectra in an unsupervised fashion into a 2-D map. This scheme was applied to the Arizona State University (ASU) thermal emission mineral spectral library and separated the samples into clusters that relate to distinct chemical and structural groups. Three hierarchical levels of distinction arise from the SOM classification; mineral class (e.g. silicates vs. oxides), mineral subclass (e.g. inosilicates vs. nesosilicates) and mineral group (e.g. pyroxenes vs. amphiboles). In addition, we investigated spectral mixing such as occurs in natural samples. An 'unknown' spectrum, constructed from a linear mix of ASU library spectra, is 'classified' by locating it in the 2-D SOM map. Through an iterative process the end members composing the unknown mixture were recovered by evaluating the members of the library that were closest in proximity to the unknown on the 2-D SOM map. We will show our SOM classifications of the ASU library and illustrate, with several mixture examples, how the SOM map is used to zero in on library end members.

B61A-0712 0830h POSTER

Autonomous Science Analysis with the New Millennium Program-Autonomous Sciencecraft Experiment

Thomas Doggett[1] (480-965-9872; thomas.doggett@asu.edu); Ashley Gerard Davies[2] (818-393-1775; Ashley.Davies@jpl.nasa.gov); Rebecca A. Castano[2] (818-393-5344; becky@aig.jpl.nasa.gov); Victor R. Baker[3] (520-621-7875; baker@hwr.arizona.edu); James M. Dohm[3] (520-626-8454; dohm@hwr.arizona.edu); Ronald Greeley[1] (480-965-7045; greeley@dione.la.asu.edu); Kevin K. Williams[4] (202-357-1424; williamsk@nasm.si.edu); Steve Chien[2] (818-393-5320; chien@aig.jpl.nasa.gov); Robert Sherwood[2] (818-393-5378; Rob.Sherwood@jpl.nasa.gov)

[1]Arizona State University, Department of Geological Sciences, Tempe, AZ 85287, United States

[2]Jet Propulsion Laboratory-California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, United States

[3]University of Arizona, Department of Hydrology and Water Resources, Tucson, AZ 85721, United States

[4]Center for Earth and Planetary Studies, National Air and Space Museum, Smithsonian Institution, Washington, DC 20560, United States

The NASA New Millennium Program (NMP) is a testbed for new, high-risk technologies, including new software and hardware. The Autonomous Sciencecraft Experiment (ASE) will fly on the Air Force Research Laboratory TechSat-21 mission in 2006 is such a NMP mission, and is managed by the Jet Propulsion Laboratory, California Institute of Technology. TechSat-21 consists of three satellites, each equipped with X-band Synthetic Aperture Radar (SAR) that will occupy a 13-day repeat track Earth orbit. The main science objectives of ASE are to demonstrate that process-related change detection and feature identification can be conducted autonomously during space flight, leading to autonomous onboard retargeting of the spacecraft. This mission will observe transient geological and environmental processes using SAR. Examples of geologic processes that may be observed and investigated include active volcanism, the movement of sand dunes and transient features in desert environments, water flooding, and the formation and break-up of lake ice. Science software onboard the spacecraft will allow autonomous processing and formation of SAR images and extraction of scientific information. The subsequent analyses, performed on images formed onboard from the SAR data, will include feature identification using scalable feature "templates" for each target, change detection through comparison of current and archived images, and science discovery, a search for other features of interest in each image. This approach results in obtaining the same science return for a reduced amount of resource use (such as downlink) when compared to that from a mission operating without ASE technology. Redundant data is discarded. The science-driven goals of ASE will evolve during the ASE mission through onboard replanning software that can re-task satellite operations. If necessary, as a result of a discovery made autonomously by onboard science processing, existing observation sequences will be pre-empted to obtain data of potential high scientific content.

Flight validation of this software will enable radically different missions with significant onboard decision-making and novel science concepts (onboard decision making and selective data return).

This work has been carried out at the Jet Propulsion Laboratory-California Institute of Technology, under contract to NASA.

URL: http://ase.jpl.nasa.gov

B61A-0713 0830h POSTER

An Interoperable Framework for Mining and Analysis of Space Science Data (F-MASS)

Yubin Matt He[1] (256-824-5161; mhe@itsc.uah.edu); Rahul Ramachandran[1] (256-824-5157; ramachan@itsc.uah.edu); Xiang Li[1] (256-961-7969; xli@itsc.uah.edu); Sara Graves[1] (256-824-6064; sgraves@itsc.uah.edu); Wladislaw Lyatsky[2] (lyatsky@cspar.uah.edu); Arjun Tan[2] (atan@aamu.edu); Glynn Germany[3] (256-824-5129; germanyg@cspar.uah.edu)

[1]Matt Yubin He, TH S331 The Information Technology and System Center, The University of Alabama in Huntsville,, Huntsville, AL 35899, United States

[2]Wladislaw Lyatsky, Department of Physics Alabama AM University, Normal, AL 35762, United States

[3]Glynn Germany, 3Center for Space Plasma, Aeronomy, and Astrophysics Research, University of Alabama in Huntsville, Huntsville, AL 35899, United States

An Interoperable Framework for Mining and Analysis of Space Science Data (F-MASS) is currently being created by an interdisciplinary team composed of information technology and space science researchers. Information technology researchers on the F-MASS team are working with space scientists to adapt the innovative Algorithm Development and Mining system to provide a comprehensive framework for space science data investigations. The research will focus on problems such as the handling of the large datasets inherent in the space science communities, the heterogeneity of these data, and the incorporation of specialized data mining algorithms and domain-specific user interfaces into the framework. A set of three case studies has been selected to drive this effort. The first case study being researched is identification of polar cap boundaries. The polar cap boundary is an important parameter that determines how much energy is stored in the magnetotail. The stored energy is sporadically released, leading to the formation of strong geomagnetic disturbances, so-called substorms, which are associated with strong ionospheric currents in high-latitude ionosphere and auroras. Spacecraft UV auroral images are being analyzed using different heuristic and mining algorithms for 2-D patterns of auroras. One of the difficulties in developing such algorithms is the uneven and dynamic form of the polar cap boundary. Initial results from this case study will be presented.

B61B MCC: Hall C Saturday 0830h

Application and Validation of Land Surface Products From the MODIS Sensor Posters

Presiding: **D P Turner,** Oregon State University**; S W Running,** University of Montana

B61B-0714 0830h POSTER

Overview of the Validation of Land Surface Products From the MODIS Sensor

Jeffrey T. Morisette[1] (201-614-6676; jeff.morisette@gsfc.nasa.gov)

Jeffrey L. Privette[1] (301-614-6630; jeff.privette@gsfc.nasa.gov)

Christopher O Justice[2] (301-405-1600)

[1]NASA's Goddard Space Flight Center, Code 923 Building 33, Room G325, Greenbelt, MD 20771, United States

[2]University of Maryland, Geography, 2181 Lefrak Hall, College Park, MD 20742, United States

The MODIS Land team (MODLAND) is producing a suite of global land products whose uncertainty will be estimated through validation activities. The MODIS Land team will base its validation work on the comparison of its products to similar products derived from independent sources. The independent products are derived by combining in situ (field and tower) data with imagery from airborne and spaceborne sensors. In many cases the imagery and airborne data can be used to validate more than one product. With this, the EOS Land Validation Core Sites have been established. These sites provide easy access to multi-scale imagery as well as field and tower measurements. The poster will present the data available for the Core sites and the web-based infrastructure to access these data. It will