

Appendix B

Module Descriptions

Co-PI: Greg Baker

Institution: The Ohio State University

Discipline: Mathematics and Computer Science

Module 1: Vector Spaces and Subspaces: The challenge we start with is an appropriate description for a structure, for example, a crane. Vectors can then be introduced as locating positions, in particular, in three-dimensional space. Here computer graphics will be useful to illustrate vectors and how they point to locations. Next, we introduce the concept of relative position, which can be used to draw the sides of the crane, for example. This is also the time to discuss the origin of vectors. As we fill in a description of a crane (using a simplified diagram), it will become clear that projections are useful ways to see the space. This will lead to the idea that we want to perform operations on vectors, and thus we need matrices. Having illustrated addition (and subtraction) by using vectors to locate points, the next stage is the idea of zooming in (or out) to see what we have. Thus, scalar multiplication will be introduced. Remember all these concepts can be illustrated through computer graphics. Finally, we will introduce the mathematical concepts of vector spaces, drawing on the experiences already gained.

Module 2: Linear Transformations: Based on the module on vector spaces, where we introduced the description of a mechanical structure, a simple crane, we now ask how we might study the forces acting on the structure as the crane tries to lift a body. This module will revise and strengthen the concepts of vectors, as we introduce the ideas of forces as vectors. Ultimately, we will obtain a linear system of equations that gives the balance of forces. With this example in mind, we can describe the methods for solving linear equation, in particular, Gauss' method. Computer animation will play a large role in showing how the algebra is performed. Several windows can be used to illustrate the algebraic steps both on the linear system and the consequence on the matrix describing the linear system. This example is also the perfect opportunity to discuss how several steps can be performed in parallel. This is a natural example of the use of parallel computations and should help students see the ideas beyond parallel algorithms.

Module 3: Eigenvalues and Eigenvectors: By using the example of a singular system in the previous module, we may introduce formally the idea of the null-space. The null-space is composed of vectors with the remarkable property that multiplication by the matrix produces a zero result. This raises the question of the presence of other vectors that would return some multiple of them when multiplied by the matrix. The incredible value of these vectors (eigenvectors with non-zero eigenvalues) may be illustrated with several interesting examples. First, we note that if the forcing vector is an eigenvector, then we know the solution to the linear system is automatically some multiple of the eigenvector. We may continue to develop this idea to show that eigenvectors effectively transform the matrix to a diagonal one. Of course, the solution to a diagonal system is easily found, provided known of the eigenvalues are zero (this provides another view of the difficulties in seeking solutions to singular matrices). After this mainly algebraic presentation, it is useful to give some geometric illustrations of eigenvalues. A good example is to draw the connection between eigenvalues and the symmetries in a geometrical shape. This leads naturally to the introduction of Householder rotations to calculate the eigenvalues of a matrix. Finally, we point out that the largest eigenvalue in magnitude describes the largest possible response to the application of the matrix. From an engineering perspective, the matrix may be viewed as a black box. Then the largest response possible comes when the eigenvector with largest eigenvalue in magnitude is used as the input or stimulus.

These ideas will be applied to the design of the crane so that it can pick up the largest possible load.

Module 4: Curve Fitting: Theoretical considerations often give the form for a result but may not give precise values for parameters. For example, we know that we can stretch a substance by applying an appropriate force. How much it stretches depends on the strength of the force and the material properties of the substance. Engineers determine the constant of proportionality through specifically designed tests. The results are a series of measurements, and the objective is to fit the “best” line through the data points. Because there are always experimental errors, the data never fall exactly on a straight line. The least squares fit gives one way to calculate the “best” fit and this is as good a place to introduce the QR algorithm. Many of the concepts from the module on eigenvalues will be useful here. By introducing Bayesian statistics we can show how knowledge of the errors in the data can be used to ascertain the reliability of the fit. At this stage, we are ready to consider the general case of fitting some curve to data points. When the data is essentially free of error, as might be the case when evaluating a known function, then we can use more elaborate representations for the fit. The most common example is a high-order polynomial. The value of such fits is that the standard operators from calculus, differentiation and integration, can be applied to the polynomial representation. The dangers of using high order polynomials are easily revealed with some well known examples, and they motivate the need for piecewise approximations and composite integration rules. From a geometric viewpoint, we can use computer graphics to highlight the approximate representations, and their derivatives and integrals.

Module 5: Fourier Series: Eigenvectors are very useful in the study of finite-dimensional vector spaces. The theme in this module is to establish the value of trigonometric functions as eigenfunctions for infinite-dimensional vector spaces, such as periodic functions. By using symbolic manipulators, we can calculate the Fourier coefficients (or amplitudes) and display them graphically in Fourier space for various examples. We can show how properties of the functions are revealed in the spectrum, for example, the presence of sharp transitions in the function. Symbolic manipulations are effective only for a limited range of functions. Thus we introduce the approximations associated with the discrete transformation. We can start by relating the calculation of the amplitudes with the process of fitting a finite trigonometric series to a function. The remarkable thing is that the matrix for the linear system of equations that determine the coefficients can be factored into a product of simple matrices. This decomposition can be related to elementary properties of trigonometric functions. Unfortunately, the discrete spectrum has some quirks that require careful explanation. Once again, computer graphics can be used to demonstrate the relationship between functions and their discrete spectrum, and to remove the sources of confusion. With this knowledge, students will be able to calculate solutions to many different problems. For example, many systems exhibit persistent periodic patterns, and the Fourier series provides a natural and effective representation for these patterns.

Co-PI: Kim Baldrige

Institution: San Diego Supercomputer Center

Discipline: Chemistry

Module 1: Statistical Mechanics: Because statistical mechanics is a topic that is given the least amount of time in a typical physical chemistry course, this portion will be the most extensive of the three areas. It will be designed to be as self-instructive as possible so that

motivated students could learn the material independently in chemistry departments where it was not possible to offer the course. The tutorial is divided into two main sections. Section 1 will allow users to choose from a list of individualized topics in statistical mechanics from which to study. Section 2 will involve an in-depth study of a particular system as chosen by the user. The user plays a major role in the direction of the calculation by choosing which approximations and correction factors to include. This module will also include Thermodynamics as it arises from derivations of statistical mechanics phenomenon.

Module 2: Quantum Mechanics: The quantum mechanics component will include a series of tutorial modules covering the many areas of chemical bonding studied in physical chemistry. The most basic example is the one- and three-dimensional particle-in-the-box models. The wave and probability functions and energy level expression are obtained for the model, and how the results are related to the translational motion of the molecule. Similarly, a harmonic oscillator tutorial will allow study of the vibrational motion of a molecule. The hydrogen atom is a model for electronic motion in molecules. The hydrogen molecule-ion will also be addressed. Such programs allow study of fundamental concepts relating interference, bonding, and antibonding. Many electron systems will be included, as these systems are found to be complicated and may not always be studied in great detail in introductory physical chemistry courses. An extensive analysis involving the lithium hydride molecule will serve to expose the more advanced and motivated students to this area of quantum mechanics. The program will lead the user through a simple, minimal basis set Hartree-Fock SCF calculation of the LiH molecule

Module 3: Kinetics: The kinetics component will include tutorial modules that allow one to study some of the most general types of reaction processes. The selection includes study of simple first- and second-order reactions, consecutive irreversible first-order reactions, reversible first-order reactions, competitive first-order reactions, and the steady-state approximation. Two programs which involve numerical methods for solving rate expressions will be included particularly for more advanced to research levels. The fourth order Runge Kutta method used (a) to solve a simple first order reaction; demonstrating the comparison between the analytic solution and the numerical solution, and (b) to give concentration-time data for a nested set of elementary reactions which form a complex mechanism. This latter feature is actually useful for analysis of research data. Also included in this component will be study of the Arrhenius equation, transition state theory, and three simple programs involving the kinetic-molecular theory of gases.

Co-PI: Wayne Becketl

Institution: Capital University

Discipline: Chemistry/ Biochemistry

Module 1: Tools for Genomics and Proteomics: The twin areas of genomics and proteomics essentially concern the management of information concerning nucleic acids and amino acids. The analysis of large quantities of data from DNA or protein sequences is an aspect of “data mining” in ordered databases. In addition, the assembly of small fragments of DNA into a consistent, large sequence is essential in forming entire genomes. For protein analysis, the identification of proteins by mass spectroscopy and by two-dimensional gel electrophoresis can be automatically carried out by several different public domain software packages. This module will introduce the student to the NCBI and EMBL packages, as well as to stand-alone mass

spectroscopy analysis software. This module will complement modules created by Charles Daniels.

Module 2: Visualizing Protein Structures and Computing Structural Properties:

The analysis of protein structure and function involves not only a good understanding of the chemistry of amino acids and their interactions, but also the three-dimensional structural motifs associated with different classes of proteins. It is now possible to make significant progress in this analysis by using internet-based or public domain software. The overall problem involves the understanding of the structure and function of known enzymes and proteins and how the three-dimensional structure of the reactive side chains contributes to protein function. This module introduces the student to chemical and computational aspects of protein structure and function. Both internet-based and stand-alone software will be examined.

Module 3: Predicting Protein Structure and Function from Sequence: Long considered the “Holy Grail” of protein chemistry, the prediction of structure and function from amino acid sequence has taken on increased importance with knowledge of the complete genome sequences in a number of organisms. This is particularly true because: (1) proteins with very similar amino acid sequence often have completely different structures and/or functions and (2) proteins with very different sequences sometimes have similar structures and/or functions. The manner in which the primary amino acid sequence determines the final, folded structure has been called the *second genetic code*. It is fair to say that we are still learning to read it. This module introduces the student to many of the techniques of protein structure prediction. These include homology modeling and motif analysis. The particular family of proteins to be examined is the histone-like proteins from Archeons. These organisms exist in surroundings, which range from room temperature to over 100° C.

Co-PI: Charles Daniels

Institution: The Ohio State University

Discipline: Microbiology

Module 1: Gene Finding: How does an investigator identify and distinguish the biologically significant protein coding information in a string of DNA sequence information? Biological background for this module will include a review of the basic principals of DNA structure, gene structure and the genetic code. A progressive approach will be used to illustrate how computational tools are applied to the problem of identifying gene regions within DNA sequences. Current methods use simple rule based approaches, statistical fitness based on codon usage patterns and the application of probabilistic methods. The probabilistic approaches combine the rule-based and statistical methods with search tools that identify other specific biological features, such as intron-exon boundaries, and rank the likelihood of the predictions. These methods will illustrate both the complexity of the biological information, in terms of information content, and the challenge faced by computational biologists in identifying the “true” information. These are central questions for the bioinformatician and the molecular biologist. Computer science and mathematics students will see how statistical methods are applied to the analyses of DNA sequences and how web-based tools are used to present complex text and graphic images. All of the needed programs are freely available as web-based applications. Data for analysis will be constructed from models designed to illustrate basic concepts. Homework problems and projects will be constructed from data presented at the National Center for Biotechnology Information (NCBI). Students will be able to compare the

results of the predictive methods with the actual annotation that is provided by the submitter of the sequence. Unannotated DNA sequence data is also freely available on the WWW for “discovery-based” projects. In this case students can perform primary analysis using raw genomic sequence data. Use of the NCBI database will also act as an extension of the first module on biological databases and build a broader understanding of this resource.

Module 2: Gene Identification: How is the biological identity of a potential gene product predicted? This is one of the primary problems faced by bioinformaticians. The genetic code provides rules for the prediction of open reading frames; however, these data do not allow assignment of a function to the gene product. Current predictive methods depend on the identification of homologs or related sequences that have already been identified and are present in the NCBI databases. The problem is reduced to the questions: Are there any related sequences present in the databases? And if so, is this relationship sufficiently significant that an assignment of function can be made? Methods for the identification of gene products are based on a series of tools that measure the relatedness of nucleic acid or protein sequences. This is achieved by finding the best alignment between two macromolecules. The alignment problem will be present to the students as a progression of methods starting with simple matrix comparisons visualized as dot plots, then extended to introduce global alignments, which utilized dynamic programming methods. The dynamic programming approach is similar to the familiar “traveling salesman” problem. The final step is to apply these methods to query a large database, such as NCBI, searching for related molecules. In this later step we will also address the problem of determining the significance of the match. For this we will introduce concepts of probability. Biological science students will see the difficulties and the limitations in assigning gene identities based solely on sequence information. CS/Math students will see the applications of statistical tools to the problem of pattern matching. The importance of data management and the visualization of sequence alignment will also be emphasized. The problems will be designed to illustrate the basic concepts of alignment and then extended to searching larger databases. The NCBI site provides a search tool, Basic Local Alignment Search Tool or BLAST, which allows users to query the NCBI databases. This is the commonly used method. There are numerous options for homework and project assignments, and many of these can be built to be extensions of the assignments from the preceding module. The NCBI site also provides tutorials with documented examples.

Co-PI: Lisette de Pillis
Institution: Harvey Mudd College
Discipline: Mathematics

Co-PI: Ami Radunskaya
Institution: Pomona College
Discipline: Mathematics

Module 1: Pattern Formation in Biological Systems: In this module we focus on the biological question of how animal coat patterns form. For example, how would spots develop on a leopard's tail? The mathematical modeling component of this module leads to the development of a system of partial differential equations. We would then simplify the model to look only at the steady-state solutions of these equations. This will give rise to an elliptic boundary value problem. We derive analytical solutions for only the simplest geometries. We then focus on finding computational solutions. This will involve straightforward spatial discretizations, which in turn lead to a potentially large system of linear equations. The efficient numerical solution of linear equations is still a topic of research, but there are several well-established fundamental approaches with which the students will become familiar. Since the linear system that arises

from spatial discretizations will be sparse, we will introduce the students to iterative methods for solving these systems, in particular, Jacobi, Gauss-Seidel, optimal SOR, and conjugate gradients. The students will then compare the relative efficiency of the four numerical solution approaches. A parallel computing component will then be introduced by way of parallelizing the matrix-vector multiplications and vector-vector additions intrinsic to these iterative schemes. The students will then have an opportunity to compare the relative efficiencies of the four iterative schemes in a parallel environment. Finally, the project is concluded by coming back to the original question of how the development of patterns on a surface may be modeled, and the students will engage in a discussion of their results.

Module 2: Stochastic Models of Cell Growth: This module is based on stochastic models for growth, in particular we model the growth of bacteria on a petri dish. The diffusion of a drop of nutrient in the dish, and the ensuing pattern of bacterial growth can be explored with a random-walk model. If the students have already seen other deterministic models of diffusive systems (see, for example Module 1 on pattern formation), this module will clarify the relation between random and non-random models. In developing the mathematical model we will derive the random walk and its representation as a Markov Chain. The simulation of Markov Chain models will then be discussed, and in this context we will investigate numerical techniques for the multiplication of positive-definite matrices, as well as the computation of eigenvalues and eigenvectors. We will also derive the relationship between random walk Markov Chains and the normal distribution. This connection can be important in modeling, since powerful statistical methods can be used with normal distributions, enabling the modeler to estimate parameters and validate the model. Finally, we will introduce the idea of a general probability distribution and Monte Carlo methods for their simulation, including parallel implementations.

Module 3: Modeling Tumor-Immune Interactions: In this module we introduce the Tumor-Immune system interaction equations developed by Kuznetsov(1994). This model is of particular interest because the mathematics are straightforward, yet the dynamics of the system are quite rich. We will begin by developing each component of the differential equations from biological principals. We will then take the students through an elementary qualitative analysis of the system, finding critical points and stabilities. The computational component will then be introduced, and this will allow us to familiarize the students with simple explicit, implicit and specialized stiff numerical ODE solvers. The students will learn about the benefits and drawbacks of each category of solver, and will be guided through a basic stability analysis of each solver. Once we start running numerical experiments, we will have the students experiment with parameter changes, and discover the impacts on the dynamics of the system. The system we will be working with has several bifurcation points, equilibria appear and disappear and stability properties change. We will examine the numerical problems encountered in the analysis of these bifurcations by investigating methods of numerical continuation. Continuation is also important in the numerical detection of unstable equilibria and separatrices. We will then guide the students through a discussion of the biological interpretation of these parameter changes. For example, this model allows for an explanation of tumor dormancy, a phenomenon that is still not fully understood biologically.

Module 4: Optimizing Chemotherapy Protocols with Dynamic Programming and Genetic Algorithms: In treating cancer patients with chemotherapy, it is important to reduce the tumor size while keeping the normal cell and immune cell populations at healthy levels. In this module we will introduce optimal control to design cancer chemotherapy protocols. We begin by developing the three population model of cancer growth developed by de Pillis and Radunskaya

(2001) that describes the growth of a tumor cell population that is stimulating an immune response and competing with a surrounding normal cell population for resources. We will guide the students through a preliminary qualitative analysis of the system, and then generate some numerical solutions for certain parameter ranges. (Note: the specific numerics of various ODE solvers will be addressed in more detail in the module on Tumor-Immune interactions.) Once the students understand how the model is expected to behave in the absence of outside interference, we then introduce mathematical terms describing the administration of chemotherapy, which will detrimentally affect all three-cell populations at different rates. Now the question becomes: How can we optimally administer the chemotherapy in order to most rapidly reduce the tumor cell population without making the patient too ill? At this point we introduce the concept of optimal control. We will examine two numerical methods for solving the optimal control problem. The first is classical dynamic programming, systematized by Bellman (1952), that reformulates the problem to one that can be solved through recursive relations. The second approach involves genetic algorithms, which treats each potential protocol as an individual in a population that evolves according to a set of fitness criteria. The most fit candidates are sifted out of a large search space. This algorithm is ideally suited to parallelization, since it is a straightforward exercise to divide a search space up among several computing processors. Additionally, this algorithm is fairly time consuming on a traditional serial processor, and is likely to benefit from nearly linear speed-up when parallelized. We complete this module by having the students compare the solutions found by their optimal control programs to the solutions they would get when administering traditional pulsed chemotherapy. This will lead to a discussion of the potential benefits and drawbacks of implementing the optimal control protocols.

Module 5: Using Fourier Transforms to Understand Heart Conditions: Experienced cardiologists will actually listen to an electrocardiogram to gain insight into the condition of a patient's heart. When the cardiologist actively listens, she is picking up on distinct frequencies of the beating heart. These frequency patterns can reveal important information about the health of the patient's heart. This procedure can be automated, using data from ecocardiograms. In this module, we will take actual data from the ecocardiograms of over 100 patients with various types of heart conditions, and show how the Fourier Transform on these data can be used to help classify these conditions. The students will learn the mathematics behind the Fourier Transform, and will then be introduced numerical approaches for computing the transform of a given time series. We will then acquaint the students with the Fast Fourier Transform (FFT), which is an ideal candidate for parallelization. The students will develop their own FFT programs in a parallel environment using MPI (Message Passing Interface). Finally, we will position this problem in the broader context of general time series, and investigate the period doubling route to chaos commonly observed in many physical systems. For example, it is believed that the phase space trajectories generated by a normal heart would lie on a strange attractor.

Co-PI: Eric Grosfils

Institution: Pomona College

Discipline: Geology

Module 1: Thermal Conduction: Thermal conduction is a fundamental physical process, one which controls many aspects of the volcanic and tectonic evolution of bodies within our solar system. Using transmission of thermal energy through the crust of the Earth as an

initial, physically intuitive conceptual model, the module's background material will (a) help students deduce the thermal conduction equation—a second order differential which can be constructed from first principles, (b) evaluate volume-adjusted conduction incorporating internal heat generation and temperature change, and (c) explore special forms of the equation such as steady state conduction and thermal diffusion. Analytical solutions of the problem for a semi-infinite half-space require introduction of an error function and yield direct insight into useful physical concepts like the characteristic thermal diffusion distance; solution illustrations will include calculation of a thermal boundary layer thickness and surface heat flow. Problems to be tackled by the students could include, for example, (a) assessing whether conduction alone is responsible for cooling the Earth, an approach which yields an age vastly at odds with known values and which leads students to recognize the importance of radioactive heating, (b) constraining how long magma flowed through a conduit to produce an observable, temperature-dependent thermal alteration zone, (c) assessing the isotherm which best defines the base of the cooling oceanic crust as it moves away from a mid-ocean ridge spreading center, and (d) using thermal wave propagation to constrain the depth to which subsurface ice on Mars can be melted solely by daily and seasonal atmospheric temperature variations. As an advanced topic, to be developed if time permits, students will be introduced to time-dependent solutions for temperature within host rock adjacent to a magma body using Fast Fourier Transform methods. All problems will be solved using interactive graphical techniques in Excel. Because of the broad applicability of these equations and problems within a geological context, extra time will be spent to develop separate versions of the module suitable for use in introductory as well as more advanced geology courses; only some elements will translate well into an introductory setting. For the more advanced/complete version of the module, I will test it at Pomona College in a course with minimal lecture (Introduction to Geomathematics), and thus as part of the final product I will plan to provide materials that help instructors lead their students through exploration of thermal conduction in an interactive, hands-on exploration mode.

Module 2: Volcanic Ballistic Trajectories: Under the right conditions volcanic eruptions and impact cratering events can scatter large quantities of rock into the surrounding countryside along ballistic trajectories. In this module students will explore the ballistic process using both forward and inverse modeling techniques. Invoking an eruption from a small volcano and ignoring atmospheric drag, the module's initial background material will help students explore (a) how to derive and employ standard ballistic equations, (b) the relationships between ejection angle, ejection velocity, eruption energy and distance traveled in the horizontal plane, and (c) how variations in gravity from one planet to the next can affect volcano shape and size. Building on these initial investigations, students will then introduce atmospheric drag into the problem and explore the implications for distance, velocity and energy; even for linear drag the resulting equations, though simple in form, do not have a straightforward analytical solution. Problems to be tackled by the students could include, for example, (a) predicting how the size of ballistically emplaced particles will vary as a function of distance from an impact crater, (b) exploring the eruption conditions required to emplace a large block found some distance from a volcanic vent with and without consideration of atmospheric drag, (c) performing a ballistic hazards assessment for a farmer living near a volcano, and (d) assessing the distance at which curvature of the Earth becomes significant in a ballistic calculation. As an advanced topic, to be developed if time permits, students will explore the tektite problem – how can small blebs of impact-derived glass be emplaced more than a thousand miles from the crater where they originated? All problems will be solved using interactive graphical techniques in Excel. I will

test the module at Pomona College in a course with minimal lecture (Introduction to Geomathematics), and thus as part of the final product I will plan to provide materials that help instructors lead their students through the ballistic emplacement module in an interactive, hands-on exploration mode. In addition, because some components of the module require less mathematical or geological sophistication than others, I will design the no-drag portions of the module to serve in a stand-alone format suitable for use in an introductory geology course.

Co-PI: Andrea M. Karkowski
Institution: Capital University
Discipline: Psychology

Module 1: Modeling Temporal Aspects of Behavior. Although we are relatively unaware of it, humans and other animals possess a keen sensitivity to timing. Precision timing is most evident in individuals who have honed this skill, such as musicians and athletes; however, even simple everyday tasks like driving and eating require a temporal organization of behavior. When our timing system is disrupted due to brain injury, illness, or attentional deficits, the results can range from minor embarrassment to serious injury or death. In this module, students will explore the Time Derivative Model of conditioned responding. Establishing a conditioned response requires associative learning, thus this model can serve as a natural extension of a module created for the NSF CCLI grant (DUE: 9952806), Modeling Associative Learning Using the Rescorla-Wagner Model. The Time Derivative Model will be explored from both behavioral and neural network representations.

Module 2: Modeling Scheduled Reinforcement Contingencies of Behavior. Most behavioral scientists treat typical schedules of reinforcement, such as Variable Interval and Variable Ratio schedules, as separate and discrete entities. While this approach may facilitate the development of an understanding of such schedules, both in the classroom and in the laboratory, it belies the true character of scheduled reinforcement as it exists in the natural environment. That is, few behaviors are mediated by a pure interval or ratio schedule. To alleviate this discrepancy, Berger (1988) established a continuum of behavioral-temporal reinforcer contingencies, the Interactive Schedule: $f_o = F_b^x/C$ where f_o is the instantaneous frequency of reinforcement, F_b is the mean response rate since the last reinforcer, x identifies the point on the continuum that the organism is experiencing, and C is the set of contingencies established for the value of x . Using the Interactive Schedule, students will explore the resulting relationship among responses, time, and reinforcement along the continuum.

Co-PI: Terry Lahm
Institution: Capital University
Discipline: Environmental Science

Module 1: Spatial Data Analysis in Environmental Science: The need to interpolate between known data values is common in many areas of Environmental Science such as spatial distribution of geologic, ecologic, and atmospheric properties and in cartography (Davis, 1986). We will examine a variety of methods of interpolation and visualization of spatial data. Data will be field-based examples of topography analysis, atmospheric temperature and pressure analysis, and geologic material analysis. Use of visualization and computational skills, such as image rastering data in two and three dimensions and matrix algebra techniques from prior coursework

will be used to produce two and three-dimensional representations of numerical results. We will also make comparisons to the “ground true” data in each of the three example data sets. The module goals are to (i) understand spatial statistical analysis as it applies to interpolating spatial data, and to (ii) understand the differences in each interpolation method as they apply to the three data sets. Students will also understand the data needs for spatial analysis problems by comparison to real data sets.

Module 2: Watershed Data Analysis and Visualization: This module will examine the computational methods available to represent the static and dynamic data concerning watershed models. Geographic information systems (GIS) allow the manipulation and analysis of large-scale problems within the Environmental Sciences. We will examine specifically how GIS can be used to visualize environmental data at this large scale and how computational methods are used to address common environmental problems. The goals for this module are to (i) understand problem and solution techniques employed for watershed modeling, (ii) understand the use of GIS software to visualize different types of environmental data, (iii) develop a working knowledge of the integration GIS and other types of computational models of the environment.

Module 3: Extension of Groundwater Flow Modeling: This module will be an extension of a previous model of the glaciofluvial aquifer system within the Aberjona River valley in Woburn (created for NSF CCLI 9952806). We will use public domain software (MODFLOW and MODPATH) to extend our understanding of contaminant transport at this US EPA Superfund site. Pathline analysis will be used to visualize the movement of groundwater through the subsurface geologic material. The module will be designed for those institutions that would like to examine the numerical model of groundwater flow in greater detail. The goals for this module are to (i) understand the sensitivity of physical and chemical controls on contaminant transport with the groundwater environment, (ii) develop a working knowledge of how computational methods can be used to predict the movement of groundwater and contaminants, and (iii) understand the visualization techniques used with the hydrologic sciences to understand contaminant transport.

Co-PI: Robert Lawson

Institution: Capital University

Discipline: Finance

Module I: Cash Flow Analysis: In the first portion of this module will analyze Cash Flow Analysis, employing NPV and IRR concepts for valuation. Cash Flow data will be used to generate forecasting of a project’s profitability when compared by means of NPV and IRR applications. The concepts of Present Value and Future Value computation will be acquired prior to the Cash Flow analysis. The second portion of this module will emphasize evaluation, generation, and interpretation of the Capital Asset Pricing Model (CAPM). To facilitate financial analysis and valuation of risky assets based on the model it will be necessary to introduce the concept of return-risk analysis and linear regression analysis prior to its application. Data will be used to generate the returns and a linear regression analysis will determine the risk valuator benchmark to be used in the analysis. Applications of the model will be assessed, and possible numerical adjustment and manipulation to the benchmark will be discussed.

Module II: Option Pricing: This module will analyze the topic of option pricing. The Black-Scholes formula will be used to price call options on an underlying asset. The put call parity theorem will be used to value put options. Background information on the model will be

provided as well as the theoretical implications of the model. Assessment of the model will be evaluated. The module will also try to evaluate the differences between European and American options using mathematical models. The second portion of the module will examine derivatives and expansions of the model, like adding dividends or adapting the model for foreign exchange option valuation.

Co-PI: Raghu Machiraju

Institution: The Ohio State University

Discipline: Computer Science

Module 1: Volume Visualization: Volume visualization will be presented as a way to post-process simulation data and, if possible, to monitor an evolving simulation. The thrust of the modules is to convey that visualization is more than a way to generate pretty pictures of scientific phenomenon. The module will stress on the analysis prowess of visualization algorithms in addition to the graphics rendering techniques that are required to create meaningful visualizations. For scalar fields, iso-contouring in 2D and 3D will be illustrated through known examples of topographic relief plots of terrains and an iso-surface extraction of features from well calibrated CT medical data set respectively. Vector field visualization will exploit local iconic techniques, streamline and path-line tracing and application-specific feature extraction (e.g., vortices, shocks in fluid dynamics simulations). Emphasis will be accorded to comparative advantages of each technique. A 2D Rankine vortex example will be used to underscore the differences. Volume rendering through the use of optical models and transfer functions will be used to underscore the ability to allow for multiple surfaces to be juxtaposed in natural visibility order. A well-calibrated CT data set will be again used to drive home the need to design meaningful transfer functions. The modules will use VTK and OpenGL. Homework and laboratory assignments will include exercises to locate features and structures in data sets. Not much will be revealed in terms of the actual nature of the data sets. The students will be required to conduct exploratory exercises to glean the important structures. Another set of assignments will be used to impress upon the students the need to reduce the computational burden of the rendering algorithms. Simplification and multi-resolution techniques will be deployed to reduce the number of primitives rendered. The module will really be sub-divided into various sub-modules that call several VTK C++ routines. Additional material in terms of assessment and evaluation tools will be made available under the auspices of this grant.

Module 2: Imaging Pipeline: The imaging pipeline will be presented as data-preparation step for visualization and analysis. 2D images and 3D data sets will be considered for noise removal, enhancement, edge detection, morphological operations and transform coding. Emphasis will be placed on the use of linear shift-invariant convolution operations to achieve many of the above-stated operations. Low-, band- and high-pass spatial filtering operations will be included for considerations. The frequency space will be used to illustrate the efficacy of these operations. Thus, suitable treatment of Fourier transforms will be provided in appropriate sub-modules. Wavelets for multi-resolution and multi-scale processing will be employed. Finally, the efficacy of various morphological operators (e.g., dilation) for feature preservation will be illustrated. Laboratory assignments will include the deployment of spatial filtering, Fourier methods and wavelet transforms for noise removal and image enhancements. The computational and accuracy tradeoffs of filter choices will be examined for a set of well-known images and a variety of operations. Both Fourier and wavelet transforms will also be examined

for comparative assessment for a class of images with sharp transient features. Scalar quantization schemes will be used. Finally, a laboratory assignment will include the de-cluttering of an image replete with shapes of various sizes. The assignments and modules will use VTK routines for many of the image processing functionalities. The wavelet transform sub-modules will use an enhanced version of *QCCPack* which is under development under the supervision of the co-PI and his collaborators at Mississippi State University. This C++ API is being developed under the auspices of a grant from the National Science Foundation and will be released as open-source software.

Co-PI: David Reed
Institution: Capital University
Discipline: Computer Science

Module 1: Object-Order Projection Visualization: The modules for the National Science Foundation Grant (DUE 9952806) are being developed using the Visualization Toolkit (VTK) to allow students to produce visualizations without having to write all the low level visualization code themselves. By using the VTK, the students will be able to concentrate on the concepts rather than the details; however VTK does not support a number of visualization methods. In this module, suitable for a final project, students will write their own visualization software (without using VTK) and learn many of the intricacies of writing graphics software that VTK hides. For this module, students will implement an object-order projection visualization technique such as splatting. The design of VTK does not allow object-order projection visualization techniques to correctly handle transparency. Splatting is one such technique that can correctly handle transparency and is of a reasonable level of difficulty for advanced undergraduate students when limited to parallel projection methods. The background material for this module will present the concepts of splatting, along with the additional problems that perspective projections present for splatting. The students will implement a parallel projection splatting method without using the VTK software.

Co-PI: Linda Reinen
Institution: Pomona College
Discipline: Geology

Module 1: Friction and Faulting: Tectonic forces produce a wide variety of observable geologic phenomenon, including earthquakes and faulting. In this module, students will explore the application of classical Newtonian mechanics to the shallow layers of the Earth in order to understand the formation and movement of faults. Background information provided with this model will include information on Newtonian mechanics, rock rheology, plate tectonics, Mohr diagrams and Coulomb failure. Students will explore the interrelations between principal, normal and shear stresses; the angle between the fault plane and the plane normal to the applied force; and the coefficient of friction of the fault. Students will use the model they develop to determine why Anderson's hypothesis for fault formation generally applies to natural systems. Students will test their models by determining the frictional strength of a natural fault from data collected from the southern segment of the San Andreas Fault. Possible advanced topics, if time permits, include: (a) the effects of finite displacement on the strength and orientation of normal faults, (b) models of fault strength with a time-dependent term, and (c) slip-rate dependent models of

dynamic friction. All problems will be solved using interactive graphical techniques in Excel. Some of the advanced topics may require the use of another package, such as STELLA. I will test the module at Pomona College in the laboratory section of the Structural Geology course.

Module 2: The Influence of Mechanical Layering in Rock Deformation: Under an applied stress, rocks can deform via localized deformation (e.g. faulting) or distributed deformation (e.g. distortion, strain). How does mechanical layering influence this deformation? In this module, students learn about how the physical properties of different rock types influence the mechanical behavior observed during deformation. Background information provided with this module will include information on rock rheology, failure modes (localized vs. distributed deformation), and strain styles (pure and simple shear). Initially, students will develop models of a homogeneous, isotropic medium to explore the depth distribution of stress under different tectonic regimes, and rock deformation by pure and simple shear. Students will then develop models exploring the influence of mechanical layering on fold geometries (distributed deformation), strain distribution within folds, and jointing (localized deformation). Students will test their models by determining the strain recorded in naturally deformed rocks. If time permits, students will develop models to investigate the influence of mechanical layering on cleavage orientation and refraction. All problems will be solved using commonly available software such as Excel or Adobe Illustrator. I will test the module at Pomona College in the laboratory section of the Structural Geology course.

Co-PI: Karl Romstedt

Institution: Capital University

Discipline: Biology

Module 1: Mathematics in Neurophysiology: Most students have a real interest in understanding how their bodies work and in seeing medical consequences for pathologies resulting from impaired function. This can also be extended to show how pharmacological therapies impact physiological systems. In this module, the function of neurons will be modeled using electronic circuitry. This means that neuronal activity can be used to show the intersection of a variety of disciplines including mathematics, biology, physics, chemistry, physiology, pathology and pharmacology. Chemical and physical models for neuronal activity are especially amenable to mathematical analysis. This includes equations for RC circuits and the Nernst equation that quantifies the effect of ion distribution on membrane voltage. The Goldman equation shows the effect of membrane channels on the voltage and also models the events that lead to a nerve action potential. This module will emphasize student-centered discovery and discussion of models for neuronal behavior. The study of neurophysiology will be illustrated with hands-on experiments using simple electronic components. It will also be supplemented with animated software including ADAM interactive physiology that shows the behavior of the molecules involved. Virtual physiology laboratories will employ PhysioEX software from Benjamin Cummings.

Module 2: Diffusion Across Cell Membranes: Cell membranes are essential to life because they compartmentalize the basic molecules involved into interactive communities that carry out life functions. Cell membranes must be selectively semipermeable to allow necessary entry of required molecules and to permit the exit of unwanted, toxic materials that accrue during normal metabolic activities within the cell. This module will explore the structure of membranes and show how individual components contribute to its function. The functions will be modeled

mathematically. Models will include equations for diffusion and membrane flux for solutes. Osmotic properties of cells will also be discussed and illustrated using mathematical models. Students will be required to design laboratory experiments and use empirical data to derive equations that describe the movement of molecules in model systems. For example, models for diffusion rates may be derived by observing movement of colored molecules under different conditions in an agarose gel. Students will plot potential models along with their data and determine which has the best fit statistically.

Module 3: Gas Exchange in Living Systems: This module will demonstrate the process of gas transfer in living systems. It will emphasize lung function but exchange involving cutaneous transfer and gill function will also be included for comparison. The principle gases to be studied will be oxygen and carbon dioxide. Basic gas laws (Boyle, Dalton and Henry) will be reviewed and mathematical models will demonstrate the effect of surface area, permeability and concentration on the exchange rate across respiratory membranes. The relationship between blood pH, bicarbonate ion concentration and pH will be demonstrated as well as the effect of ventilation and hyperventilation on blood pH. Oxygen dissociation curves will be used to show subunit cooperativity of hemoglobin protein and demonstrate the Bohr Effect caused by changes of pH on oxygen delivery in tissues. The relationship between various lung volumes (tidal, expiratory reserve, inspiratory reserve, vital capacity, etc.) will be explored using spirometer measurements obtained from the students themselves.

Co-PI: Patrick Shields

Institution: Capital University

Discipline: Physics

Module 1: Atomic Structure of Single-Electron Elements: The structure of the hydrogen atom is a staple of introductory quantum mechanics. Students are taught to separate and solve the radial and angular parts of the Schrödinger equation. Computer algebra systems offer the opportunity to visualize these solutions in three dimensions and develop a more intuitive feel for the structure. Once the student has explored the hydrogen atom it will be possible to extend the exploration to the structure of other single-electron elements. From there it is a natural step to inquire about the structure of multi-electron atoms, thereby introducing the Hartree-Fock approximation. This technique is an example of a self-consistent field approximation and represents a starting point for many of the computational problems in the structure of atoms and molecules.

Exploring the hydrogen atom wave functions is intended to reinforce and expand the student's ability to analyze and interpret graphical information. The radial equation will produce the usual planar graph, but graphing the angular equation will introduce the student to the power of three-dimensional visualization. Use of these representations will be reinforced by the examination of other single-electron elements. The Hartree-Fock approximation will introduce the student to more advanced issues involved in solving eigenvalue problems and some of the methods of matrix algebra, including matrix transformation and diagonalization.

Module 2: Electrostatic Potentials using the Laplace Equation: Electrostatic problems translate into elliptical partial differential equations. The interconnections between the different spatial dimensions make solving these types of problems much more challenging than many physics problems, such as projectile motion, that result in ordinary differential equations. This module will examine the solution of Laplace's equation for a variety of geometries with

appropriate boundary conditions. The initial method of solution will use finite differences and the usual relaxation techniques. Subsequent to that the student will be introduced to the expression of these equations as sparse matrices that can then be solved via matrix algebra.

Module 3: Diffusion-limited Aggregation: The structure of a fern or a snowflake can be modeled by a technique known as diffusion-limited aggregation (DLA), a form of cellular automata. Implementation of the model begins with the construction of a two or three-dimensional lattice upon which the “snowflake” can grow. A “seed” particle is planted, typically at the center of the lattice. Subsequent particles are introduced from the boundary of the lattice one at a time. These particles are then allowed to diffuse through the lattice in the manner of a random walk. If they reach a lattice boundary they disappear. However, if they reach a lattice site with an occupied neighboring site, then they “stick” to that site and the structure grows. While fairly simple in design, the DLA model introduces several useful modeling concepts. Whether addressing crystal structure, electronic behavior, or magnetism, the idea of a lattice is fundamental in condensed matter physics. If the student has no prior experience with the concept, then this model offers an opportunity to introduce the lattice and connect it with the concept of an array or matrix. Modeling the random-walk opens the door for the student to explore the use of random number generation as a tool. A vast collection of techniques, generally referred to as Monte Carlo (MC) simulations, depend upon this method and have been applied to such diverse areas as traffic flow, radioactive decay, and numerical integration. Finally, the structure that results from the simulation will be a fractal. This affords the opportunity to introduce the student to the concept of fractal dimension and to explore the methods by which appropriate measures of fractal behavior can be computed.

Co-PI: Angela B. Shiflet

Institution: Wofford College

Discipline: Mathematics/ Computer Science

Module 1: Simulation of Animal Behavior in Searching for Food: According to research by Dr. Aliston K. Reid, one can simulate an animal's search for food using a diffusion process similar to the diffusion of heat ["A Dynamic Route Finder for the Cognitive Map," Aliston K. Reid, *Psychological Review*, 1998, Vol. 105, No. 2, pp. 585-601]. In this module, we will discuss the theory and an algorithm for performing the simulation. The simulation can be developed in a programming language or with a spreadsheet. Discussion questions will help the students to compare simulated results to actual data and to consider the impact of different parameter values.

Module 2: Modeling Blood Cell Population: Various diseases have an impact on white blood cells, red blood cells, and platelets ["A Blood Cell Population Model, Dynamical Diseases, and Chaos," William B. Gearhart and Mario Martelli, UMAP Module 709]. This module will discuss various models for normal and abnormal blood cell growth, as occurs in such diseases as hemolytic anemia and chronic myelogenous leukemia. The student will develop systems models using *STELLA* and simulations using a programming language. Using the models, the student will be asked to draw conclusions about the progress of the diseases.

Module 3: Tomography: Computed Tomography (CT) scanners give three-dimensional images of patients. However, raw data from the scanner is virtually useless. The data is used to create a model of the scanned body. This module will discuss how to develop such models using mathematics and *Mathematica*.

Co-PI: Edward J. Soares

Institution: College of the Holy Cross

Discipline: Mathematics/ Computer Science

Module 1: Image Reconstruction in Emission Tomography: Direct Inversion:

Single-photon emission computed tomography (SPECT) is a non-invasive imaging modality that yields information regarding the function of organs within the body. This information is in the form of pixilated images that have been reconstructed from projection measurements at various angular positions around the patient. The image, which represents the bio-distribution of radio-pharmaceuticals that have been administered to the patient, can be used as a diagnostic tool. Since cancerous or damaged tissue can have a different metabolic rate compared with healthy tissue, anomalies exhibit themselves as exceptionally dark or bright regions within the image. An important problem in SPECT is the reconstruction problem: How do we obtain the image from the projection data measurements? The most accurate model for the data acquisition process is a continuous-to-discrete transformation, since patients are continuous objects, whereas the data are measured discretely. However, to reduce the complexity of the model, the object and data are both discretized into pixels (picture elements) and regarded as vectors \mathbf{x} and \mathbf{b} , respectively. The data acquisition process can then be modeled as a matrix transformation \mathbf{A} , where the ij th element of the matrix describes how object pixel j contributes to detector measurement i . The reconstruction problem now reduces to solving a linear system of equations, $\mathbf{Ax} = \mathbf{b}$. There are a few outstanding parameters that must be set, such as the number of object and data pixels, which may be unequal, and the number and orientation of the detector acquisition angles. The inversion methods that this module will focus on are so-called “Direct Methods”, meaning 1) Gaussian Elimination/Back-substitution, 2) Direct Matrix Inversion, and 3) Singular Value Decomposition/Pseudo-inversion. Each are standard techniques to solve linear systems of equations from matrix theory, and have both theoretical and computational advantages and disadvantages. For example, Direct Matrix Inversion is not possible if the number of object and data pixels are not equal or if the matrix is singular (theoretical constraints). However, even if they are equal and the matrix is non-singular, Direct Matrix Inversion may not be practical from a numerical standpoint, as round-off errors in determining the matrix inverse would effectively make the matrix singular. The most elegant way to implement the solution methodologies would be a computer algebra system such as MAPLE or MATHEMATICA, or a computational tool such as MATLAB. The dimensionality of the problem would play a crucial role in determining the platform. For example, to demonstrate the modeling and computational aspects of the module, a simple 9×9 example would suffice, and would best be exhibited using MAPLE or MATHEMATICA. However, if one used a 4096×4096 example, students could better visualize the reconstruction as a gray scale 64×64 image and thus MATLAB would be a better choice. One assignment would be to supply the students with data measurements and have them determine the object that produced the data. The obvious challenge here is that they don’t know the correct answer (and there may be several “correct” answers). First, they need to know the acquisition positions of the detector and use simple geometry to determine the elements of \mathbf{A} . Next, they would use the appropriate software to define \mathbf{A} and \mathbf{b} , and then implement the solution methodologies to find a solution \mathbf{x} . Subsequently, the students would try to discover if indeed this was the exact solution to the linear system, or if it was an approximation (and what type of approximation).

Module 2: Principal Component Analysis of Satellite Imagery: Remote Sensing uses satellites as a means of gathering information about the Earth's environment. Low-orbiting satellites perform multi-spectral imaging of the Earth's surface, which yields gray-scale or pseudo-color reflectance of surface features for a variety of spectral bands in the electromagnetic (EM) spectrum. These images can then be used, for example, for classification purposes or to detect the effects of acid rain such as defoliation. Although several spectral bands are commonly used, the multi-band information can be redundant, due to natural spectral correlation in neighboring EM bands, topographic slope and aspect, and overlap of spectral sensitivities between adjacent EM bands. Therefore, some bands contain redundant information. The problem is to analyze the information in the most efficient manner possible and to eliminate data redundancy. A typical satellite image is a discrete array of pixels (picture elements). Each pixel in the image represents the gray-level reflectance of that particular surface feature within the particular spectral band. Assuming there are K spectral bands, each image pixel can be regarded as a K -dimensional feature vector \mathbf{x} , whose j th component describes the reflectance properties of that surface feature in spectral band j . Thus, we have a set of K -dimensional feature vectors obtained from all the pixels in the satellite images. In order to decrease the redundancy, we form a new set of K -dimensional feature vectors, where each feature component is a linear combination of the original K features. This is accomplished via a linear transformation $\mathbf{x} = \mathbf{W}\mathbf{x}$, where the ij th element of the matrix \mathbf{W} represents the contribution of original feature j to new feature i . The idea is to choose \mathbf{W} so that the only a subset of the components of \mathbf{x} contain pertinent information. This reduction in dimensionality would then allow one to use a subset of the spectral images for purposes of classification or terrain analysis. The Principal Components (or Karhunen-Loeve) transformation provides a matrix \mathbf{W} that de-correlates the features with respect to the various spectral bands. Since the i th row of the matrix corresponds to the i th eigenvector of the covariance matrix for the original set of feature vectors \mathbf{x} , \mathbf{W} will diagonalize, and so the covariance matrix for the new set of feature vectors \mathbf{x} will be diagonal, with decreasing feature variance along the diagonal. Thus, we only need to retain the first few principal components. The most elegant way to implement the solution methodology would be a computational tool such as MATLAB. Because the dimensionality of the multi-spectral images tends to be large, for visualization purposes, any platform with good image display capabilities would allow the user to better see the original and processed images. Using LANDSAT, Thematic Mapper and SPOT images from NASA (or other suitable site), one would create the feature vectors from the set of multi-spectral images. Then, the covariance matrix for the feature vectors would be computed, as well as the eigenvalues and eigenvectors for the covariance matrix. One would then apply the Principal Components transformation to the original set of feature vectors to create a new set. By analyzing the eigenvalues, one would be able to determine how many principal components to keep. After processing each feature vector, one would reconstitute the multi-spectral images, and then use them. A nice application would be to look for areas of defoliation in the Northeast or Canada in order to determine the impact of acid rain in industrial areas.

Module 3: Image Reconstruction in Emission Tomography: Iterative Inversion: Single-photon emission computed tomography (SPECT) is a non-invasive imaging modality that yields information regarding the function of organs within the body. This information is in the form of pixilated images that have been reconstructed from projection measurements at various angular positions around the patient. The image, which represents the bio-distribution of radio-pharmaceuticals that have been administered to the patient, can be used as a diagnostic tool.

Since cancerous or damaged tissue can have a different metabolic rate compared with healthy tissue, anomalies exhibit themselves as exceptionally dark or bright regions within the image. An important problem in SPECT is the reconstruction problem: How do we obtain the image from the projection data measurements? The most accurate model for the data acquisition process is a continuous-to-discrete transformation, since patients are continuous objects, whereas the data are measured discretely. However, to reduce the complexity of the model, the object and data are both discretized into pixels (picture elements) and regarded as vectors \mathbf{x} and \mathbf{b} , respectively. The data acquisition process can then be modeled as a matrix transformation \mathbf{A} , where the ij th element of the matrix describes how object pixel j contributes to detector measurement i . The reconstruction problem now reduces to solving a linear system of equations, $\mathbf{Ax} = \mathbf{b}$. There are a few outstanding parameters that must be set, such as the number of object and data pixels, which may be unequal, and the number and orientation of the detector acquisition angles. In contrast to the “I. Direct Inversion” module, this module will focus on “Iterative Methods” of the general form $\mathbf{x}^{(k)} = \mathbf{T}\mathbf{x}^{(k-1)} + \mathbf{c}$, where $\mathbf{x}^{(k)}$ is the estimate of the exact solution \mathbf{x} at the k th iteration. Some methods that fit within this framework are the 1) Jacobi, 2) Gauss-Seidel, and 3) Landweber algorithms. The first two are standard techniques to solve linear systems of equations from numerical linear algebra. Two of the main points that need to be addressed are 1) if the method converges and, if so, 2) how fast it converges. Both can be determined examining the spectral radius (magnitude of the largest eigenvalue) of the matrix \mathbf{T} . The most elegant way to implement the solution methodologies would be a computer algebra system such as MAPLE or MATHEMATICA, or a computational tool such as MATLAB. The dimensionality of the problem would play a crucial role in determining the platform. For example, to demonstrate the modeling and computational aspects of the module, a simple 9×9 example would suffice, and would best be exhibited using MAPLE or MATHEMATICA. However, if one used a 4096×4096 example, students could better visualize the reconstruction as a gray scale 64×64 image and thus MATLAB would be a better choice. One assignment would be to supply the students with data measurements and have them estimate the object that produced the data. First, they would need to know the acquisition positions of the detector and use simple geometry to determine the elements of \mathbf{A} . Next, they would have to rewrite the linear system $\mathbf{Ax} = \mathbf{b}$ in the general iterative form to determine the matrix \mathbf{T} , and then implement the solution methodologies to iterate to an approximate solution $\mathbf{x}^{(k)}$. They would determine the stopping point k by deciding on a tolerance to be met by the relative error and then use the relative error as a measure of accuracy. Convergence and speed of convergence would also be determined by computing the spectral radius of the matrix \mathbf{T} .

Module 4: Processing of Images Corrupted by Noise and Its Relation to Signal

Detection: Often, images obtained for medical or scientific purposes are corrupted by noise. This is usually unavoidable, due to the nature of how the images are acquired. Noisy images are often processed (or filtered) in order to reduce the visual conspicuity of the noise and to increase the appearance of the signal. However, a trade-off between reducing noise and visually improving the signal exists, as filtering may have the effect of introducing noise correlations that visually impair the ability of the observer of the image to detect a signal, introducing a false signal, and reducing the contrast of a true signal to the background. All three confounding artifacts can appear in medical tomographic images. The problem here is to understand the tradeoffs that underlie image filtering and to find an optimal way to design a filter to order to minimize the introduction of artifacts. Each image is a discrete pixilated representation of the object (patient) and can be regarded as a random vector \mathbf{x} . The projection data \mathbf{b} from which the

image is constituted, can also be regarded as a random vector. The projection data are corrupted by noise inherent in the photon decay and collection process. The idea is to reduce the effects of the noise in the data before it is propagated into the reconstructed image (see **Image Reconstruction in Emission Tomography** modules for descriptions of reconstruction methods). Because the image is regarded as random vector, a set of noisy images of the same object will have an associated mean vector and covariance matrix. These serve as descriptive measures for the statistical properties of the reconstructed images. Before image reconstruction, the data are first filtered to reduce the noise and to correct for reconstruction artifacts arising from filtered-backprojection reconstruction. This filter can be designed to eliminate/retain high frequency components via selection of a cut-off frequency. After reconstructing a set of noisy images, the mean and covariance matrix can be computed. Also, a receiver-operating characteristic (ROC) study will be setup to allow us to determine the effectiveness of the image as a diagnostic tool. By creating both signal present and signal absent images, we can measure how effectively the signal can be seen. This can be accomplished for a variety of cut-off frequency settings for our reconstruction filter. The most elegant way to implement the solution methodologies would be a computational tool such as MATLAB or a programming language. The dimensionality of the problem requires this, as a reasonable minimum image size would be on the order of 64×64 or higher. Once again, the image display capabilities of the software is vitally important, particularly for the ROC studies. One assignment would be to supply the students will noise-free data measurements and have them add noise to create lots of noisy data sets. Both signal present and signal absent data would be made. Then, they would design their reconstruction filters by determining cut-off frequencies, and subsequently filter the data sets. Next, they would reconstruct the projection data, creating large sets of noisy images. The students would then analyze the images mathematically, by computing the statistical parameters such as mean vector and covariance matrix. In coincidence, they would conduct the ROC experiments and analyze the response data to determine which filter setting yielded optimal detection performance.

Co-PI: Kris Stewart

Institution: San Diego State University

Discipline: Computer Science

Module 1: Performance for Steady State Diffusion with LAPACK, Part I:

Characterize performance as: 1) Amount of work needed in the computation and 2) Amount of accuracy delivered by the numerical approximation. Learning through discovery is the overarching theme of this module. Students are assumed to have had calculus and some programming in C or Fortran. The module develops an appreciation for numerical approximation, but does not assume a formal background in numerical analysis. Students choose a scientific problem with known mathematical properties of work and accuracy and then devise computational experiments to discover which of those properties can be verified by computable metrics. Where to Place Timing Computations and How to Compute Accuracy Given Fortran 90 code developed to construct the matrix, A , and right-hand-side, b , corresponding to a discretisation of the rectangular grid using LAPACK subroutines. Students consider where to place CPU timing calls to isolate the cost of factoring the matrix and solving the resulting system. Students also consider the reproducibility of data over several runs, so that averages of performance can be discussed. Students decide how to measure error, by choosing an appropriate norm. Using system documentation, students discover the timers available on their computing

platform. Because this is a machine-dependent, implementation-dependent characteristic of each computing environment, this will be a useful skill for student to acquire. Once data are gathered from the experiment above, they can be analyzed to form a synthesis report. Written laboratory reports are used to assess the students' ability to, in turn, assess the performance.

Module 2: Performance for Steady State Diffusion with LAPACK, Part II: Building on first module which examines the natural, full matrix statement of the steady-state diffusion problem, this module takes into account the sparsity pattern of the 2D grid approximation and the anticipated savings in cost. Students choose a scientific problem with known mathematical properties of work and accuracy and then devise computational experiments to discover which of those properties can be verified by computable metrics. Again, building on the first module, students now construct only the non-zero entries corresponding to the grid approximation and modify the code to use the resulting banded structure. Using system documentation, students discover the timers available on their computing platform. Because this is a machine-dependent, implementation-dependent characteristic of each computing environment, this will be a useful skill for student to acquire. Once data are gathered from the experiment above, they can be analyzed to form a synthesis report. Written laboratory reports are used to assess the students' ability to, in turn, assess the performance.

Module 3: Performance for Steady State Diffusion with LAPACK: Building on first two modules which examine the full matrix and banded matrix statement of the steady-state diffusion problem, this module takes into account the work performance enhancement using the manufacturers optimized LAPACK routines. Choose a scientific problem with known mathematical properties of work and accuracy and then devise computational experiments to discover which of those properties can be verified by computable metrics. Again, building on the first module, students now construct only the non-zero entries corresponding to the grid approximation and modify the code to use the resulting banded structure. Using system documentation, students discover the linear algebra libraries available on their computing platform. Because this is a machine-dependent, implementation-dependent characteristic of each computing environment, this will be a useful skill for student to acquire. Once data are gathered from the experiment above, it can be analyzed to form a synthesis report. Written laboratory reports are used to assess the students' ability to, in turn, assess the performance.

Co-PI: Kathryn Thorbjarnarson

Institution: San Diego State University

Discipline: Geology

Module 1: Flood Prediction: Flood insurance programs are based on prediction of the 100-yr flood plain. The recurrence interval of stream flows is based on probability analysis of the historic record at a measuring station. The cumulative frequency distribution of measured annual peak flows and a theoretical probability distribution (lognormal, Gumbel Type I extreme-value, Gumbel Type II extreme value or Pearson Type III) are used for prediction of extreme flows. As historic stream flow records are typically short (less than 50 years), the impacts of long vs. short historic records and the type of assumed probability distribution on flood prediction will be evaluated. Confidence intervals will be calculated to assess ranges in the predicted 100-yr flood estimate. Student groups will adopt a stream location with historic records on the USGS web site and conduct a flood prediction study. Visualization of the floodplain will be conducted by

plotting the topographic profile of the streambed and superimposing the stream stages associated with various stream flows.

Module 2: Environmental Pollution: Hundreds of differing chemicals can be released into the environment as potential pollutants. Students will examine the distribution of differing chemicals in the environment using equilibrium principles and a model environment. The model environment will consist of soil, water, air and pure chemical phases. In the lower level classes, distribution of chemicals within these phases will be calculated assuming instantaneous equilibrium conditions with associated algebraic equations. The model environment and solution of the equations will be completed with STELLA. In higher level classes, transfer rates between phases and in or out of the model environment can be added to evaluate more complex scenarios. After reviewing examples, students or groups of students will be assigned an environment and chemical. In this inquiry-based assignment, the students must report back to the class the distribution of the chemical (primarily in soil, water, groundwater or air) and the concentration levels. The concentration levels will be compared to maximum contaminant levels set by environmental regulators. The volume of a phase or compartment assumed to be in equilibrium has a large impact on this type of modeling. The limitations of equilibrium compartmental modeling will be assessed by varying the compartment size to see the subsequent change in predicted concentrations.

Module 3: Pollutant Transport in Groundwater: Chemicals in groundwater can be transported to water-supply wells or surface water bodies. Humans exposed to low concentrations of certain chemicals over long time periods can develop health problems. Prediction of pollutant transport in groundwater is important for evaluating travel times from pollution sites to points of exposure (wells, lakes, etc) and for evaluating clean up times for polluted sites by pumping out groundwater. The one-dimensional advection-dispersion equation governing pollutant transport will be solved by analytical and numerical methods (finite difference). Analytical solutions of one-dimensional advection and two-dimensional dispersion will also be evaluated. Simplifying assumptions and constrained initial and boundary conditions and their associated limitations for analytical modeling will be assessed. In finite-difference modeling, the effects of spatial and temporal discretization will be assessed illustrating potential numerical dispersion and oscillations for large temporal and spatial grids. After numerous examples, student groups will be assigned various pollution scenarios for modeling. Scenarios to be included will be the Borden Aquifer, Canada (a controlled tracer test) and Hinkley, CA (site famous from Erin Brockovich movie). Due to the extensive amount of material in this module, subsections will be created for instructors to select from for use in a class. Each subsection will be independent and can be used in differing combinations for low or high level classes.

Co-PI: Michael Torello

Institution: Capital University

Discipline: Psychology/ Neuroscience

Module 1: Artificial Neural Networks: This module introduces the student to the mathematical foundation, biological foundation, the structure and the function of artificial neural networks (ANNs). There are a wide variety of approaches to the construction of artificial neural networks including the multi-layer perceptron, learning vector quantization, and the Hopfield network. Some ANNs are classified as feedforward while others are recurrent (i.e., feedback) depending on how data are processed through the network. ANN types are classified by their

method of learning (or training), and some ANNs employ supervised training while others are referred to as unsupervised or self-organizing. The advantage of ANNs lies in their resistance to input data distortions and their capability of learning. Neural networks are unlike artificial intelligence software in that they are trained to learn relationships in the data they have been given. Just like a child learns the difference between a chair and a table by being shown examples, a neural net learns by being given a training set. Due to its complex, non-linear structure, the ANN can find relationships in data where humans can't. ANNs are collections of mathematical models that emulate some of the observed physical properties of the nervous system and draw on the analogies of adaptive learning in the nervous system. An ANN is composed of a large number of interconnected processing elements that are analogous to neurons that are tied together with weighted connections that are analogous to synapses. Learning in the nervous system involves adjustments to synaptic connections between neurons. This is true of ANNs as well. Learning typically occurs by example through training, or exposure to a truth set of input/output data where the training algorithm iteratively adjusts the connection weights (synapses). These connection weights "store the knowledge" necessary to solve specific problems.

Module 2: Neural Networks: Applications in the Behavioral Sciences: Although ANNs have been around since the late 1950's, it wasn't until the mid-1980's that algorithms became sophisticated enough for applications in the behavioral sciences. Neural networks have advanced to the point where we can now attempt to model complex behaviors. Today, ANNs are being applied to an increasing number of complex, real-world problems such as visual pattern recognition, speech perception, and the manifestation of psychopathology. They are useful at solving problems that are too complex for conventional methodologies. Students will explore the applications of ANNs in the behavioral sciences. Students will carefully collect behavioral data associated with aggression in three species: the domesticated cat, the domesticated dog, and human children in a playground setting. These data collection exercises will be unique and elaborate. Operational definitions of aggressive behaviors will be established, behaviors from each species will be recorded on video tape, and the relevant behaviors will be tabulated as to their latency, duration, frequency, and onset and offset ramps. Further, we shall begin to assess the validity of the data collected, since these data will ultimately serve as inputs to an inter-species neural network model of aggressive behavior.

Module 3: Modeling Aggression: In this module, we shall utilize the data collected in the course Neural Networks II. Neural networks can correctly learn only if the training set consists of examples based on valid data. Therefore, great care must be taken to present non-correlated inputs, remove outliers in the data and use prior expert knowledge to find as many relevant inputs as possible. Care must also be taken that the training set is sufficiently large and representative of the population. Half of the entire data set will be used to build and train the neural network model, which we hope will accurately predict aggressive behaviors in each species and eventually lead to a global, unified model of the manifestation of aggressive behaviors within and between species. This is a complex problem since changes in the amplitude, duration, latency and rates of onset and termination of each behavior interact over time. It is a multivariate problem, perfect for an ANN approach. We shall test the neural network model by predicting the likelihood of aggressive encounters in the second half of the data set, then validate the model by visually examining the aggressive behaviors in these recorded data. We shall then refine the model and reiterate the above. Students who take this course will be required to present their findings at an appropriate professional conference.

Co-PI: Ignatios Vakalis

Institution: Capital University

Discipline: Mathematics and Computer Science

Module 1: Elementary Partial Differential Equations: From Analytic to Numerical Techniques: This module will serve as an extensive tutorial for introducing students to the field of partial differential equations (PDEs). Typically the field of PDEs is not part of an undergraduate math/science curriculum, even though PDEs are one of the most widely used set of tools by scientists and engineers to model a plethora of phenomena. The three fundamental types of PDEs that will be examined include: i) the wave equation on a finite string (*hyperbolic type PDE*); ii) the heat equation (*parabolic type PDE*); iii) the Laplace equation (*elliptical type PDE*). For each type, a step-by-step solution will be presented, so students will develop an appreciation for the solution process. The method of finite-differences will be used to produce a numeric approximation for each of the three types of PDEs. Numerical algorithms will be implemented in Matlab. Stability and convergence analysis will also be discussed. In addition, the method of finite elements will be presented as an alternate method in solving PDEs. Throughout the module, a Computer Algebra System (Maple, Mathematica) will be used to produce visualizations and animations of the solutions, thus adding to the conceptual understanding. The tutorial will include homework and projects where students will solve (numerically and whenever possible analytically) PDEs, from the fields of Biology, Environmental Science, Engineering, Geology, Finance, Physics.

Module 2: Parallel Shortest Path Algorithms on Distributed Memory Machines: A Comparison Analysis: Shortest path algorithms are part of an area called graph algorithms. The background section of this module will review the fundamental sequential shortest path and all-pairs shortest path algorithms. The objective of the module is to present a new paradigm in designing parallel shortest path algorithms and implement them on distributed memory machines. The MPI library (freely available) will be used to implement the message passing algorithms on a collection of networked PCs and Unix based workstations. The module will cover the parallelization of: i) Dijkstra's algorithm; ii) Floyd's shortest path algorithm; iii) construction of minimum cost spanning trees; iv) Prim's algorithm; v) Kruskal's algorithm. A comparison analysis of the speedup will be presented. Students will extend the presented ideas to design and implement a parallel version of the Traveling Salesperson Problem, on distributed memory machines.

Module 3: Calculating the Electrostatic Potential in Parallel: This module is an extension of module #2 (Electrostatic Potential using the Laplace Equation) proposed for the Computational Physics course. It will be hyperlinked with the Computational Physics module to provide an analysis of the physical problem and the sequential approach to its solution methodology. The objective of the module is to parallelize the computational intensive methods produced by the finite difference approximations to the Laplace equation. We will present the construction, analysis, implementation, timings and speedup analysis of the algorithms on distributed memory machines. The freely available library MPI will be used to implement the algorithms on a collection of networked PCs and Unix based workstations.

Module 4: Modeling Traffic Flow: In this module, students will be introduced into the sub-field of traffic science known as car-following theory (CFT), by building a series of increasing complex models. CFT models the case of single-lane, relatively dense traffic on long

straight highways. Students will explore Pipe's model that assumes that vehicles travel so that the distance between two adjacent is a car length. Assuming n vehicles a system of n differential equations models the speed of each car. The s -multiplied Laplace transformation will be used to produce an analytic solution for the speed of each vehicle. The above model will be enhanced by taking into account that a driver's response time is proportional to a stimulus ($response(t + T) = sensitivity \times stimulus(t)$). A linear car-following model using a constant sensitivity coefficient will be constructed. A Computer Algebra System (Maple, Mathematica) will be used to construct an analytic solution of the model. Stability analysis of the model is a crucial part of the model's assessment. Mathematical analysis to questions such as: "What is the response of the system to a small perturbation", will characterize a system as stable or unstable. The linear car-following model will be extended to include a non-uniform constant sensitivity parameter, giving rise to non-linear car-following models. Further extensions of the models will be presented as projects where students will examine situations such as: i) two dimensional CFT-type models to deal with convoys of ships, bird or aircraft flights; ii) introducing different sensitivities for acceleration and deceleration; iii) including memory functions to represent past behavior.

Module 5: Diffusion in Biology: Diffusion is the process by which matter (salute molecules) are dissolved and transported through the solvent as a result of random molecular motion. Though diffusion many metabolites are exchanged between cells and its environment of between the blood stream and tissues. Diffusion is a direct result of the random motion (Brownian motion) of the molecules in the direction of a gradient. Using Flick's law, one and three-dimensional diffusion equations (PDEs) will be generated. Student will examine the solution methodology of the diffusion equation for: i) the one dimensional case, analyzing the steady and unsteady cases; ii) the case of a solute through a plane membrane of thickness h the two sides of which are maintained at constant concentrations. The model (PDE) is solved: i) analytically (separation of variables, Laplace transformations); ii) numerically (Matlab). A Computer Algebra System (Maple, Mathematica) is also used in the solution process. Students will extend the model by constructing the convective transport, where the solute is in a moving liquid that is transported by the flow.

Module 6: Pharmacokinetics: Analysis of Drug Distribution in Living Organisms: The mathematical theory of drug phenomena is a branch of the general theory of metabolism (a process by which in any living body, food is converted into energy). This model presents the two general aspects of the theory of drug phenomena: i) the distribution of a drug in an organism; ii) the biochemical kinetics of the interaction of the drug with different components of the organism and the mechanism of its metabolism. Compartmental problems and their associated models (one, two and of three compartments) will be presented as background information. Such problems arise frequently in biological applications. Compartments can be either physically separate or chemically distinguishable pools sharing the same space. The modeling process will begin by considering the living organism as a system of n components allowing m metabolites to be involved in the exchange between components. A system of differential equations will be generated to model the phenomenon, and the two component system with one metabolite will be solved analytically using an Computer Algebra System (CAS). Solutions will be presented in both Maple and Mathematica. The second part of the module will examine the modeling process (a two compartment, blood-tissue system) in the case where a drug is injected at a constant rate and the drug is taken up by one particular tissue on which the biochemical reaction occurs. The system of differential equations will be solved using a CAS; students will examine (graphically) the relationship of a drug on the tissue and its rate of injection for safe or lethal effects.