Fall, 2006

Ocean Informatics Matlab Working Group, Mirroring the LTER Community Approach

- Lynn Yarmey (CCE/PAL)

Matlab is a matrix manipulation software application that meets the computation and graphics needs of many data handlers in a number of different research groups. A new Matlab Working Group (MWG) represents a mechanism for sharing experiences, building community and shortening the learning curve associated with this powerful software package. The group meets a growing need for local technical training and communication within an academic research organization. MWG is developing within the Ocean Informatics environment, which is formed around the long-term arrangements of the CCE and PAL LTER sites in collaboration with the California Cooperative Fisheries Investigations (CalCOFI) and other Integrative Oceanography Division participants at the Scripps Institution of Oceanography.

The group was founded in June of this year after recognizing a recurring thread in technical data analysis discussions: there are difficulties encountered by new users starting out in Matlab. New instrumentation and data streams are overwhelming traditional local data handling capacities involving legacy 2D, Excel-based processes. As a first step toward preparing for future workflow applications, Matlab is a reasonably priced and robust alternative, but the transition to Matlab with its software environment, structured line commands, and matrix concepts is daunting without training, resources or help. Previous local Matlab communities waned with changes in infrastructure, needs and personnel. With a new influx of individuals starting out, this community working group provides a mechanism to link up with other new or more experienced Matlab users in the physical and organizational vicinity.

The current working group consists of more than a dozen active participants as well as a number of inactive members. Membership crosses social boundaries by including staff, students and researchers. Participants are involved in a wide range of interdisciplinary projects and all come with a diversity of datasets and interests. Once a month, with a different member preparing and presenting a particular topic of interest, we meet informally to discuss, share experiences, and mentor those not familiar with the material. Meetings are theme-based, with previous topics including methods of importing data into Matlab, debugging scripts and functions, and an overview on the concepts and uses of cell arrays. Rotating the meeting presenter ensures relevance, promotes engagement, prompts leadership and creates a mechanism for participants to delve deeper into targeted topics.

By building interdisciplinary partnerships amongst the participants, the group presents a unique opportunity to design and enact local variable naming conventions, coding guidelines and common workflows, which all hold the potential of contributing to a shared infrastructure. Organizationally, we are in the process of designing a community script repository to pool home-grown tools and resources, building upon the group's enthusiasm for better documentation and linking personal coding methods. Through team dialogue about concepts, problems and solutions, we are developing a common language

that can help us articulate our needs, work and accomplishments as data collectors and users, which in turn inform researchers in their data handling plans and data scoping practices. Also, the group creates an arena for building identity, prompting for feedback, and redesigning communication processes. For example, we initially discovered that Matlab resources (books, blog posts, etc) were scattered and not easily available. In response, we have developed a group website that brings many of these resources together along with locally pertinent information and presentation products from our meetings.

The group is moving slowly and in a personal and grounded way towards goals including generic data ingestion, pre-processing, processing and initial visualizations. This process of loose standardization involves community decisions based on an intimate knowledge of local data acquisition, handling, practices and common needs (i.e. specific oceanographic calculations, corrections, etc). In many ways, technology has created an obstacle to this process rather than simply adding capabilities. Matlab bugs, multiple operating system representations and a range of computing and display capacities have all served as limiting factors in bringing together the data and data processing of individuals. We are working to mediate these differences through development of platform-independent coding guidelines and optimization training as well as by making local solutions available in the form of a shared script repository.

The Ocean Informatics endeavor focuses on creating an environment supportive of collaboration and mutual learning. The Matlab Working Group contributes to such efforts as a community approach to exchanging, preserving, building and sharing knowledge. WIth an emphasis on tools, partnerships and ongoing education, MWG mirrors a number of the features of the LTER community that support site science.