Proficiency in Ocean Data Science (PODS) education through innovative learning experiences and technological support

Bruce D. Campbell, Christopher R. Kincaid**, Lucie Maranda**, Robert A. Pockalny***

* Rhode Island School of Design ** University of Rhode Island

Abstract: We designed and implemented a five-course Proficiency in Ocean Data Science (PODS) curriculum for undergraduate STEM students interested in learning through high-impact multidimensional interventions and pedagogical best practices applied to an ocean-based collection, analysis, interpretation and communication of data. We committed to designing the program to engage and retain early career STEM undergraduate students through a series of data-centric courses and research experiences that would prepare them for careers in fields critical to the ocean's component of national security. As we came together with different professional experiences that explored a wide-variety of teaching approaches, we benefitted greatly from weekly collaborations to find common ground, evaluate recent results, and generate a coherent program. We share the emergent PODS heuristics, processes, and technologies we consider to have been most successful while pursuing our goals in the first three years of PODS program work – most prominent being a continual use of a notebook-based analysis process facilitated by the Python programming/scripting language

Introduction

We four faculty members, with participation from campus faculty, designed and implemented a five-course *Proficiency in Ocean Data Science* (PODS) curriculum program intended for undergraduate science, technology, engineering, and math (STEM) students at the University of Rhode Island (URI). New data science departments, programs, and competencies have been recently announced on many campuses, but ours focuses on the ocean as an area of study. The PODS program emphasizes the interdisciplinary nature of ocean science and facilitated learning through high-impact multidimensional interventions and pedagogical best practices applied to an ocean-based collection, analysis, interpretation and communication of data. We were challenged by the Navy's concern of suboptimal rates of retention to engage and retain early career STEM undergraduate students so they were motivated and capable of finishing four-year degrees in a STEM discipline [1].

In the summer of 2016, we began attending weekly curriculum design meetings to assign overall program learning objectives, advised by a PODS taxonomy (figure 1), into five sub-courses. Weekly meetings created and adjusted syllabi to best implement an explore-focus-reflect-apply process for each learning objective. We used an Assess, Design, Develop, Implement & Evaluate (ADDIE) methodology to improve our contribution in delivering the learning objectives within the constraints of the syllabi [2,3]. During the initial three-year grant period, we continuously improved a consensus-driving whiteboard process to collaborate effectively while co-present. We supplemented in-person meetings with e-mail, web audio, video conferencing, and synchronous text messaging for impromptu communications on spontaneous, relevant topics.

We created an informational website, tens of Python notebooks and a content management system for emphasizing continuity and cohesion throughout the five courses.

Proficiency in Ocean Data Science

The PODS (Proficiency in Ocean Data Science) curriculum is comprised of five foundational courses: one at the 100-, two at the 200-, and two at the 300-level. The curriculum is intended for use in conjunction with at least one capstone research experience. The introductory 100-level course, *Field and Data Collection*, focuses on field-based environmental observations, data collection, instrumentation, and data characteristics identification. The 200-level courses, *Data Analysis and Visualization*, and *Data Analytics*, focus on data analysis and interactive visualization. The 300-level courses, *Oceanographic Data Integration I*, and *Ocean Data Integration II*, focus on developing models and simulations of data and providing tools and interfaces for performing interactive analysis. Each course has a final group project to enable students to collaboratively synthesize and apply their course experience.

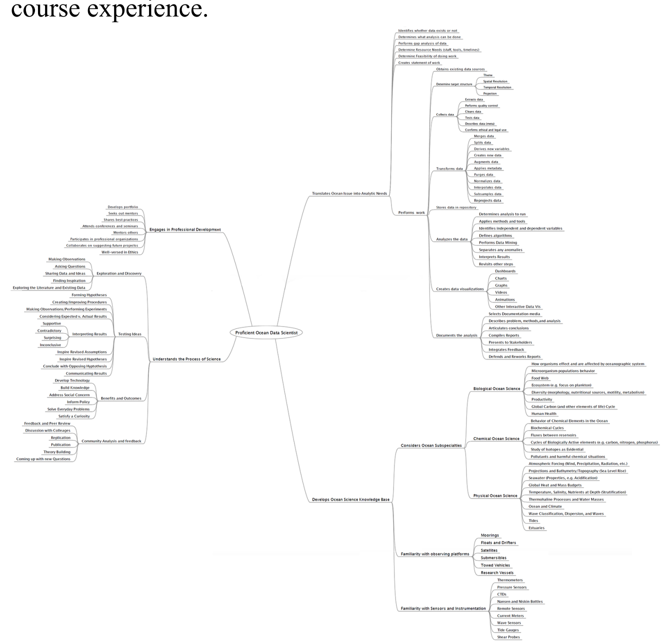

Figure 1. The PODS taxonomy used to advise curriculum (see http://pods.gso.uri.edu/images/pods-taxonomy.png).

Curriculum Design and Development

Our approach to the undergraduate curriculum is based on learning cycle models successfully applied to K-12 inquiry-based science programs [4, 5]. These learning cycle models use initial activities or focus questions to engage a student's natural curiosity, which is followed by an exploration phase to collect data or information to address the initial question. A follow-on reflection phase among peers and/or an instructor summarizes the observations and provide an explanation or interpretation of the exploration. To assess whether a student has grasped the science concept, the student is asked to apply his/her new knowledge to another situation with slightly different parameter(s). Ideally, this "final" application phase leads naturally to the student's next focus question in the learning cycle.

The iterative learning cycle closely mirrors the general approach that scientists use to understand a phenomenon and engineers use to solve problems. This approach also has a fractal attribute that can be applied to a single concept, an entire course, or even a complete curriculum.

We use the Assess, Design, Develop, Implement & Evaluate (ADDIE) model of course development and modification to infuse pedagogical best practices [2,3], while delivering in-class experiences consisting of lecture, discussion, conceptual exercises, coding exercises, and student presentations. Delivery during the initial three-year grant period resulted in the in-class time and motion breakdown seen in figure 2.

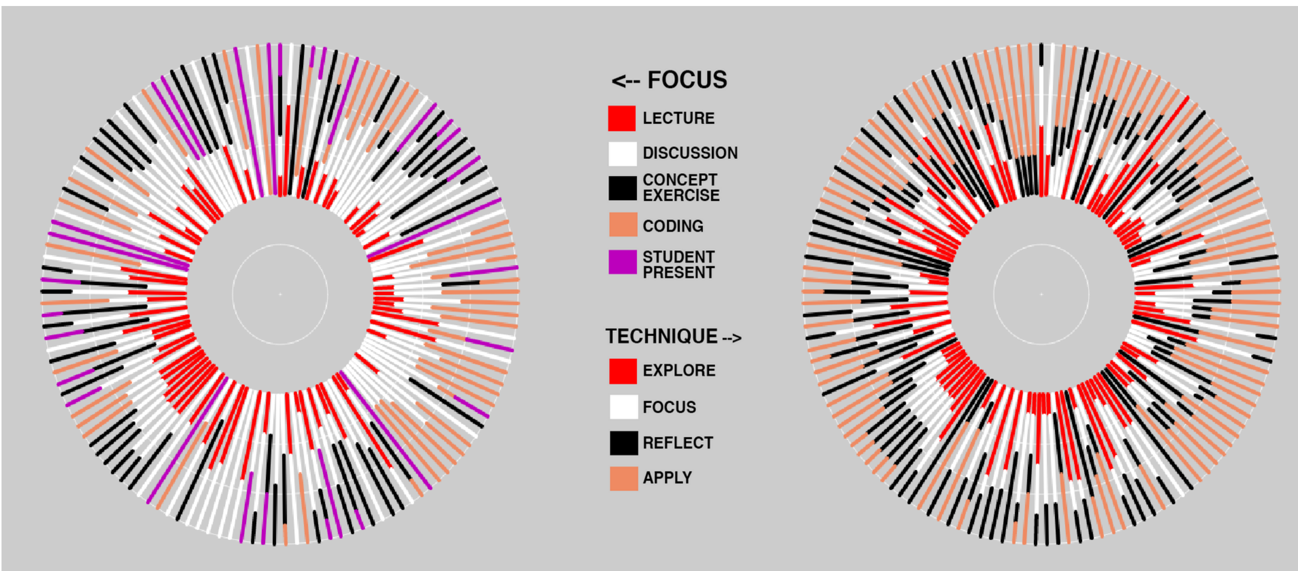

Figure 2. A time and motion study of the five PODS courses implemented during the initial three-year period. The left radial chart identifies time spent in lecture, discussion, conceptual exercise, coding, or student presentation by class session. The right radial chart categorizes time spent in the phases of a explore-focus-reflect-apply pedagogical approach. Class time and motion data begins with the first class at 12 o'clock and continues clockwise through the five courses chronologically.

We finalized our approach after reviewing research on pedagogical best practices (e.g., [6]) for making courses more student-centered (guide on the side) and less teacher-centered (sage on the stage), while continuing to apply recommendations of discipline-based research (DBER) [6, 7] and the STEM Undergraduate Model [1].

We continue to coach other participating faculty, in facilitating the PODS classroom environment, as they participate in the PODS program to make it relevant beyond grant-period course offerings. A primary and alternate instructor approach is necessary to ensure the course can be taught by prepared personnel, in case one of the responsible faculty is at sea, in the field, or on sabbatical. The combination of field studies, group projects with teams of 3-5 students, and problem-based activities throughout the sequence of courses is exactly what DBER recommends.

Multiple course strands and undergraduate research opportunities address the proposed interventions of the STEM Undergraduate Model.

Technological Support

Technological support for the PODS program includes the use of environmental sensors, guided software downloads to set up student laptop and home workstations (e.g. for Python and Google Sheets use), sessional MATLAB licenses for food web and hydrothermal modeling support, a campus-wide URI-based class-support website, a PODS coordination website (pods.gso.uri.edu), and a content management system (CMS).

The Blackfish 2.0 CMS is a high-performance, low-cost communication platform for student use throughout their academic program – to share data resources and develop skills to understand, analyze, and manipulate their physical and environmental data [8].

Through the CMS, students experience long-term data visualization and communication strategies for promoting relevant research results while gaining productivity in using an “open source, open architecture and service-oriented CMS architecture”. The open architecture affords integrated extension of CMS services, such as an online service for shared Python notebooks.

We delivered two directed learning modules focused on emerging an awareness on the value of metadata and content management systems in the support of ocean science. As URI has its own campus-wide service for asynchronous student file and communication sharing, the CMS was underutilized compared to initial expectations – as some intended function was outsourced to the URI services many students already had used competently before their first PODS course.

The CMS provides data management and discovery services for data analysis and visualization, supports a project lifecycle approach, shares project work components among students, and promotes work in a professional communications approach students can aspire to master for maximum impact of their work. The data portal grows through student use to inspire skill-development in integrating disparate data sets for hypotheses generation, similarly to how it has been used by professionals (e.g. [9]).

As the CMS is currently used to perform professional services in support of ocean sciences research involving undergraduate students and scientists, students can also learn by example from existing content (e.g., [10]).

Example Learning Modules

The PODS course pedagogy teaches through activities represented by the examples described in this section.

In one example, students are introduced to remote sensing through an activity that begins with their use of their cell phones as aerial data capture devices and creates a simulated digital elevation model from a sandbox as representative of the gridded digital elevation files they would be using in future classes.

From eight photos taken at a 45-degree angle up from the ground plane, each 45 degrees apart in a circle surrounding the sandbox, students use software to create a 3-D model and then extract a height mesh. They then use the height mesh to perform computation, visualization, and analysis steps in a Python notebook (figure 3).

A fundamental aspect of the PODS program is to use the natural draw of the local coastal region combined with a range of authentic oceanographic data collection and active learning to engage STEM students.

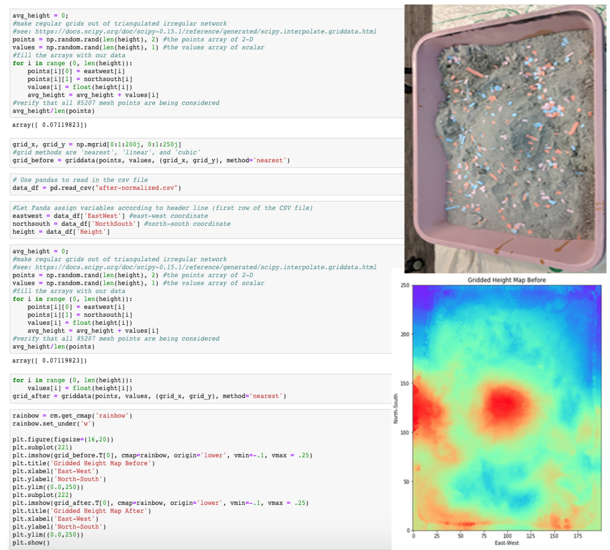

Figure 3. Python analysis of remote sensing data acquired by cell phone photography of a sand box terrain.

As a second module example, students pilot kayaks up and down the main channel of the local Pettaquamscutt Estuary, towing temperature, salinity, and pressure sensors close behind their kayaks near the water surface. After learning how to use appropriate software to extract data from the sensors, they spend weeks analyzing the data in spreadsheet exercises designed to facilitate an emergent awareness of interesting physical properties of the main waterbody of the watershed as well as temporal events, trends, and patterns.

Students are then introduced to on-line data repositories where related weather and oceanographic data are made available (e.g. https://www.ncdc.noaa.gov/cdo-web/; https://tidesandcurrents.noaa.gov/). The variety of data suggested for use via online sites increases as students progress through the PODS program.

In another module, students get out on the water to retrieve a series of moorings that had been set up in the Pettaquamscutt main channel to provide a longer term data set of temperature, salinity, and pressure. Students reinforced their sensor data extraction skills and developed Python notebook skills to perform computation on and visualization of the extracted data in unison with downloaded online data sets (figure 4).

As in the case with these examples, PODS course learning often takes place through using a Python notebook as the educational artifact of the apply portion of the explore-focus-reflect-apply pedagogical approach. The notebooks help modularize analytical thinking and support reuse of pieces.

The Collaborative Process

The co-PIs of the PODS program bring together an extensive array of research interests and educational/teaching experiences with a primary focus on ocean sciences. Our project combines the expertise of observational scientists, modelers and data visualization specialists in fields spanning from marine geology & geophysics to ocean biochemistry.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

# Use pandas to read in csv file
data_df = pd.read_csv('CO-OPS__9452210__ml.csv')

# Assign variables according to header line (i.e., first row)
msl = data_df['MSL']  #mean sea level
year = data_df['Year']
month = data_df['Month']

# quick math procedure to convert year and month into decimal year so that the value falls on
dec_year = year + month/12 - 0.5/12

# Use numpy to do a simple linear fit to time series data
model = np.polyfit(dec_year,msl,1)
predicted = np.polyval(model,dec_year)

# Detrend the data by removing the best-fit curve
detrended = msl - predicted

# Plot initial scatterplot with axes labeled
plt.figure(figsize=(8, 4))

plt.subplot(131)
plt.scatter(dec_year, msl, s=1, color='blue', alpha=0.5)
plt.plot(dec_year,predicted, linewidth=2, color='black')

plt.xlabel("Year", fontsize=14)
plt.ylabel("Mean Sea Level", fontsize=14)
plt.ylim((-0.5,0.5))
(-0.5, 0.5)

# Add an equation and correlation coefficient (R^2)of the best-fit curve
slope = model[0]
intercept = model[1]
plt.annotate('$y=%3.7sx+%3.7s$'%(slope, intercept), xy=(0.1, 0.9), xycoords='axes fraction', f

r2 = np.corrcoef(dec_year, msl)[0, 1]**2
plt.annotate('$R^2=%3.4s$'%(r2), xy=(0.1, 0.8), xycoords='axes fraction', fontsize=8)
<matplotlib.text.Annotation at 0x113506f60>

# Generate plot of detrended values (detrended = msl - predicted)
plt.subplot(132)
plt.scatter(dec_year,detrended,s=1, color='red', alpha=0.5)
plt.xlabel("Year", fontsize=14)
plt.ylim((-0.5,0.5))
(-0.5, 0.5)

# Generate histogram of residuals (e.g., detrended values) and compare to normal (i.e. Gaussia
plt.subplot(133)
import matplotlib.mlab as mlab

# Plot tide histogram
plt.hist(detrended, 50, color='red', normed=1, orientation="horizontal")
plt.xlabel("Count", fontsize=14)
plt.ylim((-0.5,0.5))
(-0.5, 0.5)

# Create random distribution data
mu = np.mean(detrended)   # mean values
sigma = np.std(detrended) # standard deviation
x = mu + sigma * np.random.randn(5000)
n, bins, patches = plt.hist(x, 50, facecolor='green', alpha=0)


# Plot random data curve
y = mlab.normpdf( bins, mu, sigma)
l = plt.plot(y, bins, 'g--', linewidth=1)

plt.tight_layout()
plt.show()
```

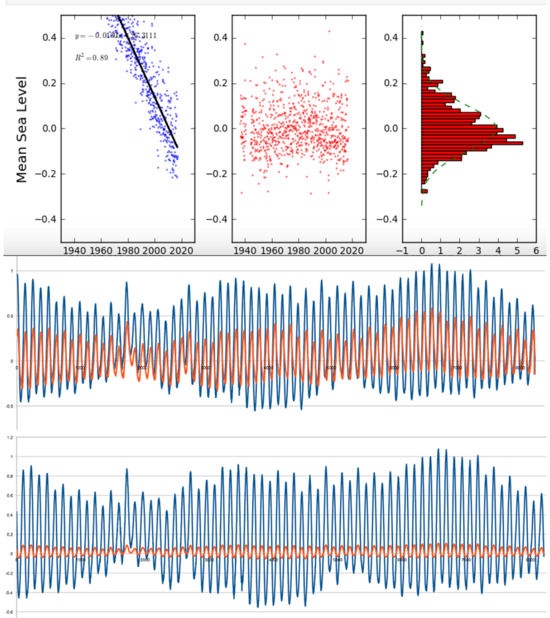

Figure 4. The Python notebook-based analysis of sea level data from kayak and mooring pressure data at depth. Multiple locations are analyzed in a combined study.

As a result, the co-PIs are actively learning from each other in preparing for class and within the classroom – fostering an active learning for everyone present, regardless of prior life experience.

During the three-year grant period, we met weekly to formulate the activities and code and perform the activities the students would be performing in upcoming classes. We challenged each other in ascertaining the students' learning capacity for the take away goals of each activity. We hoped to anticipate mistakes students would make so we could identify teachable moments whenever they occurred.

In-class experiences coalesce through these weekly faculty meetings. We share photographs and personal notes from each class session and often stay after class to review what went right and what could be done better next time. We also adjust the upcoming classes on the syllabus to make time for interjecting additional time on pursuing expected learning results that we conclude have not yet been satisfied.

Potential to Expand the Reach of the Curriculum

The PODS program has allowed us to achieve our objectives and enhance current capabilities at URI. We have gained a deeper insight into developing, implementing and sustaining a data-centric curriculum to recruit, engage and retain STEM majors. Our primary goal is to prepare students for STEM careers in the Navy and related industries, but the PODS curriculum provides important program development information and serves as a model for developing other similar courses at URI, other universities, and other education levels besides the undergraduate.

The design and implementation of the PODS program and integration with a CMS-based curriculum represents a systemic change to undergraduate education at URI and most universities. For systemic change in education to be successful, several factors need to be in place [11]. Top-down support is needed while bottom-up support is essential. A vision for change and educational best practices are required, which the co-PIs provide.

Most importantly, for sustainability, a curriculum-based program is required to be flexible, cost-effective and able to achieve the stated goals [11]. While sustainability is difficult to guarantee at any one institution, we are very confident that continued implementation of the PODS program will likely be achieved, given that all of the critical pieces are available.

One of us has had more extensive experience in creating distance-learning courses that include physical and software-based learning activities that are then shared by way of on-line critique through an asynchronous learning portal.

Although the nature of distance learning courses is most related to just one of the five PODS courses developed so far, packaging the PODS courses as a distance-learning experience would likely expand the reach of a potential learning community that could then share an expanding geographical and temporal reach of data and analysis performed through consistently delivered activity-based exercises. The maintenance of a shared CMS allows for communicating program results across supporting organizational boundaries.

Opportunities to Enrich Early Ocean-related Careers

As we move forward from the initial three-year grant phase, we envision the Navy and other potential employers participating in a STEM careers presentation as part of the course sequence. We acknowledge the Navy's contribution at the beginning and end of each course and provide additional literature or website information about programs or careers to any interested student. As part of an internship-readiness program, we inform students of Navy and other internship opportunities and provide assistance with preparing and submitting the application material in a timely manner.

The modular approach to course design enables flexibility by allowing the courses to adapt to new data, new instruments and new regions. All intensive learning exercises can be adapted for other bodies of water and virtual on-line experiences that would simulate physical exercises have been designed but not yet implemented.

An estimated 20% of the North American economy is based on ocean-related activities [12]. Current trends in research and industry are not well reflected in many educational programs and clear links are missing between education curricula, marine careers, and the marine economy. There is a lack of readily available information about careers in non-academic marine science and technology positions. Many ocean occupations are not classified by the U.S. Department of Labor and therefore are hidden from educators and students.

As a result, we have injected the PODS curriculum with motivational perspectives for caring about the ocean – reinforcing the importance of analytical results in the context of the services the ocean provides that are critical to quality of life on our planet. In that context, a basic sequence of courses on ocean data science can build the requisite ocean awareness suggested by the importance of ocean services – while at the same time providing data science methods that are appropriate for terrestrial data sources as well.

We keep an eye on the potential costs of maintaining the curriculum by seeking and testing low-cost alternatives to more expensive data systems. Our approach is to keep per-student costs reasonable and below the $200 cost of a textbook. Having invited many subject matter specialists to participate in short lectures and discussions, course materials have been generously provided as online versions of presentation materials.

Metrics and Evaluation

We incorporate the services of an experienced program evaluator to consult with us regarding both the effectiveness of the program and the potential for use among a wider audience. The evaluator's primary function on the team is to support adaptive learning and clarify team discussions by asking evaluative questions and gathering data to provide feedback and support developmental decision-making and course correction. As such, a formal three-year evaluation will provide both formative and summative evaluation components – collecting and analyzing qualitative and quantitative data.

Evaluation activities have been embedded in the program planning and implementation and have been performed at all stages of the project. During program design and planning, the evaluation activities clarified program strategy, identified expected outcomes and developed assessments. During the implementation stage the evaluation activities have centered on the development of a program management plan and the collection of baseline data to help establish connections between actions and results. During the strategic reporting stage, the evaluation activities will document our initial three-year course accomplishments, organize data, prepare reports and define the variances between the initial planned program and the produced program.

The ongoing process of investigating, understanding and improving the PODS program, processes, products, policies, performance and systems are facilitated by the evaluator. We would recommend an evaluator role in any implementation of the PODS curriculum whereby the evaluator would:

- participate in the design of project record keeping to ensure that data are collected in a manner that will provide information usable in the evaluation design.
- identify, design, administer and analyze evaluation and assessment instruments to measure the growth in participants' competence, knowledge, proficiency level and persistence.
- analyze records of curricula, program outreach, recruitment and dissemination to inform project development.
- provide continuous feedback to guide the project's activities, strategy and impact.
- integrate data and evaluative thinking to help adapt learning strategies.
- complete evaluation reports.

Conclusion and Future Work

As we are about to finish our initial PODS course sequence offerings in May 2019, we reflect back upon the experience the coordinated team teaching approach has provided our students and us as faculty: We are better at coordinating student activities and setting expectations for the explore-focus-reflect-apply components to our approach; we have arrived at a consensus for best practices in the classroom; we are better at motivating students to participate and invest a personal ownership into their progress in pursuing their STEM major; we verified the appropriateness of using an ocean science focus for the data science learning; and we are significantly more comfortable in using spreadsheet and electronic notebook tools for in-class data processing and analysis activities.

We have noticed a difference in how students and faculty enjoy each course's final exam period, which consists of student oral and written reports, more so than our traditional lecture and text-based courses and expect that enjoyment to linger into each semester of a STEM major's degree pursuit.

We have experienced a sense of community in the classroom that provides a comfort level as students give their final presentations, and anticipate benefits that will come from building an even better sense of community throughout the PODS program. We sense we will have to work hard to make sure that sense of community can expand into any distance-learning version of the materials we extrapolate from the in-person classes, and that synchronous tools will have to afford a serendipitous interaction among students and faculty.

We are generally capable in anticipating performance through the interaction style by which we were assessing each student during every conceptual exercise, coding session, and short presentation. This ability to fully assess performance may be more difficult in a distance-learning format.

Providing materials for students to consider ocean-related careers may be be useful earlier in a K-12 context. Providing materials for the general public to develop skills that build community awareness of ocean-related issues will likely help communities respond to potential climate change events such as biodiversity change, sea level rise, and ocean acidification – or at least develop an empathy for those communities experiencing stress from those events.

We are very fortunate to have a marine environment abutting campus and an ideal river-estuary system within a short drive from the classroom. We anticipate a growth in virtual reality systems that can deliver virtual experiences that better simulate physical experiences in the future. We prefer outdoor activities and have discussed coordinated activities for potential distance-learning students – perhaps requiring students to travel to locations where intensive fieldwork can be experienced. As ocean services awareness provided by one shared global ocean is relevant everywhere, we encourage interested educators to contact us for help in developing a PODS program for their target student base.

PODS can be a useful program for anyone interested in our teaching approach, anyone trying to bring similar technologies into realtime activities in the classroom, or anyone considering an ocean-focus for satisfying STEM education goals.

References

[1] *Business-Higher Education Forum (2013). The U.S. STEM Undergraduate Model*. PDF retrieved from http://www.bhef.com/publications/us-stem-undergraduate-model-applying-system-dynamics-help-meet-president-obamas-goals, pp 1-23;

[2] Branson, R. K., Rayner, G.T., Cox, J. L., Furman, J. P., & King, F. J. (1975). *Interservice procedures for instructional systems development: executive summary and model.* Florida State Univ. Tallahassee Center for Educational Technology.

[3] Hannum, W. (2005). *Instructional systems development: A 30 year retrospective*. Educational Technology-Saddle Brook then Englewood Cliffs NJ, 45(4), 5.

[4] Bybee, R. W., Taylor, J. A., Gardner, A., Van Scotter, P., Powell, J. C., Westbrook, A., & Landes, N. (2006). *The BSCS 5E instructional model: Origins and effectiveness.* Colorado Springs, CO: BSCS, 5, pp. 88-98.

[5] National Science Resources Center. 1997. *Science for All Children: A Guide to Improving Science Education in Your School District*. Washington (DC): National Academy Press.

[6] Kober, N. (2015). *Reaching Students: What Research Says About Effective Instruction in Undergraduate Science and Engineering*. National Academies Press.

[7] Singer, S. R., Nielsen, N. R., & Schweingruber, H. A. (2012). *Discipline-based education research*. Washington, DC: The National Academies.

[8] Blackfish CMS (2016). *The CMS for the Center for Environmental Visualization at the University of Washington*. Accessed from http://cev.washington.edu/story/BlackFish.

[9] *Networked Observations and Visualizations of the Axial Environment* (2015). Axial Volcano Workshop 2015. Accessed from http://novae.ocean.washington.edu/story/Workshop_2015.

[10] *River Systems Research Group* (2016), University of Washington. Accessed on March 2, 2016 from http://boto.ocean.washington.edu/.

[11] Adelman, H. S., & Taylor, L. (2003). On sustainability of project innovations as systemic change. *Journal of Educational and Psychological Consultation*, 14(1), pp. 1-25.

[12] *State of the U.S. Ocean and Coastal Economies* (2016). Middlebury Institute of International Studies at Monterey, Center for the Blue Economy, pp. 7-11. PDF acquired from http://midatlanticocean.org/wp-content/uploads/2016/03/NOEP_National_Report_2016.pdf