Title: Demonstrating an Effective Marine Biodiversity Observation Network in the Santa Barbara Channel

Agreement number: NNX14AR62A

PI: Robert Miller – Marine Science Institute, UCSB

Co-PIs: David Siegel (UCSB), Craig Carlson (UCSB), Kevin Lafferty (USGS), Andrew Rassweiler (UCSB), Daniel Reed (UCSB), Phaedon Kyriakidis (UCSB), Deborah Iglesias-Rodriguez (UCSB), B.S. Manjunath (UCSB), Milton Love (UCSB), Andrew Thompson (NOAA SWFSC), John Hildebrand (UCSD), Douglas McCauley (UCSB)

Date: July 29, 2016

Performance Period: October 1, 2015 to September 30, 2016

Project Goals and Objectives:

Although we spend millions annually monitoring marine resources, these efforts are uncoordinated and have major information gaps. We are developing a scalable and transferable demonstration Marine Biodiversity Observation Network (MBON) in the Santa Barbara Channel (SBC), one of the most monitored areas of the world. SBC MBON will network existing monitoring efforts and fill the major remaining information gaps. Our focus on SBC allows us to effectively cover the complete spectrum of biodiversity from ecosystems to microbes within a reasonable scope of funding. This is due to the profusion of existing biological monitoring and research programs by our partners including government agencies, universities and NGOs. Yet there are significant gaps in our knowledge of even this relatively well-studied area, such as microbial diversity. Our overall objective is to provide a complete picture of biodiversity in SBC using a transferable system that integrates and augments existing monitoring programs including the NSF-funded SBC LTER program, Channel Islands National Park, and NASA Plumes and Blooms. Broad goals are to:

A. Integrate biodiversity data to enable inferences about regional biodiversity. Synthesizing information relevant to biodiversity requires integrating highly heterogenous data collected at widely different temporal and spatial scales. We employ advanced techniques in spatial statistics for this synthesis and will provide multiple biodiversity-related data products, including holistic indices that will provide easily interpretable measures of ecosystem health.
B. Develop advanced methods in optical and acoustic imaging and genomics for monitoring biodiversity in partnership with ongoing monitoring and research programs to begin filling the gaps in our knowledge. A key element of our plan is a ‘pincer movement’ using two classes of methods that approach diversity observation from opposite directions: optical and acoustic imagery from the ecosystem scale down to the species level, and molecular biology from the genetic scale up through community level.

C. Implement a tradeoff framework that optimizes allocation of sampling effort. An effective marine BON will require targeted sampling to address key data gaps, while making best use of existing sampling efforts, thereby gaining a complete description of biodiversity while minimizing cost. Optimal decisions about data collection will require a framework for balancing costs and benefits of alternative sampling. Such a framework will be used to make recommendations for how resources should be allocated in a full-scale MBON as a function of the program’s goals and anticipated funding level.

In the Cooperative Agreement with NASA, the project agreed to fulfill the following more specific goals.

1. Develop a scalable and transferable demonstration marine biodiversity observation network (BON) in the Santa Barbara Channel.

2. Integrate and augment existing monitoring programs to provide geographically-integrated time-series metrics of biodiversity and ecosystem health, a transferable BON data management system, and a sampling cost-benefit optimization framework useful for designing a BON anywhere.

3. Disseminate products and results from the demonstration marine BON to a wide range of end users from scientists to school children.

4. Develop and implement advanced methods in optical and acoustic imaging and genomics for biodiversity monitoring.

5. Integrate this activity into the international Group on Earth Observations Biodiversity Observation Network (GEO BON).

6. Prepare for, participate in, and respond appropriately to reviews of the project.

7. Participate in scientific meetings or professional society meetings identified by U.S. Government program managers as pertinent to project goals.

8. Inform the relevant U.S. Government program managers of any results nearing publication and the release of final data products so that they might prepare for the announcement and any associated publicity and/or public outreach.

9. Data produced under this cooperative agreement will be available without restriction as to its disclosure, use or duplication except as otherwise negotiated by NASA and the University of California, Santa Barbara. The goal is full and
open exchange of data and metadata with minimum possible cost, delay and restriction.

Year 2 Progress:

Here we describe our approach and progress organized by the three broad goals above.

A. Integrate biodiversity data to enable inferences about regional biodiversity.

We have assembled data from 173 coastal sites monitored by 4 different programs, many of which have 30+ year data series. These datasets document abundance of more than 350 distinct taxa including fish, invertebrates and algae. To facilitate data synthesis, we have developed a taxonomic database to enable comparison of biodiversity data from different sources. The taxonomic database is scalable to accommodate the integration of additional data sets. Biodiversity datasets that are being added currently and in the next year include data from deep-water benthic sampling, data from open water plankton sampling, and data from ichthyoplankton sampling. Specific data that we have acquired include the following:

NPS Kelp Forest Monitoring (KFM)
- 37 sites across Channel Islands (San Miguel, Santa Rosa, Santa Cruz, Anacapa, Santa Barbara, and San Clemente Islands)
- Annual sampling from 1982 to present
- 118 distinct taxa
- Benthic density, benthic cover, fish density data, fish size-frequency, kelp tagging, subtidal temperature

Partnership for Interdisciplinary Studies of Coastal Oceans (PISCO)
- 118 sites across Southern California Bight
- Varied sampling from 1999 to present
- 131 distinct taxa
- Benthic density, benthic cover, fish density, fish size-frequency

San Nicholas Island (SNI)
- 7 sites around San Nicholas Island
- Biannual sampling from 1980 to present
- 325 distinct taxa
- Benthic density, benthic cover, fish density, kelp tagging

Santa Barbara Channel Long Term Ecological Research (SBC LTER)
- 11 sites across north side of Santa Cruz Island
- Varied sampling from 1982 to present
- 307 distinct taxa
- Benthic cover, fish density, kelp tagging, subtidal temperature

We have begun to incorporate citizen science data from Reef Check California to compare with expert-collected data from the above programs.
Deep-water fish survey data provided by Milton Love and colleagues have also been incorporated into the BON data management framework.

This year we set up a transferable data integration flow that can be applied to other datasets and MBON projects. The basic steps of data integration consist of (1) Data cleaning, including adding missing information (e.g. sampling method), standardizing time/date format, and correcting any data entry errors; (2) Data standardization, including homogenizing column names across projects and checking species’ scientific names to meet the requirement of current convention; (3) Data integration, including extracting information from all projects (e.g. site locations, transect depths, and taxonomy), as well as combining datasets based on data categories (e.g. taxonomic categories); (4) Dataset production (sometimes including summary statistics) for ultimate dissemination and analysis.

Additionally, we are in the process of combining biological data with physical data to allow analysis of the physical drivers of biodiversity. Because our goal is to use these results to develop regional scale inference about patterns of biodiversity, we focus on physical data for which we have continuous coverage within the region. These data include 1) bathymetry and substrate data derived from sonar surveys, 2) sea-surface temperature, chlorophyll and local kelp canopy derived from satellite imagery, 3) wave disturbance and ocean current data derived from regional models, and 4) data on the shape of the coastline calculated from maps. Physical data that have thus far been processed and compiled for predicting biological spatial-temporal variations include:

- Chlorophyll a http://coastwatch.pfeg.noaa.gov/
- Ocean connectivity Siegel lab
- Wave disturbance SBC LTER
- Landsat-derived kelp biomass SBC LTER/Siegel lab

We recruited a data specialist, Li Kui, to replace Jacob Staines (who is going on to grad school), to continue acquiring and integrating biodiversity data in ways that are sustainable into the future (clearly annotated code, versioning, full metadata). MBON postdoc Thomas Lamy started in July 2015. Dr. Lamy came to us from a postdoc in the lab of renowned spatial statistician and ecologist Pierre Legendre, University of Montreal. Lamy is developing predictive models of marine biodiversity using physical and biological data derived from remote sensing. In collaboration with Dr. Brian Kinlan, a marine spatial ecologist with the NOAA Biogeography Branch, we (Miller, Rassweiler, Reed) leveraged MBON to obtain NOAA funding for an additional postdoc. We recruited Dr. Rhiannon Rognstad for this position. Rhiannon earned her Ph.D. in Biology at the University of South Carolina at Columbia, where she worked with Jerry Hilbish on marine population connectivity and how it shapes range boundaries and interacts
with climate change. On SBC MBON she will work on developing species distribution models to relate physical and biological data, starting in August 2016.

Analyses thus far show that reef community variation in the SBC is mainly explained by a linear trend, from east to west, driven almost entirely by sea surface temperature. Communities also vary at broad scales (> 60km) within the SBC and this variation is explained primarily by differences in propagule delivery (e.g. site oceanographic destination strength average over all the model domain) and by differences in pelagic primary production (mean chlorophyll a). We find that communities do not, in general, vary significantly at scales smaller than ~30 km. Each island in the region has different assemblages and it is difficult to detect significant community variation (scales of ~30-60 km) within an island. Surprisingly, wave disturbances (mean wave height and number of days waves exceed 2m) and kelp canopy biomass (mean long-term canopy biomass, inter-annual variability in kelp canopy biomass and extinction probability computed as the number of years a site had no canopy over the total number of years) turned out to be poor predictors of biodiversity. However, in a smaller high-resolution dataset (SBC LTER), biodiversity of fishes and sessile invertebrates was positively affected by kelp.

Statistical analyses have been done to test the temporal and spatial changes in deep-water fish communities. Three sites, Anacapa passage, Footprint, and Piggy Bank, for which we have time-series data, have been the main focus for these analyses. Preliminary results suggest that the fish communities are driven by habitat, especially depth and substrate. The fish species seem to have shifted in depth over the past 15 years and sizes of some fish species have decreased over time. The above results for shallow and deep benthic communities are being prepared for publication in 2016-17.

**SBC BON Remote Sensing**

Remote sensing data is critical for scaling up local observations of biodiversity and for relating physical and ecological variables to marine biodiversity. The SBC mBON is focused on three activities linking remote sensing to observations of marine biodiversity. These build on on-going NASA and NSF supported projects and span both subtidal and pelagic environments. Specifically, the remote sensing component of the SBC BON is focused on:

- The remote sensing of giant kelp populations and the application to these products to answer ecological and biodiversity questions
- The analysis of satellite ocean color observations using novel ocean color inversion approaches, and
- The analysis of planktonic biodiversity indices from the Plumes and Blooms (PnB) time-series study applied to SBC BON genomics observations.
Accomplishments from all three of these components are reported in this annual report.

Both multispectral and hyperspectral remote sensing of giant kelp canopy fronds are used to help answer ecological and biodiversity questions for the SBC BON. We have assembled a Landsat giant kelp (*Macrocystis*) canopy biomass time series from 1984 – present, spanning its dominant range in the NE Pacific (San Francisco, CA to Punta Eugenia, Baja California Sur, Mexico). This dataset has recently been used to determine the geographical variability and non-linear response of the environmental controls of giant kelp biomass dynamics (Bell et al. 2015a), test metapopulation theory, showing that well connected patches had high probabilities of colonization and lower probabilities of extinction (Castorani et al. 2015), and identify large-scale biogeographic population clusters using seascape genetic approaches (Johansson et al. 2015). We are now able to working to combine data from three Landsat sensors to produce a continuous and current time series of giant kelp canopy biomass in the North Eastern Pacific. We are combining Landsat 5 TM, 7 ETM+, and 8 OLI imagery, and verifying that biomass estimates are unbiased by comparing each sensor to field measurements as well as simulating different Landsat sensor images from hyperspectral data. We have discovered a tidal bias that we are correcting for. We have also developed a gap-filling algorithm to estimate canopy biomass in pixels with data loss due to the scan line error in Landsat 7 ETM+ imagery (2003 – present). We have optimized the relationship between canopy biomass and adult plant density for better estimations of individuals on a reef. This should be ready for submission by the end of the year

Multiple hyperspectral images of the SBC are being collected as part of the HyspIRI Preparatory Airborne Campaign. Empirical relationships between laboratory reflectance of giant kelp blades and their physiological state (Chl:C) have been developed and applied to hyperspectral images captured by the AVIRIS sensor. These images have shown that the physiological state of the canopy is positively correlated to the depth of the reef where the kelp plants attach, and may have implications for measuring the productivity and age structure of kelp forests (Figure 1; Bell et al. 2015b). We have validated the hyperspectral Chl:C algorithm with field data from sites in Santa Cruz and the SB Channel. We have applied this algorithm to the full 3 years of hyperspectral imagery taken as part of the HyspIRI Preparatory Campaign. Regional patterns of kelp canopy physiological condition are related to environmental patterns of nitrate and light. Local scale patterns seem to be related to demographic effects of the canopy. We are testing this by observing change in the kelp canopy biomass and time since first observation in high-resolution imagery from a variety of satellite sensors. Tom Bell is currently writing this manuscript for submission in late 2016/early 2017.

In connection with this work, we are also using the chlorophyll to carbon (Chl:C) field time series from sites in Santa Cruz, SB Channel, and San Diego to understand the environmental controls of kelp canopy physiological condition and
how it relates to biomass and NPP patterns. We found that Chl:C (our proxy for physiological condition) is antagonistically balanced by ambient seawater nitrate and insolation, both by non-linear relationships. We produced a temporal autoregressive function of kelp canopy biomass at each site (3-month lag) and found that the residuals were positively correlated to the Chl:C time series. This showed that changes in physiological condition are related to the accumulation and loss of kelp canopy not predicted by the canopy biomass state in the recent past. We also found that physiological condition measurements along with measurements of standing foliar crop provided simplified estimates of kelp NPP compared to present methods. These results open the possibility of estimating net primary production over large spatial and temporal scales using present and planned remote sensing technologies and the modeling of physiological condition based on environmental variables. This manuscript is with the coauthors for review.

The second goal was to characterize what controls space-time variability of optical properties of the Santa Barbara Channel and, more specifically, to understand how changes in chlorophyll concentrations (chl) and particulate backscattering coefficients (b<sub>bp</sub>) co-vary over time and space. This project has now been concluded. Satellite observations of chlorophyll (CHL) in coastal waters are often described in terms of changes in productivity in response to regional upwelling processes while optical backscattering coefficients (BBP) are more often linked to episodic inputs of suspended sediments from storm runoff. We were able to show however that the surface gravity wave resuspension of sediments has a larger role in controlling BBP than previously considered. Almost 18 years of 2-km resolution SeaWiFS, MODIS, MERIS and VIIRS satellite imagery of the Santa Barbara Channel and its surrounding waters, spectrally merged with the Garver-Siegel-Maritorena bio-optical model, were used to assess the importance of physical

![Figure 1. Map of the distribution of (a) estimated Chl:C in the surface canopy (b) depth and (c) kelp pixel fraction of Santa Barbara area kelp forests on April 11th, 2013. Insets show magnified data of the Isla Vista kelp forest and reef. Data for a. and c. from AVIRIS hyperspectral images, and b. from NOAA National Geophysical Data Center coastal relief model (Divins & Metzger, 2009).](image)
forcings on chlorophyll and suspended particle distributions. The space-time distributions of BBP and CHL estimates from the model were analyzed using Empirical Orthogonal Function analysis, and the resulting spatial and temporal patterns were compared with environmental variables. Analysis revealed that CHL blooms in the warmer portions of the domain occur in phase with SST minima, usually in early spring, while CHL blooms in the cooler regions lag SST minima and occur simultaneously to the strongest equatorward winds every year, often in the summer. Connections among CHL changes and El Niño conditions were also found, illustrating the wide range of processes that affect CHL variability. Tight coupling between BBP and CHL variability was seen in offshore areas, as expected for productive waters. However, values of BBP near the coast were primarily modulated by surface waves. The relationship between BBP and surface wave height holds throughout all seasons and extends several kilometers offshore, until about the 100 m isobath. This forcing of particle resuspension by surface waves is likely a feature ubiquitous in all coastal oceans characterized by fine sediments. The implication of surface waves determining BBP variability beyond the surf zone has large consequences for the interpretation of satellite ocean color signals in coastal waters and potentially redefines the extent of the littoral zone. Fernanda Henderikx-Freitas, a student in the Siegel lab, is conducting this work, which was submitted for review in the Journal of Geophysical Research – Oceans in July 2016. We are also collaborating with Maria Kavanaugh (WHOI) in applying these data with her seascape characterization approaches for mapping marine biodiversity indices.

The goal of the Plumes and Blooms (PnB) study is to assess and model ocean color changes in a complex coastal site and is supported by the NASA Ocean Biology and Biogeochemistry program. PnB conducts monthly day-long cruises at 7 stations crossing the Santa Barbara Channel (~45 km long transect). At each station, measurements of ocean color spectra, inherent optical properties, phytoplankton pigment, dissolved and particulate carbon, and macronutrient concentrations, and particle size spectra are measured. Samples are also collected for particulate DNA for both prokaryotic and eukaryotic (Emma Wear et al., work in progress). Our eventual goal in the SBC mBON is to statistically relate bio-optical information of planktonic biodiversity from the PnB measurement suite (phytoplankton pigment concentrations, phytoplankton absorption spectra, etc.) with signatures from the SBC mBON genomic work. We are working with Emma Wear, a graduate student in the Carlson laboratory, to compare genomic indices of prokaryotic biodiversity with the PnB bio-optical data. A new student, Dylan Catlett, is now continuing this work in collaboration with with Profs. Siegel, Carlson and Iglesias-Rodriguez. Catlett is analyzing a long time series (2006-2014) of hyperspectral phytoplankton-specific absorption spectra ($a_{\text{ph}}$) and phytoplankton pigment concentrations generated by Plumes and Blooms. Absorption features in $a_{\text{ph}}$ have been identified using a derivative analysis and these spectral signatures have been linked with individual phytoplankton pigment concentrations. Phytoplankton community structure in the SBC over this time period has been characterized via empirical orthogonal
function analysis and corroborated using cluster-based analyses, and we are currently investigating links between these estimates of phytoplankton community structure and the spectral signatures identified in the derivative analysis. Moving forward, we hope to use the signatures in $a_{ph}$ identified from the derivative analysis to model individual phytoplankton pigment concentrations and phytoplankton community structure.

B. Develop advanced methods in optical and acoustic imaging and genomics for monitoring biodiversity.

Optical imagery - field
We have used an SLR with 14mm lens mounted on a rigid frame (quadrapod) to image shallow benthic and kelp forest communities at 14 reef sites spanning the Santa Barbara Channel, including several Channel Islands National Park monitoring sites, as well as seven offshore oil platforms. In deep water, we have obtained still imagery data collected from a SeaBED class autonomous underwater vehicle (AUV) that surveyed the benthos in the Santa Barbara Channel region. Transects were conducted on two seamounts known as the “Footprint” and “Piggy Bank.” Elizabeth Clarke, who leads the AUV program at NOAA NMFS NWFSC in Seattle WA, provided the data and will be collaborating with us as we move forward.

The SeaBED AUV is a hover-capable robotic vehicle that is able to work as close as 2 m off the seafloor while maintaining precise altitude and navigation control. Its 2,000 m depth rating makes it an ideal tool for conducting surveys of reef, shelf and deep slope habitats. Its small footprint allows it to be operated from platforms ranging from global class oceanographic research ships to small vessels of opportunity. The SeaBed AUV can carry a wide variety of optical, acoustic, and oceanographic sensors for non-extractive surveys of the benthic communities in habitats that are too deep for divers and surface acoustics and too rugose for towed camera sleds and traditional bottom trawling. The area of each image is estimated from the measurement of the AUV altitude off the bottom and the specified camera field of view. This allows for density estimates of species abundance, biomass, and diversity.
The AUV data was collected during a “comparison cruise” in 2011. The objective of the cruise was to survey fish using the SeaBED AUV, the Dual Deepworker submersible, and an ROV coupled with fish detecting hydroacoustics and to compare the abundance, size distribution, biomass and diversity estimates from the different platforms to understand the capabilities of the different technologies and methods to assess West Coast groundfish populations in untrawlable areas.

The Piggy Bank is about 30 km$^2$ in area, ranging in depth from 275 to 900 meters; the Footprint Bank is about 10 km$^2$ in area, ranging in depth from 80 to 500 meters. The underwater visual surveys were planned to span from 400 m to the top of each seamount. The imagery includes stereo pairs that can be used to estimate size and biomass of organisms.

This year, we have been building the BisQue training dataset comprised of graphical annotations from AUV images collected at the two deep seamounts. We have fully or partially annotated 2600 images from 11 of the 25 survey blocks sampled by the AUV in Figure 2. Seven of the cells are located at the Footprint at depths ranging from 116 m near the top of the feature to 419 m, and four cells at the Piggy Bank ranging from 277 m at the top of this deeper feature to 455 m. We have identified at least 29 species among the 2752 fishes manually annotated in the AUV images. We have 22.5K data points of at least 60 invertebrate taxa and 15 unique types of sponge based on color and morphology located in the images. We are annotating bacterial mats, algal drift, and anthropogenic debris (e.g., fishing gear, beer cans).

In the interest of developing BisQue as a tool to both characterize habitat in imagery and explore species diversity in relation to habitat variability, we are expanding the BisQue dataset by annotating physical substrate (e.g., mud, sand, cobble, boulder, bedrock) in images. Quantifying habitat heterogeneity in imagery by traditional methods such as overlaying a grid point matrix to estimate percent cover is time consuming and problematic, particularly when resolving small-scale variability. New techniques to automate the characterization of substrate in images will be explored. An emergent goal is to develop the automation of detecting and quantifying the overlap or proximity of benthic biota and habitat.

In September 2016, we will conduct an AUV survey in collaboration with Elizabeth Clarke, her colleagues affiliated with NWFSC, and CINMS to characterize deep coral and sponge communities in the Santa Barbara Channel region. Density estimates of diversity, biomass, and species distributions from AUV imagery will allow the quantitative comparative analyses of habitats in this and other regions. Of particular interest are deep areas within or near the CINMS boundary, marine protected areas, and deep features open to fishing. The upcoming AUV survey will provide imagery for expanding the training dataset for BisQue.
**Image processing**

During year 2 we have continued our research on image analysis techniques targeted at underwater image classification on the operational level. Our main interest was targeted at finding techniques that could be trained automatically based only on provided training data and thus could be deployed automatically in our cloud-based image analysis and annotations system BisQue. Such techniques would have to overcome the complexity inherent in the multitude of existing image feature descriptions and classifications techniques. Previously we explored several feature aggregation techniques that would allow automated classification based on a large number of computable feature descriptors and a distributed cloud based system. We proposed and published in peer-reviewed conferences two novel techniques: 1) feature aggregation based on a CRF modeling of feature dependencies and 2) K-NN classification using dropout regularization. We further explored state of the art deep learning techniques using Convolutional Neural Networks on the same dataset and obtained comparable or better results. The major advantage of these techniques is that CNNs learn image features automatically (in convolutional layers) and operate directly on image pixels that can be efficiently accessed via the BisQue system. The major disadvantage is the computational complexity that requires use of the latest GPUs. We have used a very popular CNN library “Caffe” running on an nVidia K20 GPU and observed training times of 14 hours for the aforementioned dataset. During these studies we have used the extensive dataset of >2000 underwater images acquired during the shallow benthic field effort described above and manually annotated for percent coverage of sessile species in the BisQue system. Each image contained 100 annotated locations amounting to >200K data points with over 30 species. We obtained very encouraging results on automated identification of the 11 most abundant classes (over 80% of data points) with classification accuracy of over 85%.

Empowered by these results we have concentrated our efforts on the deep learning technology and implemented a scalable classification service for the BisQue system, utilizing an nVidia K40 GPU based server already acquired for another project. The first implementation offers two modes of classification. The first one is the automation of the percent cover annotations. Each automated point is marked with a novel measure of classification accuracy and allows user control of automated annotations. The second mode is a semantic segmentation where the image is partitioned in regions using a SLIC superpixel algorithm, segmented regions are later merged based on their classes and/or removed if accuracy does not match the requirements. The results produced by the new service demonstrated higher accuracies than previously observed with the average accuracy of 94.73% and the error of 3.65% on the same dataset with 11 classes.

We are currently in the process of building the model training (creation) user interface and scalability services. This will allow any user to train a new classification model based on their annotations without requiring machine
learning knowledge. In addition, we are implementing another classification mode for low-resolution partitioning of the image useful for substrate-based image partitioning. Here we use Voronoi partitioning combined with dense point based classification. Finally, our future work includes hierarchical classification to facilitate progressively fine-tuned classes e.g. functional groups leading to species. For example, this will allow to separate all anemones and then sub-classify them into proper species thus reducing the overall error.

BisQue extensions
In the past year we have also done extensive work on the cloud based image analysis system BisQue to enable better annotation of existing data and faster operation for the classification framework. Our main improvement was in accelerating and improving scalability of image services that are accessing pixel data in various formats and give pixel access for the analysis modules. For the case of percent cover based classification, the image service is accessing original 14 bit RAW image data and produces small 8 bit image patches for training. In this update various image operations were sped up by approximately 2-5 times.

We have also finished several User Interface improvements extending graphical annotation capabilities and summarization of annotated data. Improvements include full support for GeoTIFF imagery with export to GeoJson and KML, support for Homographic geometrical transformations of the percent cover annotation grid (helps annotating sub-regions of images), occurrence statistics of graphical annotations and various other improvements.

The basic BisQue services were also enhanced to accommodate new datatypes and improve user experience. The user interface has seen several changes in the image annotator aimed at reducing the time required for manual annotations. A new Canon 5DSr RAW format was added along with the full color spaces and profiles support for correct color representation in supporting operational systems and browsers. A new service supporting large tabular data was added for improved summarization of annotation and classification data.

Acoustics
Using the Channel Islands National Marine Sanctuary vessel Shearwater, Hildebrand’s lab has continued to successfully service two high-frequency acoustic recording packages (HARPs) in the Santa Barbara Channel. Both instruments are part of a seafloor array of five high-frequency acoustic recording packages (HARPs) deployed to 1 km depth with a maximum horizontal aperture of 1 km that has been used to estimate the directionality of underwater sound radiated by current commercial ships. This array has returned good acoustic data for 2015 that has been analyzed to examine ship noise in SBC and the wider Southern California Bight. The underwater sound radiated by commercial ships is an unintended by-product of their operation and one of the
most significant contributors to man-made noise at low frequencies in the ocean. As a ship of opportunity passed over the HARP array, the directions from the ship to each HARP along with the corresponding source levels were estimated for each ship location. Ships were tracked via satellites (Automatic Identification System—AIS). The directionality estimates of contemporary commercial ships exhibit significant stern-bow asymmetries among other quantitative characteristics. These results have important implications for managing ship traffic to minimize impacts on marine mammals, particularly whales, in the SBC.

Genomics
Taxonomic identification of microorganisms has traditionally been challenging because, in addition to limited morphological characteristics, less than 1% of the microbial diversity has been cultured successfully. The introduction of genomic approaches over the past two decades has allowed microbiologists to overcome these limitations and assess microbial diversity in terrestrial and aquatic ecosystems. The use of genetic markers with high-throughput Next Generation (Next Gen) sequencing provides high-resolution taxonomy for phylogenetic analyses as well as data for biogeographical distributions of marine microbial plankton.

Identification of metazoans has traditionally been based on morphology, which suffers from well-known limitations including phenotypic plasticity, morphologically cryptic taxa and life stages and the need for taxonomic expertise. In response, the use of DNA ‘barcodes’ has been increasingly adopted for the study of biodiversity. There are several genetic markers that can be used but the most common is mitochondrial cytochrome oxidase subunit I (COI), which has been successfully used for the identification of fish, crustaceans, protists, and many other organisms. We proposed to build on existing sampling efforts to implement a sustained DNA collection program to be used for metabarcoding and eDNA analyses.

Prokaryotic and eukaryotic microplankton
SBC MBON supports Emma Wear as a graduate student researcher. Emma is a microbial oceanographer who is skilled in using molecular approaches to reveal patterns in prokaryotic community structure and its link to physical and biogeochemical patterns in the coastal ocean’s water column. Emma has participated on monthly cruises across the Santa Barbara Channel as part of the Plumes and Blooms times series program. This DNA sample archive is being used to assess prokaryotic diversity as described below.
**Bacterial genomics:** We have been working to analyze our existing time-series of bacterial DNA samples, collected on the Plumes and Blooms cruise line (http://www.oceancolor.ucsb.edu/plumes_and_blooms), encompassing 15 samples per cruise from mid 2010 through mid 2014. We sequenced the full time-series through the Illumina platform using the 16S rRNA primers we purchased in conjunction with Mya Breitbart’s lab of the Sanctuaries MBON.

These sequencing results showed a strong seasonal cycle in the surface communities (Figure 3). The most common operational taxonomic units (OTUs) were strongly associated with bottom-up drivers (Figure 4); some of the more episodic OTUs were also clearly associated with ecosystem-scale events, such as Roseobacters blooming in conjunction with diatom blooms. We also observed very high spatial variability in bacterial community composition within a subset of cruises, on the order of the extent of variability seen in the seasonal cycle, which we hypothesize is associated with the seasonal current structure and complex internal circulation within the SBC; we will investigate this further using a combination of remote sensing and in situ measurements from the Plumes and Blooms program. While these primers worked well from a technical standpoint and gave us a good quality sequencing run, we saw evidence of potential amplification biases in our results. A small number of our samples had been previously sequenced using different primers on a Roche 454 pyrosequencing platform, and we saw mismatches in the relative abundance of OTUs that were common enough to have good coverage. The biases we suspect are also consistent with those reported in Parada et al. (2016; first available online around when we shipped our samples for sequencing), which critiqued an older version
of the primers we used. Our primer set has been updated to address a number of the issues highlighted in Parada et al. (specifically, issues with under-amplification of SAR11, which as the most common marine bacterial clade is important to sample correctly: Apprill et al. 2015), but it appears that some of the biases associated with less common phylotypes (especially the observed over-representation of the gammaproteobacterium SAR86) may remain with the newer version of the primers.

Figure 5: Surface map of Shannon diversity index over the time-series. Figure layout is as in Figure 2. Note the generally higher diversity in approximately the first third of the year, when upwelling is more common. Generally, the diversity decreased later in the year as seasonal stratification set in. In that context, the very high diversity observed in the summer of 2014 is intriguing, especially as this is during the 2014 Pacific warm anomaly; we expect that our ongoing primer testing will make it clear whether this is true ecological anomaly in the bacterial communities as well.

One of the results of these apparent biases is very low phylogenetic diversity within samples. While marine bacterial communities are well known for their lack of evenness, our results are more skewed towards common phylotypes than would be expected from both our previous sequencing work in this system and from the broader literature. In most samples, at least 70% of the community consisted of the three most common OTUs (Figure 3). While the Shannon diversity index did correlate well with a number of ecological variables – in particular, inorganic nutrients and temperature, suggesting upwelling has a positive effect on bacterial diversity – there were also a number of cruises with intriguing, unseasonably high diversity that we cannot currently definitively attribute to ecological causes rather than primer issues (Figure 5).

Work Forward:

To address these potential biases, we are comparing four sets of primers (our current set, one set that we have previously used, and two sets recommended by colleagues: Table 1) to assess whether a particular set of primers works better than the others in our system, given its large oscillations between bacterial communities associated with rich phytoplankton blooms during spring upwelling
and communities resembling those in the oligotrophic gyres during the stratified fall period. The primary goal of these comparisons to find a primer set that minimizes phylogenetic bias. While some bias is likely inescapable, this comparative testing will also allow us to estimate the error associated with our sequencing approach and to determine how well our new results can be compared with previously published results from the same system using older primers and sequencing platforms.

We are taking a three-pronged approach to testing our primer options. After this comparison, we will re-sequence the full time-series if there is a clear difference between primer sets such that one appears most consistent in replicating the known bacterial community.

**Environmental samples:** We are sequencing a subset of ~85 samples covering an annual cycle using each of the four primer sets; this will allow us to see how each performs over a range of sample conditions.

**Mock communities:** A clone library of nearly full-length 16S rRNA genes was prepared from 6 environmental samples, targeting both bacteria and archaea, and sequenced using Sanger sequencing. This library produced approximately 240 identifiable clones, encompassing 45 bacterial and archaeal clades. This library is being used to construct two mock communities of environmentally relevant phyla: one with 22 OTUs at equal concentrations, and one with the same OTUs at staggered concentrations from 1-25% relative abundance, approximating the copiotroph-dominated community composition expected during a phytoplankton bloom. We will amplify these communities with the four primer sets and sequence them with the environmental samples. By creating

<table>
<thead>
<tr>
<th>Primer pair</th>
<th>Reason for choosing</th>
<th>Known pros</th>
<th>Known cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>515FY 806R-B</td>
<td>promoted by Earth Microbiome Project (<a href="http://www.earthmicrobiome.org">http://www.earthmicrobiome.org</a>)</td>
<td>- shorter amplicons make for easier bioinformatics as well as more rigorous sequencing quality control due to greater overlap between reads</td>
<td>- older version of this pair has strong biases (Parrada et al. 2016); some have been corrected (Apupi, et al. 2015) but others possibly not</td>
</tr>
<tr>
<td>515FY 926R</td>
<td>promoted by Parada et al. (2016)</td>
<td>- multiple hypervariable regions</td>
<td>- long amplicons are bioinformatically costly</td>
</tr>
<tr>
<td>341F 785R</td>
<td>promoted by Klindworth et al. (2013); widely used in literature, including Illumina recommended protocols; recommended by colleagues</td>
<td>- multiple hypervariable regions</td>
<td>- long amplicons are bioinformatically costly</td>
</tr>
<tr>
<td>27F 338R</td>
<td>previously used for pyrosequencing samples from the SHC (as in Wear et al. 2015) – will allow us to check for sequencing platform bias</td>
<td>- long history in literature</td>
<td>- does not amplify archaea</td>
</tr>
</tbody>
</table>

Table 1. Primer sets being compared for prokaryotic genomic analysis.
samples of known composition, we can better assess how far the results from the various primer sets deviate from the expected community composition.

*Metagenomics:* We prepared ten metagenomes representing a broad range of conditions in the SBC. Because the metagenomes are prepared by amplification of random DNA regions rather than by using primers, they should provide an unbiased measure of the bacterial community; that is, they will allow us to assess the primers’ performance versus what is truly present in a small number of environmental samples. The metagenomes have been sequenced and are currently undergoing bioinformatics analysis in collaboration with Elizabeth Wilbanks of CalTech.

**Eukaryotic Plankton Diversity**

In April, 2016 we recruited Dr. Paul Matson as a postdoctoral scholar to work on the Eukaryotic plankton diversity. The goal is to use next-generation sequencing technology (Illumina MiSeq) to quantify the diversity of eukaryote plankton in the Santa Barbara Channel across space and time. Using DNA extracted from monthly seawater samples (both archived and newly collected), we will identify species based on the presence of genetic sequencing corresponding to three unique barcoding genes (18S, COI, and 16S). This work is composed of three parts: (1) testing the taxonomic resolution of these genes using a diverse mock community of phytoplankton of known identity; (2) applying these tools to an archival monthly time series of samples collected between 2011 and 2014; and (3) generate a new monthly time series beginning in 2016. At present, we are acquiring cultures of known plankton species for inclusion into our mock community, with a targeted assemblage of 25-30 different species across multiple taxonomic groups. In addition, we have begun collecting monthly samples for the new time series, beginning in April 2016.

**Ichthyoplankton**

The main goal of the ichthyoplankton component of the SB MBON is to develop genetic tools to efficiently identify larval fishes collected through plankton sampling. Co-Investigator Andrew Thompson, an expert in larval fish identification, utilizes traditional morphology-based methods to identify larval fish species within plankton samples. After identification, the larvae are returned to the sample and then we test whether genetic tools can accurately identify the known species. Because most of the larval fishes in California can be identified to species based on morphology, we will be able to directly assess the efficacy of metabarcoding to assess ichthyoplankton biodiversity.

In 2015 and 2016 we collected plankton samples for this project with the help of several collaborators. First, in 2015 CalCOFI collected samples with pairovet nets from four quarterly cruises at 308 stations between San Francisco and San Diego specifically for SBC MBON. Second, in May 2016 members of the SBC MBON and the Channel Islands National Marine Sanctuary collected pairovet samples in the Santa Barbara Channel at the Plumes and Blooms program stations. Third, in collaboration with the city and counties of Los Angeles ocean
monitoring groups we collected ichthyoplankton samples from Santa Monica Bay in June 2016.

We have made significant progress in morphologically identifying larvae. An SBC MBON technician has removed (i.e. sorted) larval fish from all samples, and Thompson has identified larvae to species from the January 2015 CalCOFI cruise. We are on schedule to finish morphological identification of larvae from all present samples in 2016.

Going forward, the larval fish metabarcoding efforts will be led by a postdoc, Dovi Kacev. We (Thompson, Miller) have secured two and a half years of funding for Kacev from the Southern California Coastal Water Research Project Authority (SCCWRP) and NOAA SWFSC. SCCWRP is responsible for monitoring and assessing impacts of oceanic wastewater discharge in southern California and is deeply interested in new genetic techniques to monitor nekton. Kacev began working on the project in January 2016 and has laid the groundwork to shortly begin testing the metabarcoding techniques.

C. Implement a tradeoff framework that optimizes allocation of sampling effort.

A major goal of the project is to compare the effectiveness of different methods of biodiversity sampling, exploring how the optimal mix of sampling depends on the goals of monitoring and the resources available. We will also consider how the optimal mix will change as technology improves (e.g., lower cost and better accuracy of genetic and computational methods). Much of this work will not begin until biodiversity data derived from a range of sampling methods can be assimilated and analyzed. However, we have started developing the modeling framework for these analyses, testing methods on simulated biodiversity datasets. Development of methods for analyzing tradeoffs will continue in year 3, and analysis of real data will begin as the data become available. Co-PI Andrew Rassweiler is leading this part of the project and will continue to do so in his new capacity as an Assistant Professor at Florida State University. This transition was facilitated by issuing a subaward to FSU, and will ultimately increase value to the project by allowing Rassweiler to recruit a postdoc with salary support saved as a result of his new position.

Data Management and Communication

On the SBC MBON web site, we have installed system for dataset display, and begun designing draft products from integration of data from four nearshore time-series sampling programs (the 'integrated biodiversity dataset', above). The integration project is expected to generate two types of products: a) basic taxonomic information, b) spatial abundance. The website (http://sbc.marinebon.org/) has draft examples of each. As with the integration code itself, metadata and data formats of these draft datasets can be tested for
extensibility and reuse by applying their principles to the results of the next integration step: the addition of deep-water benthic time-series.

SBC MBON organized a cross-MBON DMAC workshop, 2016 June 7-9. The goals were to (a) establish preconditions for cross-MBON harmonization by comparing expected re-use scenarios with local processing and management systems, and (b) outline potential frameworks for important variables or vocabularies. Workshop products (anticipated by 2016-07-01) include:
(1) Document describing anticipated data needs of potential data customers with high-level requirements, using NMS condition reports as an example (2-3 pages),
(2) Schematic of local processing streams and how these translate to a ‘pipeline’ of data for a customer, and (3) Initial list of dictionaries that have the potential to meet MBON needs for a structured measurement vocabulary.

Within the broader community, the SBC MBON is integrating its activities into the international Group on Earth Observations Biodiversity Observation Network (GEO BON) by joining GECOBON Working Group 8 (Data integration and interoperability, informatics and portals), and adhering to established guidelines for data management and dissemination.

**Outreach and education**

SBC BON is again participating in a 2-day workshop this summer for sixth grade teachers that will be teaching the new ocean science curriculum starting this year. The Santa Barbara School district has chosen to focus their entire 6th grade science curriculum on ocean science, and the goal is to help to support the teachers in their understanding of some of the basic science content. They have chosen to use a curriculum called the Ocean Science Sequence (OSS) developed by the Lawrence Hall of Science (http://www.erf.org/erf2015). There are 3 units in the curriculum that focus on 3 areas:

*Ocean-Atmosphere Interactions*
*Flow of Carbon through land, ocean and atmosphere*
*Causes and Effects of Climate change*

Miller will present an overview of the Biodiversity Observation Network concept and how it links to these areas of ocean science. Debora Iglesias-Rodriguez will address climate change, its link to ocean acidification, and the range of possible effects on marine organisms.

We (Emmett Duffy, Miller, Muller-Karger) organized a session at the 2016 ASLO Aquatic Sciences meeting on Ocean Observing and Data Management, titled *MBON Voyage: Integrating marine biodiversity into ocean observing systems.*

O'Brien presented an invited poster in an informatics session during the 2016 Ocean Sciences Meeting (New Orleans). The session, "OD24B: Science at Sea:
Marine Data Stewardship from Proposal to Preservation," was organized by Shawn R Smith, Cynthia L Chandler, Karen Stocks, and Robert A Arko.

Miller and Reed participated in the Channel Islands National Marine Sanctuary Condition Report meeting in June, and SBC MBON collated and contributed biodiversity data to inform the report.

We (Siegel, Muller-Karger, Schimel) organized an NCEAS workshop "Prospects and Priorities for Satellite Monitoring of Global Marine Biodiversity" working group with Dave Siegel, David Schimel, Frank Davis, and Ryan Pavlick.

Other project presentations are listed below.

**Presentations:**


Henderikx Freitas, F. and D.A. Siegel. Controls on phytoplankton and particle distributions in the Santa Barbara Channel, California. Poster presented at the NASA Biodiversity and Ecological Forecasting Team Meeting, Silver Spring, MD, May 2016.


Miller, R.J., 2016: Demonstrating an Effective Marine Biodiversity Observation Network in the Santa Barbara Channel. Oral presentation made at the NASA Biodiversity and Ecological Forecasting Team Meeting, Silver Spring, MD, May 2016.

Miller, R.J. et al., 2016: Developing an Effective Marine Biodiversity Observation Network in the Santa Barbara Channel. Poster presented at the NASA Biodiversity and Ecological Forecasting Team Meeting, Silver Spring, MD, May 2016.


Miller, R.J. and E. Duffy, 2016: EOVs, EBVs and global MBON. Oral presentation made at the U.S. Marine Biodiversity Observation Network (MBON) All-Hands Meeting, Silver Spring, MD, May 2016.


Publications:


