PREDICTING SPATIAL AND TEMPORAL RESPONSES OF HERBIVOROUS CORAL REEF FISHES TO SEDIMENT RUNOFF

DATA POLICY COMPLIANCE

The project investigators will comply with the data management and dissemination policies described in the NSF Award and Administration Guide (AAG, Chapter VI.D.4) and the NSF Division of Ocean Sciences Sample and Data Policy (NSF 17-037). In accordance with these policies, the project investigators will deposit all metadata, datasets and derived products to the Biological and Chemical Oceanography Data Management Office (BCO-DMO) within two years of collection.

DESCRIPTION OF DATA TYPES

This study will produce the following types of data:
Objective A – Field observational: • Survey site location and time, • Fish identification, abundance, and diversity, • Fish size distributions and biomass, • Benthic cover (i.e., proportion of branching corals, massive corals, plate corals, encrusting corals, macroalgae, epilithic algal matrix/turf algae, crustose coralline algae, sand, rubble, and 'other'), • Fish foraging rates per minute across survey sites and time, • Videos of surveys and fish foraging across sites and time, • Environmental parameters (i.e., suspended and settled sediment, water temperature, oxygen saturation, light levels, and salinity).
Objective B – Laboratory experimental: • Experimental treatment conditions (i.e., suspended and settled sediment, water temperature, oxygen saturation, light levels, and salinity) when determining grazing rates and sediment avoidance by herbivorous reef fishes, • Fish identification and size/mass in each treatment and trial, • Meal mass before and after feeding trials, • Video recordings of fish foraging in each feeding trial, • Total number of foraging bites per hour. • Total amount of time spent active versus inactive, • Swimming speeds and distances, • Video recordings of fish movement across treatment conditions in the preference/avoidance system, • Proportion of time each fish spends in each treatment condition in the preference/avoidance system.
Objective C – Simulations: • Python code to implement predictive model in Jupyter notebooks, • Simulation outputs.

DATA AND METADATA FORMATS AND STANDARDS

We will share data and metadata in a standards-compliant fashion, including data manipulation algorithms. All data will be collected and compiled following the best practices guidelines of the British Ecological Society for management of ecological data. Other than video and imagery, all data will be stored in .csv format or converted to .csv. Videos and images will be stored in its native format (e.g.,mpg4, jpeg. Metadata (i.e., date, time, location, experimental treatment, etc.) will be prepared in accordance with Ecological Metadata Language (EML) conventions. Codes for statistical analyses will be primarily written in R and the full models and codes will be stored in its native .rds format. All code for the predictive model will be written in Python and stored as Jupyter notebooks (.ipynb). Model output will be stored as numpy arrays (.npy) and csv files (.csv).

DATA STORAGE AND ACCESS DURING THE PROJECT

While the project is ongoing, data will be stored within the UH IT Data Center, which provides bulk storage of research data and digital media. This storage is handled by an on-campus storage cloud, which provides advanced encryption technology to ensure the safety, security, and integrity of the stored data. GitHub will be used for version control throughout the project. We implement the principles of GNU General Public License (GNUGPL) for all factual data generated. User account data that contain private information (e-mail address, user name) will not be released to third parties. Copyright of media generated through this project will be retained by the creators of such media. PIs Johansen and McManus will oversee data management implementation for this project, with input from UH-IT specialists. The lab will retain data until publication, but upon publication of scientific results, products will be disseminated to the scientific community through presentations in scientific meetings and through various forms of social media, as appropriate. All data will be made publicly accessible within GitHub repositories and BCO-DMO within two years of collection (following NSF guidelines), which

simultaneously ensures data security and enables version control of open access data (following best practices for reproducible science).

MECHANISMS AND POLICIES FOR ACCESS, SHARING, RE-USE, AND RE-DISTRIBUTION

GitHub maintains editable versions of data and code, and allows maintenance of the unique version of a dataset that was used to generate a particular publication, thereby ensuring permanently reproducible science. Upon acceptance for manuscript publication, the associated repository will be archived on Zenodo with a DOI. We encourage re-use of our data and predictive framework once it is publicly available and envision it used for meta-analyses of large-scale responses in fishes and by managers needing region- or species-specific data. Data generated under this project will be administered in accordance with UH and NSF policies.
All data will be accessible to the public and subject to usage and dissemination restrictions under the CC-0 Public Domain Dedication License.

PLANS FOR ARCHIVING

UH Data Center will be used for long-term cloud-based archival and preservation of all data. Data persistence is assured by data replication in the UH Data Center. Additionally, we will coordinate with the BCO-DMO to ensure that project data are submitted to the appropriate national data archive (no later than two years from data collection).

ROLES AND RESPONSIBILITIES

The project's PI Jacob Johansen and Co-PI Lisa McManus will be responsible for all of the data management. It is Johansen and McManus's responsibility to make sure all of the project team members are taught the proper data management skills and uphold the data management requirements. Johansen and McManus may delegate data management duties to the laboratory project postdoctoral associate and graduate students responsible for collecting and analyzing data for Objectives A, B and C, respectively.

PLANNED RESEARCH OUTPUTS

SOFTWARE - "PYTHON CODE FOR OBJECTIVE C PREDICTIVE MODEL"

Code for the numerical integration of the Coral-Algal-Herbivore ordinary differential equations model described in Objective C of the proposal. Primary inputs include demographic rates and initial values (percent cover or biomass) for the following state variables: coral, macroalgae, turf algae, scrapers, grazers, and browsers. Secondary inputs include relationships (empirically derived or hypothetical) between demographic rates and suspended and/or settled sediment. Outputs include percent cover or biomass over time for all state variables.

PLANNED RESEARCH OUTPUT DETAILS

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Python code for Objective C predictive model	Software	Unspecified	Open	GitHub Zenodo Biological and Chemical Oceanography Data Management Office		Creative Commons Zero v1.0 Universal	None specified	No	No