**Title**

MSEC: Queryable global layers of environmental and anthropogenic variables for marine ecosystem studies

**Authors**

Lauren A. Yeager1,2\*, Philippe Marchand1\*, David A. Gill1,3, Julia K. Baum4# and Jana M. McPherson5,6#

**Affliations**

1National Socio-Environmental Synthesis Center, Annapolis, Maryland 21401, USA

2Department of Marine Science, University of Texas at Austin, Port Aransas, TX, 78373, USA

3 Luc Hoffmann Institute, World Wildlife Fund International, 1196 Gland, Switzerland

4Department of Biology, University of Victoria, Victoria, BC, V8W 2Y2, CA

5Centre for Conservation Research, Calgary Zoological Society, Calgary, AB, T2E 7V6, CA

6Department of Biological Sciences, Simon Fraser University, Burnaby, BC, V5A 1S6, CA

\*both authors contributed equally

#both authors served as principal investigators for this work.

**Corresponding author**: Lauren Yeager (lyeager@utexas.edu)

**METADATA**

**CLASS I. DATA SET DESCRIPTORS**

**A. Data set identity:** Environmental and anthropogenic covariates for marine ecosystem studies

**B. Data identification code:** msec\_npp\_mean.nc, msec\_npp\_min.nc, msec\_npp\_max.nc, msec\_npp\_sd.nc, msec\_npp\_sdinter.nc, msec\_npp\_flag.nc, msec\_wave\_mean.nc, msec\_wave\_sd.nc, msec\_wave\_sdinter.nc, msec\_wave\_windfetch.nc, msec\_wave\_resolution.nc, msec\_reefarea\_15km.nc, msec\_landarea\_15km.nc, msec\_reefarea\_200km.nc, msec\_landarea\_50km.nc, msec\_distmarket.nc, msec\_humanpop\_20km.nc, msec\_humanpop\_50km.nc

**C. Data description:** This dataset contains environmental and anthropogenic variables that are relevant for a suite of marine ecosystem processes and functions. The variables presented were calculated to represent ecologically-relevant measurements of biophysical conditions, landscape context, and human impacts for marine ecosystems and were selected to complement existing data products. Variables were derived based on remote-sensing products and open-access global data products. Data are provided as zipped global layers and are also queryable for point locations through a complementary web application.

**1. Originator(s):**

Lauren A. Yeager

National Socio-Environmental Synthesis Center, Annapolis, Maryland 21401, USA; Department of Marine Science, University of Texas at Austin, Port Aransas, TX, 78373, USA

Philippe Marchand

National Socio-Environmental Synthesis Center, Annapolis, Maryland 21401, USA

David A. Gill

National Socio-Environmental Synthesis Center, Annapolis, Maryland 21401, USA;

Luc Hoffmann Institute, World Wildlife Fund International, 1196 Gland, Switzerland

Julia K. Baum

Department of Biology, University of Victoria, Victoria, BC, V8W 2Y2, CA

Jana M. McPherson

Centre for Conservation Research, Calgary Zoological Society, Calgary, AB, T2E 7V6, CA; Department of Biological Sciences, Simon Fraser University, Burnaby, BC, V5A 1S6, CA

**2. Abstract**

Biophysical conditions, including climate, environmental stress, and habitat availability, are key drivers of many ecological processes (e.g., community assembly and productivity) and associated ecosystem services (e.g., carbon sequestration and fishery production). Furthermore, anthropogenic impacts such as coastal development and fishing can have drastic effects on the structure and function of marine ecosystems. Scientists need to account for environmental variation and human impacts to accurately model, manage, and conserve marine ecosystems. Although there are many types of environmental data available from global remote sensing and open-source data products, some are inaccessible to potential end-users because they exist as global layers in high temporal and spatial resolutions which require considerable computational power to process. Additionally, coastal locations often suffer from missing data or data quality issues which limit the utility of some global marine products for coastal sites. Herein we present the Marine Socio-Environmental Covariates (MSEC) dataset for the global oceans which consists of environmental and anthropogenic variables summarized in ecologically relevant ways. The dataset includes four sets of environmental variables related to biophysical conditions (net primary productivity models corrected for shallow-water reflectance, wave energy including sheltered-coastline corrections) and landscape context (coral reef and land cover within varying radii). We also present two sets of anthropogenic variables, human population density (within varying radii) and distance to large population center, which can serve as indicators of local human impacts. We have paired global, summarized layers available for download with an online data querying platform that allows users to extract data for specific point locations with finer control of summary statistics. In creating these global layers and online platform, we hope to make the data accessible to a wide array of end-users with the goal of advancing marine ecosystem studies.

**D. Key words:** net primary productivity, wave energy, land area, coral reefs, human impacts, climate, human population density, habitat connectivity, coastal systems

**CLASS II. RESEARCH ORIGIN DESCRIPTORS**

**A. Project description**

**1. Identity:** A database of marine environmental and anthropogenic layers related to biophysical conditions, landscape context, and human impacts.

**2. Originator(s):** as above.

**3. Period of study:** 1979-2020

**4. Objectives:** Provide relevantsocio-environmental data suitable for analyses of marine ecosystem studies in forms that enhance ease of use for potential end-users.

**5. Abstract:** as above.

**6. Source(s) of funding:**

This work was funded by the National Socio-Environmental Synthesis Center (SESYNC), though funding received from the National Science Foundation (NSF) grant DBI-1052875, and builds on ideas first developed by the Environment and Organisms Working Group supported by the National Center for Ecological Analysis and Synthesis (NCEAS), a Center funded by NSF (Grant #EF-0553768), the University of California, Santa Barbara, and the State of California. DAG was supported by a Luc Hoffmann Institute Post-Doctoral Fellowship. JKB was supported by an NSERC Discovery Grant and a Sloan Research Fellowship. JMM benefitted from a National Sciences and Engineering Council of Canada (NSERC) postdoctoral fellowship. We thank the authors of original data sources for their contributions of open data used in these products including: ESRI, NOAA CoastWatch, NOAA NOS, and NOAA's NWS Monterey Regional Forecast Office for producing the NPP data. We also thank Greg Guannel for helpful suggestions pertaining to fetch and wave energy calculations.

**B. Specific subproject description**

**1. Site description:** Data were obtained from satellite measurements of the ocean (net primary productivity), climatology model hindcast reanalyses (wave energy), global habitat (coral reef area) and land layers (land area), human population models, and locations of world cities (geographic extent: 90°S to 90°N, 180°W to 180°E).

**Site type:** Ocean coordinates.

**Habitat:** All marine surface waters, ranging from coastal systems to the open ocean.

**2. Research methods**

We created gridded, global layers of ecologically relevant biophysical and anthropogenic variables, the Marine Socio-Environmental Covariates (MSEC) dataset, thought to be important in regulating marine ecosystem structure and function. We focused on variables that we deemed to be high priority in studies of marine ecosystems to complement other existing products (e.g., BioOracle Tyberghein et al. 2012; MARSPEC Sbrocco and Barber 2013). The variables derived from seven primary data sources: (1) NOAA CoastWatch modelled primary productivity estimates (<http://coastwatch.pfeg.noaa.gov/>), (2) global coral reef habitat area maps from Reefs at Risk Revisited (Burke et al. 2011), (3) shorelines and land areas derived from the Global, Self-consistent, Hierarchical, High-resolution Shoreline database (GSHHS v2.3.5, http://www.ngdc.noaa.gov/mgg/shorelines/gshhs.html; Wessel and Smith 1996), (4) a high-resolution, global bathymetry model (SRTM30\_PLUS V6.0; http://topex.ucsd.edu/WWW\_html/srtm30\_plus.html; Becker et al. 2009), (5) NOAA WAVEWATCH III hindcasts based on gridded wind data from climate reanalysis (Chawla et al., 2013), (6) global models of human population counts from Gridded Population of the World, V3-4 (CIESN 2005, 2016), and (7) the ESRI World Cities data set.

For global, gridded layers, we calculated all variables on a 2.5 arcmin grid (~4.6 km at the equator). Where appropriate, grid cells corresponding to a land mask based on the GSHHS v2.3.5 dataset were set to NA. However, we found in our research that nearshore points are often classified as land in global layers at a similar resolution. This is because raster cells are typically assigned as land or ocean based on the mid-point value of the cell, and thus coastal cells with some water will be classified as land approximately half of the time within land masks, including cells with midpoints up to ~2.3 km away from land. To reduce this precision error, we created a high-resolution raster from the GSHHS full-resolution shoreline (high-resolution for Antarctica), with a grid cell size of 0.25 arcmin (1/10th of our standard grid, ~ 460 m resolution) and then considered a cell within the standard 2.5 arcmin grid to be water if any of the cells within the given grid cell were classified as water in the high-resolution grid. This minimized missing data values for coastal cells in our layer, although it will in some cases result in assigning marine covariate values to cells that are mostly land. This should not be problematic, however, as it is assumed that users are querying data for marine locations only. Note also that this was done only for assigning land-containing grid cells as NA within the land mask, and not for land area calculations described below.

To enhance ease of use and allow users increased control over derived variables, we paired these global, gridded derived layers with an online platform (MSEC: Marine Socio-Environmental Covariates) that allows users to query data for specific point locations located at [shiny.sesync.org/apps/msec](file:///C%3A%5CUsers%5CLauren%5CDropbox%5CMarine_covars_paper%5Cshiny.sesync.org%5Capps%5Cmsec). Specific research methods for deriving each variable are detailed below, as well as possible user settings on the online platform.

***Biophysical variables***

*Net primary productivity*

Net primary productivity (NPP) was extracted from 8-day composite layers from 2003-2013 produced by NOAA CoastWatch (http://coastwatch.pfeg.noaa.gov/erddap/griddap/erdPPbfp28day.graph?productivity). NPP was modelled on a 2.5 arcmin grid based on satellite measurements of photosynthetically available radiation (NASA's SeaWiFS), SST (NOAA's National Climatic Data Center Reynolds Optimally-Interpolated SST), and chlorophyll a concentration (NASA’s Aqua MODIS; <http://coastwatch.pfeg.noaa.gov/erddap/griddap/erdPPbfp28day.html>) (Behrenfeld and Falkowski, 1997). In order to account for poor data quality in shallow sites due to bottom reflectance, we filtered out grid cells with minimum depth <30 m based on the STRM30 plus bathymetry layer (0.5 arcmin resolution, http://topex.ucsd.edu/WWW\_html/srtm30\_plus.html) following Gove et al. (2013). For cells with missing data following depth filtering, we interpolated values from the closest three surrounding cells within a 125km search radius. For shoreline cells (those defined as within 9.3 km, ~ 2 grid cells, of land), we gave preference to cells closer to shore which should be more similar in NPP to the missing cells than cells farther offshore. We searched for cells with data within 125km of the focal cell but preferentially chose cells within 9.3 km of shore (MARSPEC, Sbrocco and Barber 2013) or a reef location (as low-lying atolls we sometimes not recorded as land, reefs defined based on Reef at Risk Revisited Base Layer, Burke et al. 2011), if these were present within our specific search distance. This allowed us to improve data quality, and still capture the higher NPP that characterizes coastal zones. The method used for interpolation is recorded for each grid cell within the msec\_npp\_flag layer.

If missing data values within annual time series were concentrated in certain parts of the years (e.g., during warmer months, rainy seasons, or seasonal peaks in productivity), this could bias annual estimates of NPP. To minimize potential bias, we first calculated summary statistics (e.g., mean, standard deviation, min, max) for each sampling week (8-day composites) within each grid cell across all years in the time series. We then calculated corresponding intra-annual summary statistics based on these weekly summary values within years. We report intra-annual mean, minimum, maximum, and standard deviation in NPP, as well as the standard deviation of annual means (inter-annual s.d.). Each of these summary statistics is queryable in the MSEC online platform by latitude/longitude.

*Wave energy*

The WAVEWATCH III hindcast dataset (http://polar.ncep.noaa.gov/waves/CFSR\_hindcast.shtml) is the product of the namesake wave forecasting model based on wind input from the National Centers for Environmental Prediction’s Climate Forecast System Reanalysis (CFSR) (Chawla et al., 2013). The significant height and peak period and direction of waves, as well as the speed and direction of wind, is available at a 3 hour temporal resolution for a span of 31 years (1979-2009). It is composed of 14 nested grids including a 30 arcmin global grid, along with 10 arcmin and 4 arcmin grids for specific coastal areas.

For every point and time step in the original dataset, we calculated the wave energy flux (the power transmitted per unit of wavefront width) as:

$E\_{W}=\frac{ρg^{2}}{64π}H\_{s}^{2}T\_{p}$ ,

where *H*s is the significant wave height, *T*p the peak wave period,  the water density (998 kg m-3 at 20°C) and *g* the gravitational acceleration (9.81 m s-2). We calculated the mean wave energy (kW/m) for each year and each day of the year, and from these values produced three aggregate layers (mean, intra-annual standard deviation and inter-annual standard deviation) for each of the 14 WAVEWATCH III grids. Finally, we transferred the values from each grid to our final 2.5 arcmin grid using bilinear interpolation (with the ‘resample’ function in the raster package). For each cell in our final grid, we kept the value from the highest-resolution WAVEWATCH III grid covering that cell.

Due to the relatively low spatial resolution of the WAVEWATCH III data, especially the global grid, the values obtained by the method above might be inaccurate for sheltered locations, i.e. those separated from the open sea by land obstacles at a range less than the size of wave model cell. To determine which cells are sheltered, we calculated the fetch (distance from the point to the nearest shoreline) along 16 equally-spaced bearings (22.5 degree apart) for all cells with a distance to land < 50 km. To obtain a more robust estimate along each of the 16 bearings, we computed fetch for 9 sub-bearings (2.5 degrees apart and centered on the main bearing) and took the average of these values, weighted by the cosine of the angle between the sub-bearing and main bearing (Sharp et al. 2016). Sheltered locations were those for which more than half (8/16) fetch lengths were under a threshold based on the WAVEWATCH resolution at that point (50 km for the 30 arcmin grid, 18.3 km for 10 arcmin grids and 7.3 km for 4 arcmin grids).

At sheltered locations, we estimated the height and period of locally-generated waves using the method described by Resio et al. (2003):

$$u\_{f}^{2}=0.001\left(1.1+0.035U\_{10}\right)U\_{10}^{2}$$

$$\frac{gH\_{s}}{u\_{f}^{2}}=min\left(0.0413\left(\frac{gF}{u\_{f}^{2}}\right)^{{1}/{2}},211.5\right)$$

$\frac{gT\_{p}}{u\_{f}}=min\left(0.651\left(\frac{gF}{u\_{f}^{2}}\right)^{{1}/{3}},239.8\right)$,

where *F* is the fetch, *U10* is the wind speed at 10 m above sea level and *u*f is the friction velocity. The two terms inside the minimum functions represent the fetch-limited and fully-developed sea conditions, respectively. As recommended by Resio et al. (2003), we also implemented a limit for the peak wave period in shallow waters (with depth *d*):

$T\_{p\left(max\right)}=9.78\sqrt{\frac{d}{g}}$.

Wind speeds and directions were extracted from the highest-resolution WAVEWATCH III grid covering the point, and we used for each time step the fetch value of the bearing nearest to the actual wind direction. Depth value at each point was extracted from the bathymetry layer described above. The functions to calculate fetch and wind energy from the variables available from WAVEWATCH III were implemented as part of the waver R package by a subset of the authors (Marchand and Gill, 2016).

After calculating the wave energy at each time step for each sheltered point, we computed the same time aggregates as above and incorporated them in the corresponding cells of the final layers. We included a “wind-fetch” flag layer to indicate these sheltered locations where wave energy was computed from wind and fetch values rather than from the WAVEWATCH III height and period values.

***Landscape variables***

*Coral reef connectivity*

For global layers, we calculated coral reef connectivity as the total amount of coral reef area within the surrounding landscape at two buffer distances: 15 km and 200 km. The 15 km buffer distance represents the upper range of larval dispersal distances for most reef fishes, which are largely constrained to 5-15 km (Green et al., 2015). The larger buffer distance of 200 km corresponds to the upper end of larval dispersal distances and home range size for large-bodied species (although longer dispersal distances are possible, but rare, Green et al. 2015). Reef area was estimated from the 500 m resolution coral reef map from the Reefs at Risk Revisited data set created by the World Resources Institute (Burke et al. 2011). For our global grid, area estimates were made by creating a buffer of the corresponding size (15 km or 200 km) around the midpoint of each grid cell using the ‘gBuffer’ function from the rgeos package (Bivand and Rundel, 2016) projected to the Cylindrical Equal-Area (Lambert) projection (Central Meridian: -160. Datum: WGS 1984; same projection as reef layer) in R (R Core Team 2016). The reef layer was then clipped to the buffer using the ‘extract’ function in the raster package (Hijmans, 2016a) and the area in km2 estimated as the number of reef cells falling with in the buffer multiplied by the area of a cell (0.25 km2). For the MSEC online platform, buffers will be drawn from the user-input geographic coordinates and user can specify their own buffer distance at 1 km intervals ranging from 1 to 200 km.

*Land area*

Terrestrial nutrient and energy flows into marine ecosystems may impact local productivity and food web structure. Previous studies have found that nutrient inputs from land-derived sources are commonly detectable within primary producers with up to 15 km from shore (Lapointe and Clark 1992). However, riverine plumes may bring terrestrial-derived dissolved organic nutrients 50 km or more from the coast (Delvin and Brodie 2005). In most cases, effects of terrestrial sediment run-off are limited to within ~10 km of shore (Fabricius 2005, Devlin and Brodie 2005). Therefore we calculated the amount of land area within two buffer distances (15 km and 50 km) as one metric of the magnitude of terrestrial subsidies into marine ecosystems.

Similar to reef area estimates, a buffer around the midpoint of each raster cell was drawn in azimuthal equidistant projection centered on the point (to give true distances) and then projected to the WGS84 geographic coordinates system (coordinate system of the land area raster). We then extracted and summed the area of all land-based grid cells (from the 0.25 arcmin, GSHHS-derived land raster) within the buffer with the ‘area’ and ‘extract’ functions of the raster package. As with reef area, user may specify their own buffer distance for land area at 1 km intervals ranging from 1 to 200 km in the MSEC online platform.

***Anthropogenic variables***

*Human population density*

Many human impacts on marine systems, like coastal development, nutrient loading, and harvest pressure, are expected to scale with human population size. Human population density, coastal hydrology, economic development, and the nature of local fisheries will all determine the geographic scale over which the local human population impacts marine ecosystems. For example, a 20 km radius will capture the travel distance that most fishers take in subsistence/artisanal fisheries (Clark et al. 2002, Chuenpagdee et al. 2006) as well as the scale over which localized land-use change may have the strongest impacts on nutrient loading and sedimentation (see above). A 50 km radius may correspond to the upper limit of small-scale or semi-commercial coastal fisheries (Chuenpagdee et al. 2006) and watershed-scale impacts of nutrient loading and sedimentation on coastal biogeochemistry (Delvin and Brodie 2005).

 To assess human population size, we used the Gridded Population of the World, Population Count Grid v3 (GPWv3; CISEN 2005) and Gridded Population of the World v4 (GPWv4) Population Count Adjusted to Match 2015 Revision of UN WPP Country Totals (CISEN 2016) datasets produced by the Socioeconomic Data and Applications Center (SEDAC). These data layers consist of global grids at a resolution of 0.5 arcmin (~1 km). Population count estimates are provided for years 1990 and 1995 (GPWv3) and years 2000, 2005, 2010, 2015, and 2020 (GPWv4). To estimate the human population count within 20 km and 50 km radii of each grid cell, we created buffers of the corresponding radius from the mid-point of each grid cell from our base raster layer projected in an azimuthal equal-area projection. We then extracted all grid cells within the corresponding SEDAC population layer that fell within the buffer and summed the population counts within the extracted cells. We report the data as total human population count within a given radius, but data could be converted to average population density be dividing by 1,256 km2 or 7,850 km2 for the 20 km and 50 km buffers, respectively. Again, these global layers are queryable in the MSEC online platform based on user-input latitude/longitude at radii ranging from 1 to 200 km.

*Distance to provincial capital*

In additional local human population density, access to national and international seafood markets can be an important predictor of fishing effort and target species. Following Cinner et al. (2013), we used distance to provincial capital as a standardized metric of market access, as provincial capitals are likely to provide fishers access to national and export markets. We used the ESRI World Cities data set to identify provincial capitals. To calculate geodestic distances from the midpoint of each raster grid cell to the nearest provincial capital, we used the ‘distGeo’ function in the geosphere package (Hijmans, 2016b). This calculation can be performed for user-defined points in the MSEC online platform.

A limitation of this metric is that it evaluates only straight-line distance to provincial capital and does not take into account other factors that would surely affect fisher access to a given location including travel time, available technology, and physical characteristics of the coast (Marie et al. 2016). Nonetheless, straight-line distance to the provincial capital has been found to be an important predictor of the condition of reef fisheries in broad-scale studies (Cinner et al. 2013).

**CLASS III. DATA SET STATUS AND ACCESSIBILITY**

**A. Status**

**1. Latest update:** 10 January 2017.

**2. Latest archive date:** 10 January 2017.

**3. Metadata status:** up to date as of 10 January 2017.

**4. Data verification:** up to date as of 10 January 2017.

**B. Accessibility**

**1. Storage location and medium:** Data are uploaded as zipped data layers in the corresponding supplement through Ecological Archives. All data are also queryable through the MSEC website hosted at shiny.sesync.org/apps/msec.

**2. Contact person(s):** Lauren Yeager or Philippe Marchand

**3. Copyright restrictions:** None.

**4. Proprietary restrictions:** None.

**5. Costs:** None.

**CLASS IV. DATA STRUCTURAL DESCRIPTORS**

**A. Data set files**

**1. Identity:** The files’ name template is

msec\_[var]\_[subvar].nc, where 'var' is the covariate (e.g., npp, wave energy) and 'subvar' is either the specific statistic, the radius size, etc.

**2. Size:** Each file is up to 81 megabytes, depending on the variable and its number of sub layers (see Table 1).

**3. Format and storage mode:** TheMSEC dataset is provided in netCDF format in Supporting Information to this *Ecology* publication, as well as being queryable on the MSEC online platform. Each file is a global grid representing a temporal or spatial summary for a given variable/sub-variable combination as detailed in Table 1.

**B. Variable information**

**Variable identity and definition:** See Table 1.

**C. Data anomalies**

Errors within the global data layers arises from three primary sources: 1. Measurement and process error in original data products, 2. Bias in missing data values within original data products, and 3. Errors association with grid resolution and spatial interpolation. We further discuss each of these error types below.

***Measurement and process error in original data products***

NPP and wave metrics were modelled as a function of other measured biophysical variables (see methods). Any process error associated with these models will be propagated to our summary measures. Assumptions and sources of error for each model can be found in the manuscripts corresponding to model development for each (NPP model described in Behrenfeld and Falkowski 1997; wave forecasting model described Chawla et al. 2013).

Unequal sampling and accuracy of sampling across the globe will also lead to variation in data accuracy in base layers for reef area and human population density (Burke et al. 2011, Doxsey-Whitfield et al. 2015). For example, for SEDAC GPW datasets, census unit size varied across countries and continents, yielding the highest resolution data for North America and Oceania relative to other continents. Furthermore, the resolution on the base grid for the GPWv4 increased relative to GPWv3 (cell size decreased to 30 arc-sec from 2.5 arc-min) and the number of census input units increased dramatically in GPWv4 from GPWv3, leading to generally better accuracy in the GPWv4 data set (Doxsey-Whitfield et al. 2015).

***Bias in missing data values within original data products***

Models of NPP are derived from remote sensing products mapping sea surface temperature, Chlorophyll a concentrations, and irradiance. However, like all remote sensing products, missing data values are common when cloud and ice cover obscure the ocean surface from satellite view. Ice cover is obviously only an issue for locations closer to poles, but a high number of NA values from polar regions may limit the utility of NPP models for these areas and often lead no NPP estimates for the high latitude regions. We have included an additional grid of the number of 8-day intervals (out of 45 total) with no data for NPP across the times series for each grid cell as a metrics of data quality/quantity (Figure 1). Again, polar regions suffered from high numbers of 8-day intervals with no data and hence data from higher latitudes should be used with caution. If seasonal patterns in regional climate cause cloud cover to be highest in certain seasons, this could bias annual estimates if this was not taken into account. As described in the methods section, we first summarized NPP metrics across 8-day intervals for each grid cell across years before deriving inter-annual summary statistics to reduce this potential bias. Many coastal cells also suffered from missing data points or were filtered out due to shallow depths. The interpolation of coastal cells from surrounding grid cells as described above limited some of this bias, although in areas with wide, shallow continental shelves (e.g., Gulf of Mexico or U.S. East Coast) cells used for interpolation came from farther offshore. The msec\_npp\_flag layer indicates where cells used for interpolation came from nearshore cells (within 9.3 km of the coast) or farther away.

***Errors association with resolution and spatial interpolation***

For most of the variables considered herein, they exhibit in situ variation at smaller scales than the resolution of the original data layers and our gridded data products (especially for the biophysical and landscape variables). For example, water column NPP likely varies over the scales of meters or less and these global layers will not pick up small-scale peaks in NPP associated with coastal eutrophication. Thus, these global layers will be best suited for comparison across larger landscape and biogeographic gradients than for quantifying differences among sites for studies with small spatial extent. As mentioned in the methods, we replaced the original NPP values for shallow cells with the mean of nearby cells with depth greater than 30m. There were also bands of missing cell values around the prime meridian in the original data layers that thus required interpolation in derived NPP layers. All values interpolated are indicated in the NPP flag layers.

For the wave energy layers, we included a “wind-fetch” flag layer to indicate sheltered locations where wave energy was computed from wind and fetch values (see methods) rather than from the WAVEWATCH III height and period values. Given the multi-grid nature of the WAVEWATCH data, we also included a layer indicating the original grid resolution for each cell in our wave energy layers.

The global layers of reef and land area, human population, distance to market and wave energy (for sheltered points) all depend on distances calculated from the midpoint of each cell. Therefore, using the global layers to estimate their values at other points will induce uncertainties that depend on the specific interpolation method used (nearest neighbor, bilinear, etc.). Note that for all these metrics, except wave energy and NPP, the MSEC online platform uses the exact input points provided when computing distances. True distance to market is also dependent on the existing road network between the shoreline and the market or population center, which we did not calculate here (see Cinner et al 2016).

Finally, there may be some small error associated with land area estimates based on performing the calculations in the original geographic coordinate system of the data. Specifically, area calculations were done using the ‘area’ function in the raster package. The ‘area’ function uses a flat geometry approximation to compute the surface area of unprojected cells by multiplying cell height (which is constant among all cells) by cell width (which varies with latitude) of the midpoint of each cell. Ignoring the curvature of the earth will introduce some error into these estimates, which will be most pronounced at higher latitudes, but minimized when cells are small. At 83° N (approximate latitude of the northernmost point in Greenland), our cells are about 170m x 460m and the flat geometry calculation should be quite accurate. In fact, for these northern most cells, cell area estimations were found to vary only ~0.6% between calculations using the ‘area’ function versus projecting raster cells to an equal area projection. Furthermore, transforming the GSHHS to an equal area projection would have introduced its own small distortions to the original shorelines and thus we preserved the original projection of the land layers for the area calculations.

**CLASS V. SUPPLEMENTAL DESCRIPTORS**

**A. Data acquisition**

**1. Data forms or acquisition methods:** N/A.

**2. Location of completed data forms:** N/A.

**3. Data entry verification procedures:** N/A.

**B.** **Quality assurance/quality control procedures:** See Class IV.C – Data anomalies section. Additionally, accuracy of all code and data layers were checked by the first two authors (LAY and PM).

**C.** **Related material:** N/A

**D. Computer programs and data-processing algorithms:**

Data processing methods are described above and the R scripts used to process the data layers are available at:  <http://github.com/pmarchand1/msec>.

**E. Archiving:** N/A.

F. **History of data set usage:**

Gill, D., Mascia, M.B., Ahmadia, G., Glew, L., Lester, S., Barnes, M., Craigie, I., Darling, E., Free, C., Geldmann, J., Holst, S., Jensen, O., White, A.T., Basurto, X., Coad, L., Gates, R.D., Guannel, G., Mumby, P.J., Thomas, H., Whitmee, S., and S. Woodley, Fox, H.E. In review. Capacity shortfalls hinder the performance of marine protected areas globally. *Nature*.

Robinson, J.P.W., Williams, I.D., Edwards, A.M., McPherson, J.M., Yeager, L.A., Vigliola, L., Brainard, R.E., and Baum, J.K. 2017. Fishing degrades size structure of coral reef fish communities. *Global Change Biology* **23**: 1009-1022.

Yeager, L.A., Deith, M.C.M., McPherson, J.M., Williams, I.D., and Baum, J.K. Under revision. Scale dependence of environmental controls on the functional diversity of coral reef fish communities. *Global Ecology and Biogeography*.

**Literature Cited**

Behrenfeld, M.J., and P.G. Falkowski, 1997. A consumer's guide to phytoplankton primary productivity models. *Limnology and Oceanography* **42**:1479-1491.

Bivand, R. and Rundel, C. 2016. rgeos: Interface to Geometry Engine - Open Source (GEOS). R package version 0.3-20. https://CRAN.R-project.org/package=rgeos.

Burke, L. M., K. Reytar, M. Spalding, and A. Perry. 2011. Reefs at risk revisited. World Resources Institute Washington, DC.

Center for International Earth Science Information Network - CIESIN - Columbia University, and Centro Internacional de Agricultura Tropical - CIAT. 2005. Gridded Population of the World, Version 3 (GPWv3): Population Density Grid. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4XK8CG2. Accessed October 17, 2016.

Center for International Earth Science Information Network - CIESIN – Columbia University. 2016. Documentation for the Gridded Population of the World, Version 4 (GPWv4). Palisades NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4D50JX4 Accessed 17 October 2016.

Chawla, A., Spindler, D.M. and Tolman, H.L. 2013. Validation of a thirty year wave hindcast using the Climate Forecast System Reanalysis winds. *Ocean Modelling* **70**:189-206.

Chuenpagdee R, Liguori L, Palomares MD, Pauly D (2006) Bottom-up, global estimates of small-scale marine fisheries catches. *Fisheries Centre Research Report* **14**:1–112.

Cinner, J.E., Graham, N.A., Huchery, C. and Macneil, M.A., 2013. Global effects of local human population density and distance to markets on the condition of coral reef fisheries. *Conservation Biology* **27**:453-458.

Cinner, J. E., Huchery, C., MacNeil, M. A., Graham, N. A. J., McClanahan, T. R., Maina, J., … Mouillot, D. (2016). Bright spots among the world’s coral reefs. *Nature* 1–17. https://doi.org/10.1038/nature18607.

Clark, B.M., Hauck, M., Harris, J.M., Salo, K. and Russell, E., 2002. Identification of subsistence fishers, fishing areas, resource use and activities along the South African coast. *South African Journal of Marine Science* **24**:425-437.

Devlin, M. J. and Brodie, J. 2005. Terrestrial discharge into the Great Barrier Reef Lagoon: nutrient behavior in coastal waters. *Marine Pollution Bulletin* **51**: 9-22.

Doxsey-Whitfield, E., MacManus, K., Adamo, S.B., Pistolesi, L., Squires, J., Borkovska, O. and Baptista, S.R., 2015. Taking advantage of the improved availability of census data: a first look at the gridded population of the world, version 4. *Papers in Applied Geography*, **1**:226-234.

Fabricius, K. E. 2005. Effects of terrestrial runoff on the ecology of corals and coral reefs: review and synthesis. *Marine pollution bulletin* **50**:125-146.

Green, A. L., Maypa, A. P., Almany, G. R., Rhodes, K. L., Weeks, R., Abesamis, R. A., Gleason, M. G., Mumby, P. J. and White, A. T. (2015), Larval dispersal and movement patterns of coral reef fishes, and implications for marine reserve network design. *Biological Reviews* **90**:1215–1247.

Gove, J.M., Williams, G.J., McManus, M.A., Heron, S.F., Sandin, S.A., Vetter, O.J. and Foley, D.G., 2013. Quantifying climatological ranges and anomalies for Pacific coral reef ecosystems. *PloS one* **8**:e61974.

Hijmans, R.J. 2016a. raster: Geographic Data Analysis and Modeling. R package version 2.5-8. https://CRAN.R-project.org/package=raster.

Hijmans, R.J. 2016b. geosphere: Spherical Trigonometry. R package version 1.5-5. https://CRAN.R-project.org/package=geosphere.

Lapointe, B. E. and Clark, M. W. 1992. Nutrient inputs from the watershed and coastal eutrophication in the Florida Keys. *Estuaries* **15**:465-476.

Marchand, P. and Gill, D.A. 2016. waver: Calculate Fetch and Wave Energy. R package version 0.2-0. [https://CRAN.R-project.org/package=waver](https://CRAN.R-project.org/package%3Dwaver).

Maire, E., Cinner, J., Velez, L., Huchery, C., Mora, C., Dagata, S., Vigliola, L., Wantiez, L., Kulbicki, M. and Mouillot, D., 2016. How accessible are coral reefs to people? A global assessment based on travel time. *Ecology letters*, **19**:351-360.

 R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Resio, D. T., Bratos, S. M., & Thompson, E. F. (2003). Meteorology and Wave Climate, Chapter II-2. Coastal Engineering Manual. US Army Corps of Engineers*,* Washington DC, 72pp.

Sharp, R., Tallis, H.T., Ricketts, T., Guerry, A.D., Wood, S.A., Chaplin-Kramer, R., Nelson, E., Ennaanay, D., Wolny, S., Olwero, N., Vigerstol, K., Pennington, D., Mendoza, G., Aukema, J., Foster, J., Forrest, J., Cameron, D., Arkema, K., Lonsdorf, E., Kennedy, C., Verutes, G., Kim, C.K., Guannel, G., Papenfus, M., Toft, J., Marsik, M., Bernhardt, J., Griffin, R., Glowinski, K., Chaumont, N., Perelman, A., Lacayo, M. Mandle, L., Hamel, P., Vogl, A.L., Rogers, L., and Bierbower, W. 2016. InVEST +VERSION+ User’s Guide. The Natural Capital Project, Stanford University, University of Minnesota, The Nature Conservancy, and World Wildlife Fund.

Sbrocco, E.J. and Barber, P.H., 2013. MARSPEC: ocean climate layers for marine spatial ecology. *Ecology* **94**:979-979.

Tyberghein, L., H. Verbruggen, K. Pauly, C. Troupin, F. Mineur, and O. De Clerck. 2012, Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography* **21**:272–281. doi: 10.1111/j.1466-8238.2011.00656.x

Wessel, P. and Smith, W.H., 1996. A global, self-consistent, hierarchical, high-resolution shoreline. *Journal of geophysical Research* **101**:8741-8743.

Table 1. Data variable descriptors.

|  |  |  |  |
| --- | --- | --- | --- |
| NetCDF file name | Description | Units | Data Type\* |
| msec\_npp\_mean.nc | Net Primary Productivity of Carbon, Overall Mean | mg C m-2 day-1 | real |
| msec\_npp\_min.nc | Net Primary Productivity of Carbon, Mean Annual Minimum | mg C m-2 day-1 | real |
| msec\_npp\_max.nc | Net Primary Productivity of Carbon, Mean Annual Maximum | mg C m-2 day-1 | real |
| msec\_npp\_sd.nc | Net Primary Productivity of Carbon, Intra-Annual Standard Deviation | mg C m-2 day-1 | real |
| msec\_npp\_sdinter.nc | Net Primary Productivity of Carbon, Inter-Annual Standard Deviation | mg C m-2 day-1 | real |
| msec\_npp\_flag.nc | NPP Flag: not interpolated (0), interpolated from nearby coastal/reef (1) or other (2) cells | n/a | integer |
| msec\_wave\_mean.nc | Wave Energy Flux, Overall Mean | kW m-1 | real |
| msec\_wave\_sd.nc | Wave Energy Flux, Intra-Annual Standard Deviation | kW m-1 | real |
| msec\_wave\_sdinter.nc | Wave Energy Flux, Inter-Annual Standard Deviation | kW m-1 | real |
| msec\_wave\_windfetch.nc | Wave Energy Calculated From Wind and Fetch (binary: 0/1) | n/a | integer |
| msec\_wave\_resolution.nc | WAVEWATCH III Resolution: 30 arcmin (1), 10 arcmin (2) or 4 arcmin (3) | n/a | integer |
| msec\_reefarea\_15km.nc | Coral Reef Area within a 15 km Radius | km2 | real |
| msec\_landarea\_15km.nc | Land Area within a 15km Radius | km2 | real |
| msec\_reefarea\_200km.nc | Coral Reef Area within a 200 km Radius | km2 | real |
| msec\_landarea\_50km.nc | Land Area within a 50 km Radius | km2 | real |
| msec\_distmarket.nc | Distance to Market - Nearest Provincial Capital | km | real |
| msec\_humanpop\_20km.nc | Human Population Count within a 20 km Radius | individuals | integer |
| msec\_humanpop\_50km.nc | Human Population Count within a 50 km Radius | individuals | integer |
| \* Note: All real variables are 4-byte floating-point numbers (6 significant digits) |

Figure 1. Number of 8-day windows over which no data was recorded 11-year time series for NPP.

