2.2 The GOMWIR Inventory

M. R. Besonen[1], K. Withers[2], G. Gold-Bouchot[3]

[1]Harte Research Institute for Gulf of Mexico Studies, Texas A&M University-Corpus Christi, [2]Department of Life Sciences, Texas A&M University–Corpus Christi, [3]Department of Oceanography and Geochemical and Environmental Research Group (GERG), Texas A&M University

2.2.1 Introduction and Background

The US Department of the Interior Bureau of Ocean Energy Management (BOEM) has a long-standing research program on marine ecosystem science with relevance to energy management for the northern GOM, and the information is extensive and widely available. Recognizing the need to understand the Gulf as a single LME instead of just regional seas marked by political boundaries, one of the main objectives of the first GOM Workshop on International Research (GOMWIR) was to gather similar information about the state of marine ecosystem science for the SGOM. One specific GOMWIR goal to support this objective (see 1.1 Preface and 1.2.1 Introduction by Larry McKinney in this volume for further details about GOMWIR goals) was to develop an inventory of existing marine ecosystem science for the southern portion of the Gulf. This exploratory effort, the GOMWIR Inventory, would serve as a foundational document for the workshop to provide an initial assessment of the science, which in turn would help identify gaps in knowledge, and inform discussions and planning during the workshop, and into the future. This paper provides information about the considerations that guided the development of the Inventory, some of the technical details behind the effort, and an analysis and discussion of the results that were gathered, and the process in general.

2.2.1.1 Guidance for the Development of the GOMWIR Inventory

The GOMWIR Inventory was conceived as a comprehensive, cross-disciplinary inventory of LME research that would include:

1. An annotated bibliography of peer-reviewed literature, reports and other publications,
2. An annotated listing of Mexican research programs, and
3. An annotated listing of Mexican data sources.

Each of these annotated resources would address one or more of the three BOEM thematic areas of interest, i.e. Baseline Studies, Fates and Effects Studies, or Environmental Monitoring Studies (see 1.2.1 Introduction, this volume), and be focused on the waters of the SGOM. “Annotated” in this context meant including key words, geographic locations, key variables measured or described, and electronic resource locators (URLs, DOIs, etc.), as available.

Though no particular format for these annotated materials was required, a design goal for the GOMWIR Inventory was that it should match the structure of the preexisting BOEM EcoSpatial Information Database (ESID, or “ee-sid”) web application as much as possible. The ESID was a web-based database that stored full-text marine ecosystem literature and data sets in support of BOEM’s National Environmental Policy Act (NEPA) requirements (Madsen et al. 2014). In addition to regular text searches, it had an easy-to-use mapping interface for visual/geospatial searching. The system used PostgreSQLas a backend database and had comprehensive schema with the main resource table composed of approximately 40 metadata fields per bibliographic record (some mandatory, some optional). The ESID was already populated with several thousand records from the United States Atlantic and Gulf Coasts in areas under BOEM’s purview. Thus, the aspiration for matching the ESID structure was that the GOMWIR Inventory could eventually serve as a pluggable data module for the ESID, but with data for the SGOM.

Several important realities also guided development of the GOMWIR Inventory, and especially the design goal of matching the preexisting BOEM ESID structure:

1. Time frame: Very little time was available to actually develop the Inventory between the project start date (September 2016) and the actual Workshop event (March 2017). This time frame also squarely included the end-of-year holiday season with extended break time in both the United States and Mexico. No preexisting solution was available—simply the ESID database schema. Thus, we were charged with developing and implementing an international data collection effort with colleagues from the ground up.
2. Distributed effort: The Harte Research Institute (HRI) and Center for Coastal Studies at Texas A&M University—Corpus Christi (TAMUCC) had a significant amount of knowledge about certain elements of SGOM ecosystem science and databases from extensive work in the past, much of it related to the efforts and research of Dr. J. "Wes" Tunnell. But it was also clear that we were missing many resources of Mexican origin that were not necessarily indexed in international databases, nor directly available to us. Therefore, it was absolutely essential to coordinate the data gathering efforts with Mexican colleagues who either did have direct access to material of interest, or at least knew of its existence.
3. Rigorous nature of ESID records: As mentioned above, the ESID had a comprehensive database schema with approximately 40 metadata fields per bibliographic record. This comprehensive documentation was necessary given that the ESID's purpose was to support BOEM's NEPA requirements. Beyond the comprehensive metadata, every ESID record required an OCR'd or digital-born PDF copy of the reference being cited (personal communication, J. Blythe, BOEM Environmental Studies Program, Scientific Data Manager), which was an impossibility for GOMWIR Inventory given the time frame and distributed effort realities mentioned above. Thus, the goal for the GOMWIR Inventory was to match the ESID's metadata fields as completely as possible.

2.2.2 Methods

Efforts to gather and populate the GOMWIR Inventory with information about bibliographic materials, research programs, and data sources focused on the SGOM occurred both locally, and via critical partnering with a team of Mexican colleagues. At TAMUCC, the efforts were undertaken with the assistance of Justine Thomas (Research Assistant). The Mexican colleagues with whom we partnered had a wide range of expertise and experience in the SGOM. These international colleagues included Dr. Adolfo Gracia Gasca and León Felipe González Morales of UNAM (Mexico City), Dr. Victor Manuel Vidal Martínez and Daniel Aguirre Ayala of CINVESTAV (Mérida, Yucatán), Dr. Sharon Herzka and Mónica Cecilia Mozqueda Torres of CICESE (Ensenada, Baja California), and Dr. Eustorgio Meza Conde and Sergio Gabriel Jiménez of UAT (Tampico, Tamaulipas). We also established a basic anchor website for information, and widely announced the workshop itself along with a call for volunteer data contributions via mailing lists, presentations, and by personal requests to appropriately connected colleagues who could widen the circles of distribution.

Given the realities that guided the development of the GOMWIR Inventory as discussed earlier, it was decided that a mechanism for data collection using web-based electronic forms would provide the best compromise instead of e-mail document exchanges between multiple parties. We developed a solution using Google Apps Script (GAS) to produce stand-alone, web-based data entry forms that would collect data into a backend Google Sheets spreadsheet. We used this GAS solution because it was readily available to us, benefitted from Google's resilient infrastructure, and also allowed scientists with different technical skill levels to deal with the data in a familiar form (i.e., either a simple web form or spreadsheet).

An AJAX-based mechanism (Asynchronous JavaScript and XML) was used to submit data from the forms into the spreadsheet, but also preserve the form's content for reuse or resubmission instead of having to reenter a very comprehensive set of metadata from scratch. There was a two-fold rationale behind this decision. First, we suspected that many data contributions by a single person would be related; for example, different chapters from the same book, or papers by the same series of authors, or with the same focus, key words, geographic footprint, etc. At the same time, we also anticipated possibly receiving contributions from other areas like Cuba, where Internet connectivity is often patchy or problematic. The AJAX-based mechanism, by preserving form contents between submission attempts—whether they are successful or not—would allow contributors to either reuse or resubmit a form with an intact set of comprehensive metadata instead of ending up with a blank form and having to laboriously reenter all the extensive data from scratch.

Collected data were first inspected for duplicate records, and then cleaned and prepared by a mix of semiautomated and manual editing methods. Cleaning and preparation was not needed for fields with a controlled vocabulary, i.e. fields represented by check boxes or radio buttons on the web forms, but it was necessary for any free text entry field. This cleaning and preparation did not focus on content, but simply considered the mechanical and formatting aspects of the contributed data. For example, many data were apparently entered by copying and pasting text directly between PDF documents and the web forms. This resulted in errant line breaks, gibberish or extraneous characters, and other odd defects. As the same time, certain fields were reformatted to a specific and consistent convention instead of the freeform styles that were received (for example, all author names were reformatted into "Lastname, Initials of first name" format).

After data cleaning and preparation, an attempt was made to standardize the content of the geographic footprint metadata field. Whereas ESID geographic footprints included both textual descriptions and GIS polygons delineating the geospatial extent of study areas, the realities of the GOMWIR Inventory data collection effort only allowed for textual descriptions. These textual descriptions were provided in a variety of formats including as geographic coordinates, physiographic or bathymetric descriptions, political descriptions, and were also mixed between Spanish and English. Thus, this metadata field was split into three separate subfields and standardized formats to facilitate both plotting onto maps, and text-based searches.

A simplified subset of data (just seven fields) was extracted for presentation at the GOMWIR event in March 2017. Data were presented to the workshop participants in simple list form for the three breakout sessions, and also plotted on three large maps (80" x 48", ~2 m x ~1.2 m) for visualization. The maps also served as the basis for a live, on-the-fly data gap analysis exercise during the workshop. For this exercise, workshop participants were provided with small stickers with personally-identifying numbers to link the stickers back to them. Participants who knew of other SGOM resources that had not yet been identified and included in the GOMWIR Inventory could affix their stickers to the maps so they could be contacted after the workshop to provide further details about the particular resources that were missing from the Inventory. Following the workshop, the location of each of these stickers on the three maps was digitized and linked back to the participants who were individually solicited by e-mail to provide further details about the data and resources missing from the Inventory.

2.2.3 Results

After duplicate record removal, the GOMWIR Inventory ended up capturing 897 bibliographic references, 33 records about research programs, and five records about data sources. The bibliographic records were almost exclusively entered into the Inventory via the web forms either under our own local effort (~68% of total records), or by the international participants (~32% of total records) with whom we had partnered, i.e., there were only two volunteer contributions of data, despite our advertising efforts. The totals above only include information that was submitted via the web forms as data was suitably

parsed into specific metadata fields, and generally complete enough to be useful. In turn, in several instances, typical bibliographic/publication lists were supplied, but these lists do not provide the comprehensive metadata that is needed to make the information immediately actionable for inclusion into the GOMWIR Inventory. For example, such bibliographic and/or publication lists are missing basic metadata characteristics like categorization by one of the three main BOEM thematic areas, eight resource categories, key words, or geographic footprint covered by the resource. The bibliographic references provided via these lists are being assessed as time permits, and if it is possible to fill in basic metadata gaps, these resources will be included in future revisions of the Inventory.

The actual GOMWIR Inventory in spreadsheet format is provided as a special electronic addendum to this document. The intended final format of the Inventory is not a spreadsheet, and this is further discussed in the Discussion/Analysis section of this document.

In the following paragraphs, summarized results from the GOMWIR Inventory collection effort are provided using simple percentage statistics. To facilitate summarizing the bibliographic references, which were contributed in large numbers, the bibliographic references are split into three subsets by BOEM thematic area. In turn, the smaller number of records received about research programs and data sources means that no splitting is necessary.

2.2.3.1 Bibliographic Contributions by Baseline Studies Thematic Area

The vast majority of annotated bibliographic contributions fell under the Baseline Studies thematic area (571 out of 897 total records, or ~64%; see Figure 45 for map of locations). In comparison, the Fates and Effects Studies thematic area was flagged for just ~19% of the records, and the Environmental Monitoring Studies thematic area for ~28% of the records. It is noted that references could be flagged for more than one of the thematic areas; thus, the percentages sum to greater than 100%.

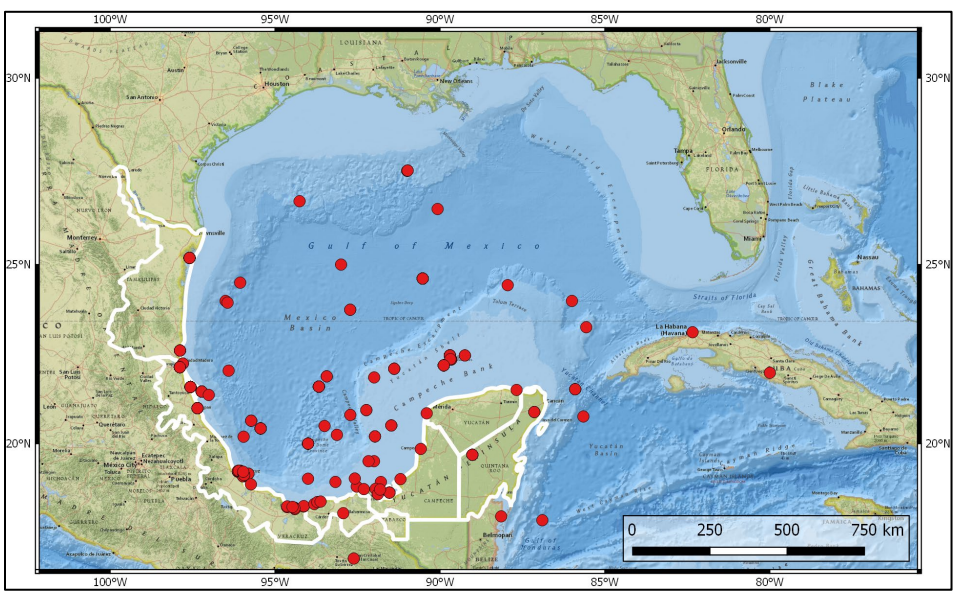

Figure 45. Locations of Baseline Studies in bibliographic records.
Note that the scale of this map, and the fact that many records share common geographic place names or coordinates (see Discussion section) precludes showing all 571 points separately. Most red dots on the map actually represent multiple bibliographic records from the Inventory. Basemap from ESRI Basemap layers collection.

With regard to the eight possible resource categories for Baseline Studies contributions, “Pelagic ecology” (~36%) was the most commonly indicated category followed by “Infauna/Meiofauna” at ~29% (Figure 46, panel A). “Water quality” and “Coral and Hardbottom” were the next most commonly indicated categories at ~17% and ~16%, respectively. Finally, “Physical processes,” “Geology,” and “Demersal fish” all hovered around the 10% mark, while “Seagrass” was only indicated for ~6% of the contributions.

“Journal article” was the most common (~40%) type of bibliographic resource documented in the Inventory followed by “Dissertation/Thesis” at ~20% and “Abstract” at ~17% (Figure 46, panel B). In decreasing order from ~8% down to ~1%, other resource types were specified as “Conference Proceedings,” “Book, Section/Chapter,” “Report,” “Book, Whole,” and “Other.” Only one “Map” and zero “Web Page”-types were indicated for records in the Inventory.

The bulk of reported Baseline Studies bibliographic contributions are from the 1980s (~30%) and 1990s (~24%) (Figure 46, panel C). Contributions from each of the two decades prior (1960s and 1970s), and each of the two subsequent decades (2000s and 2010–present), also each account for about ~10% of the inventory. Very little (~3%) of the literature is from before the 1960s.

With respect to geographic distribution by Mexican (MX) state, the vast majority (~49%) of contributed literature records are associated with Veracruz, with the state of Yucatán being the next most commonly indicated at ~12% (Figure 46, panel D). Very little of the literature is linked to Tamaulipas or Campeche, just ~6% and ~8%, respectfully. Tabasco and Quintana Roo are hardly represented at ~2% apiece. On a related note about geographic footprints, a simple frequency analysis of the ten most commonly indicated geographic place names for records in the Baseline Studies thematic area is provided in Table 4.

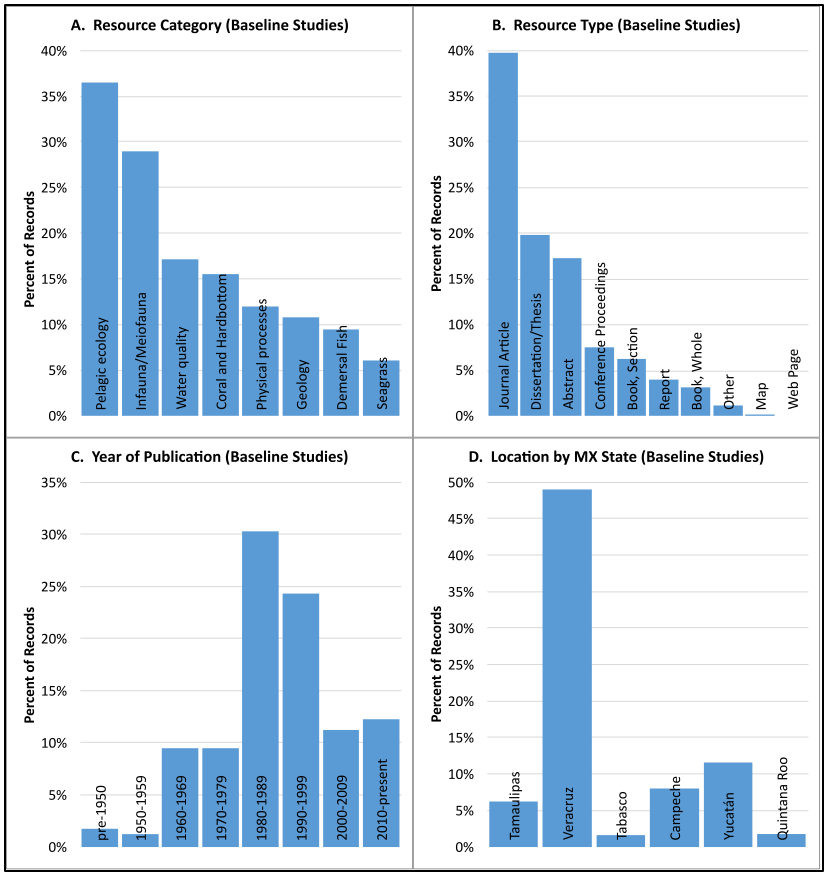

Figure 46. Simple percentage statistics for Baseline Studies bibliographic records.

Table 4. Top 10 geographic place names indicated for Baseline Studies bibliographic contributions.

Name	Counts
Laguna de Tamiahua	99
Laguna de Tampamachoco	67
Gulf of Mexico	46
Laguna Pueblo Viejo	29
Laguna Madre	25
Alacrán Reef	20
Southern Gulf of Mexico	18
Campeche Bank/Yucatán Shelf	17
Veracruz	16
Enmedio Reef	12

2.2.3.2 Bibliographic Contributions by Fates and Effects Studies Thematic Area

The GOMWIR Inventory ended up with 166 records (~19% of 897 total records; see Figure 47 for map of locations) of bibliographic resources that were linked to the Fates and Effects Studies thematic area."Water quality" was the most commonly indicated category (~47%) for these records followed by "Infauna/Meiofauna," "Physical processes," and "Pelagic ecology" at ~40%, ~34%, and ~25%, respectively (Figure 48, panel A). The "Coral and Hardbottom," "Demersal Fish," and "Seagrass" categories each accounted for about 10% of the records, while "Geology" was the least-mentioned category at 7%.Most of the records for Fates and Effects Studies bibliographic contributions were of type "Journal Article" (~45%) while ~23% of the records were linked to "Dissertation/Thesis"-type resources, and~13% to typical "Reports" (Figure 48, panel B). Regarding "Conference Proceedings," "Book, Whole," "Abstract," and "Book, Section/Chapter"-types, each of these accounted for ~5% each of the total Fates and Effects Studies bibliographic contributions. No "Map," "Other," or "Web Page"-types were indicated for any of the Fates and Effects Studies records.

The publication year results for Fates and Effects Studies bibliographic contributions is interesting (Figure 48, panel C). There are no records present that are from prior to the 1960s, then just ~1% in the 1960s, and ~4% in the 1970s. However, by the 1980s and 1990s, the quantity of southern Gulf Fates and Effects Studies literature increases tremendously, and these decades account for ~19% and ~22%, respectively, for literature records in this thematic area. The amount continues to expand moving forward in time, and some 28% of the Fates and Effects Studies literature captured by the GOMWIR Inventory was produced during the 2000's. Finally, in the eight years since that point, i.e., 2010–2017, ~27% of the literature can be linked to this decade.

Similar to the Baseline Studies literature records, Veracruz is the most commonly indicated Mexican state for Fates and Effects Studies literature at ~26% (Figure 48, panel D). Campeche State has the next most abundant focus at ~15% with Tabasco indicated slightly less frequently for ~12% of the Fates and Effects Studies records. Yucatán was the geographic focus of ~8% of the bibliographic contributions from this category whereas there was little literature associated with Tamaulipas (~2%); there are no records for this category from Quintana Roo. A simple frequency analysis of the ten most commonly indicated geographic place names for Fates and Effects-focused literature is provided in Table 5.

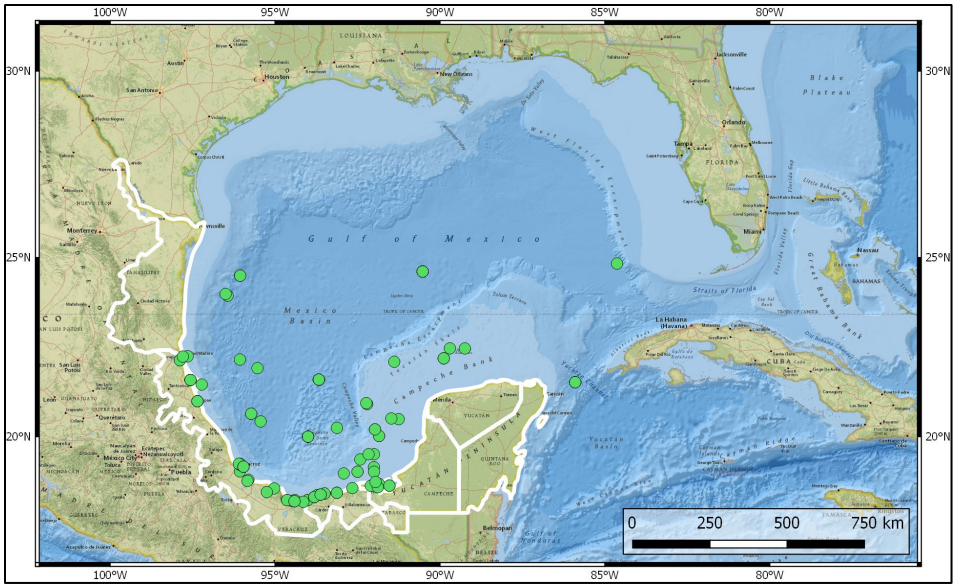

Figure 47. Locations of Fates and Effects Studies in bibliographic records.
Note that the scale of this map, and the fact that many records share common geographic place names or coordinates (see Discussion section) precludes showing all 166 points separately. Most green dots on the map actually represent multiple bibliographic records from the Inventory. Basemap from ESRI Basemap layers collection.

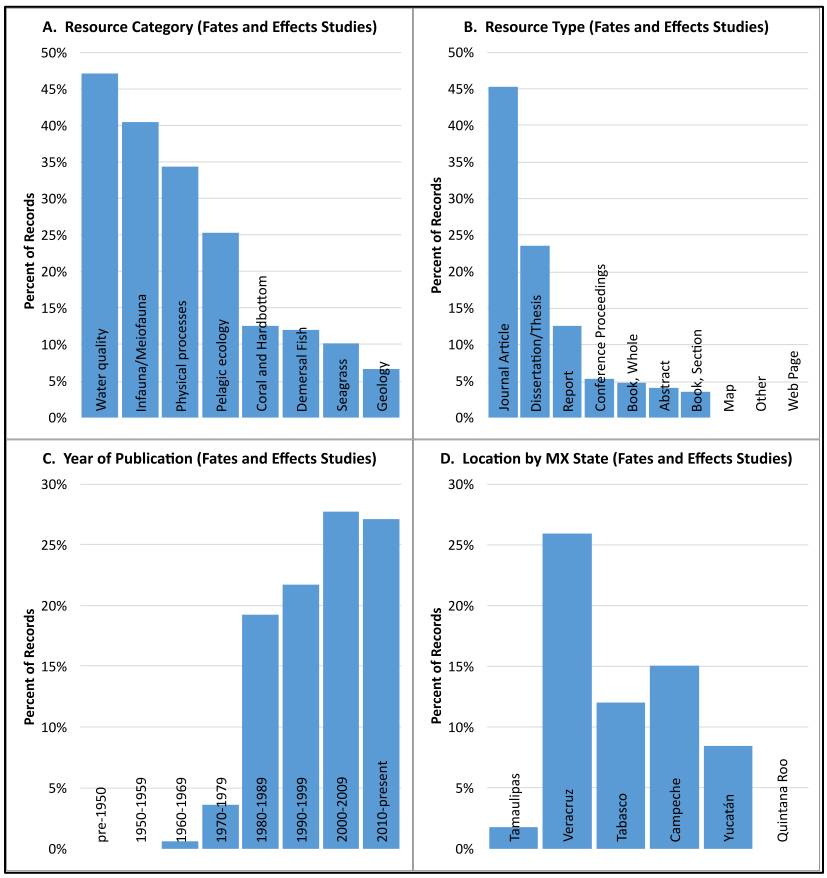

Figure 48. Simple percentage statistics for Fates and Effects Studies bibliographic records.

Table 5. Top 10 geographic place names Indicated for Fates and Effects studies bibliographic contributions.

Name	Counts
Gulf of Mexico	30
Southern Gulf of Mexico	13
Loop Current (Yucatán Channel)	8
Laguna de Tamiahua	6
Laguna de Tampamachoco	6
Sonda de Campeche	6
Veracruz Reef System	6
Bay of Campeche	5
Campeche Bank/Yucatán Shelf	4
Coatzacoalcos	3

2.2.3.3 Bibliographic Contributions by Environmental Monitoring Studies Thematic Area

The GOMWIR Inventory ended up with 250 records (~28% of 897 total records; see Figure 49 for map of locations) of bibliographic resources that were linked to the Environmental Monitoring Studies thematic area.

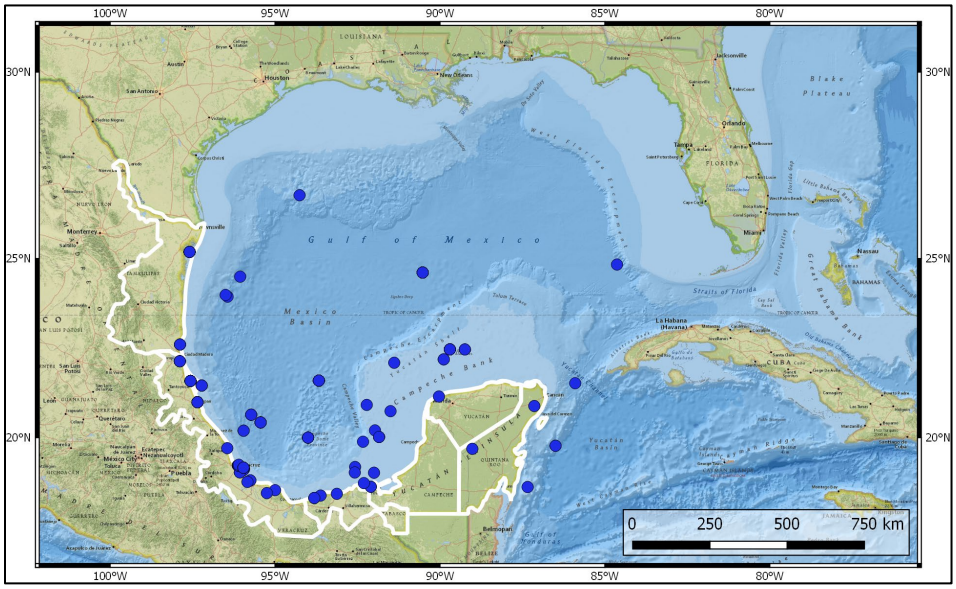

Figure 49. Locations of Environmental Monitoring Studies in bibliographic records.
Note that the scale of this map, and the fact that many records share common geographic place names or coordinates (see Discussion section) precludes showing all points separately. Most blue dots on the map actually represent multiple bibliographic records from the Inventory. Basemap from ESRI Basemap layers collection.

With respect to resource category, “Water quality” and “Pelagic ecology” together account for almost 75% of the Environmental Monitoring Studies literature at ~41% and ~33%, respectively (Figure 50, panel A). The “Infauna/Meiofauna” resource category is also well-represented with ~28% of the contributed literature records. “Coral and Hardbottom,” “Physical processes,” and “Demersal fish,” are indicated for ~16%, ~14%, and ~13% of bibliographic records. Finally, the “Seagrass” and “Geology” categories are the least commonly designated categories at just ~10% and ~7%, respectively.

The distribution of bibliographic resource types for literature in the Environmental Monitoring Studies thematic area is relatively distinct from the two other thematic areas. The “Journal Article”-type of resource is still the most common type at some ~27%, but the “Abstract”-type is the second most abundant at ~24% (Figure 50, panel B). “Dissertation/Thesis,” “Conference Proceedings,” and “Report”-types are indicated in approximately the same magnitude at ~13%, ~11%, and ~11%, respectively. Environmental Monitoring Studies literature designated as the “Book, Section/Chapter”-type accounts for ~6% of the records, while the remaining “Book, Whole,” “Other,” and “Web Page”-types all contribute less than 5% of records in this thematic area. Finally, as was the case for the other two thematic areas, there are no “Map”-type records registered in the Inventory.

Regarding the year of publication for bibliographic records in this thematic area, there are clear peaks in the decades of the 1980s (~28%) and the 1990s (~31%) (Figure 50, panel C). The earliest records in this Inventory thematic area are from the decade of the 1960s (~4%) and the 1970s (~6%). The 2000s and period from 2010 to present account for ~19% and ~11%, of the records.

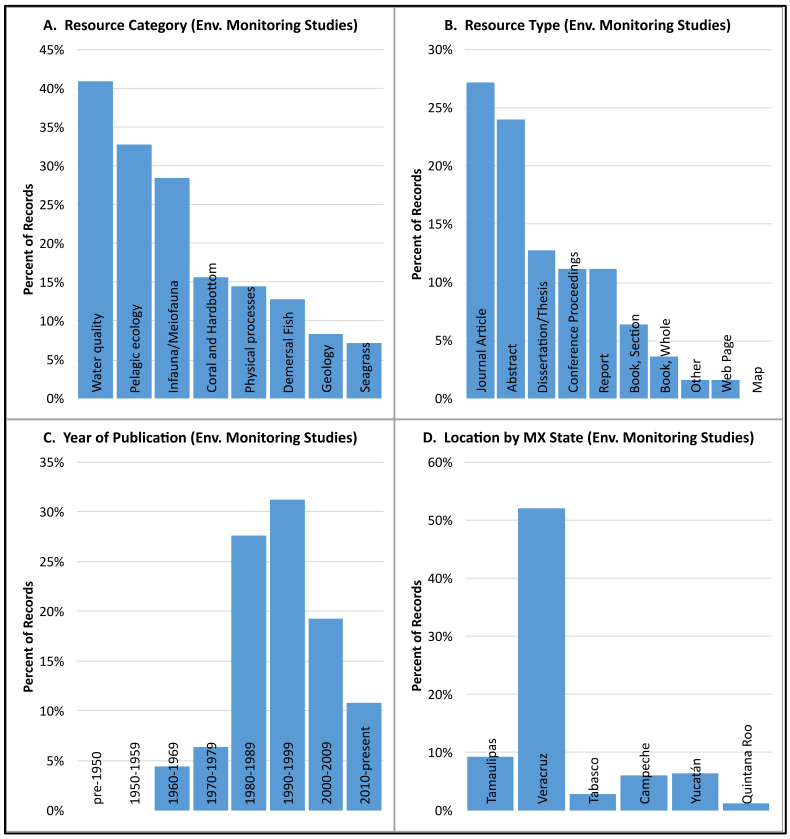

Figure 50. Simple percentage statistics for Environmental Monitoring Studies bibliographic records.

With respect to the distribution of Mexican states designated for the bibliographic records in this thematic area, Veracruz is again the most commonly specified state at ~52% (Figure 50, panel D). Tamaulipas is the second-most represented state at ~9% of the records, whereas the states of Campeche and Yucatán each account for ~6% of the contributed records in the Environmental Monitoring Studies thematic area. The states of Tabasco and Quintana Roo are poorly represented at ~3% and ~1%, respectively. A simple frequency analysis of the ten most commonly indicated geographic place names for Environmental Monitoring literature is provided in Table 6.

Table 6. Top 10 geographic place names Indicated for Environmental Monitoring studies bibliographic contributions.

Name	Counts
Laguna de Tamiahua	44
Laguna de Tampamachoco	36
Gulf of Mexico	19
Laguna Madre	19
Southern Gulf of Mexico	13
Veracruz Reef System	10
Alacrán Reef	8
Port of Veracruz	8
Laguna Pueblo Viejo	6
Campeche Bank/Yucatán Shelf	5

2.2.3.4 Inventory Results for Research Programs Focused on the Southern Gulf of Mexico

Data collection efforts for the GOMWIR Inventory captured 33 records about research programs focused on marine ecosystem science and related themes in the SGOM. The vast majority of research programs (28 out of 33) were classified as academic in nature, being the product of a university or an academic research organization. A small quantity (just four out of 33) was classified as governmental in nature, being a product of the government or a government agency, such as the Mexican Ministry of the Navy, the Ministry of Environment and Natural Resources, or similar. Finally, information about a single research program that was the product of a nongovernmental organization was also provided to the Inventory effort.

With respect to the organizational scope of these research programs, nine were identified as being broadly organized at the level of a consortium, collaboration, or working group of researchers at multiple organizations, institutions, or agencies. Another 12 of the research programs were mid-tier in size and identified as a collaboration or working group within a single organization, institution, or agency, such as multiple departments within the same organization. Finally, the remaining 12 research program contributions involved a working group at the departmental or lab level, including multiple researchers in the same department, a faculty member directing student research, or similar.

2.2.3.5 Inventory Results for Data Resources Focused on the Southern Gulf of Mexico

The GOMWIR Inventory data collection process received information about five data sources with marine ecosystem science content and related themes in the SGOM. These include:

1. CONABIO Integrated Publishing Toolkit for Biodiversity Data
2. CONACYT-SENER Hydrocarbon Fund Project 20144 (restricted availability, no URL provided)
3. IUCN Red List of Threatened Species
4. Florida Fish and Wildlife Conservation Commission Open Data System
5. REEF Volunteer Fish Survey Project

Data sources 1 and 2 were identified as academic, 3 and 4 as governmental, and 5 as nongovernmental, using the same definitions of these entities in the preceding section focused on Research Programs. All five data sources house biological data. Two of the data resources maintain additional datasets with one also housing oceanographic, water column, and chemical datasets (source 2), while the other includes oceanographic and geospatial data in addition to biological data (source 4). Three of the data sources supposedly allow remote electronic access to data with no mention of restrictions (sources 3, 4 and 5). Two of the data sources (1 and 2) provide data with restricted availability, in which the data resource is not private, but has some restriction for access like payment, membership in a certain organization, or some other factor.

2.2.3.6 Results in Sticker-Based Data Gap Analysis and Follow-Up

The live, on-the-fly, data gap analysis exercise using stickers on the three large BOEM thematic area maps finished with 495 data points being added to the three maps (Figure 51). Workshop participants placed stickers on the maps to indicate locations of known resources (bibliographic resources, data resources, etc.) that were not included in the GOMWIR Inventory presented at the workshop. Some miscommunication caused participants to indicate locations throughout the entire Gulf instead of just focusing on the southern portion, but nonetheless, all contributed sticker data points were maintained.

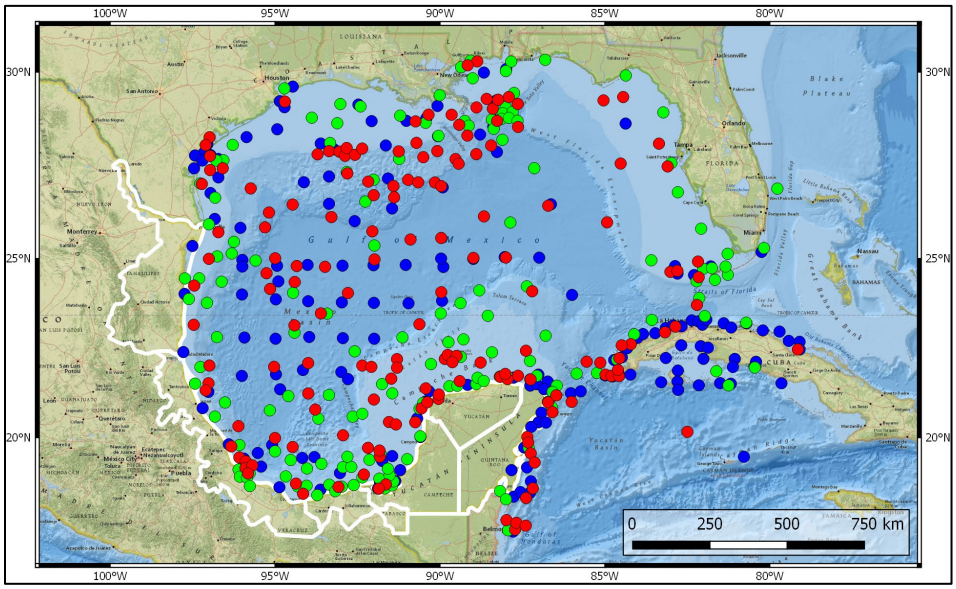

Figure 51. Locations of 495 resources not included in the GOMWIR Inventory but identified by the data gap analysis sticker exercise.
Red dots represent stickers placed on the Baseline Studies map, green dots represent stickers placed on the Fates and Effects Studies map, and blue dots represent stickers placed on the Environmental Monitoring Studies map. Basemap from ESRI Basemap layers collection.

There was a very equitable distribution between stickers on the three thematic area maps:

1. The **Baseline Studies** map received 165 out of 495 stickers, or ~33%,
2. The **Fates and Effects Studies** map received 157 out of 495 stickers, or ~32%, and
3. The **Environmental Monitoring Studies** map received 173 out of 495 stickers, or ~35% of the total.

The follow-up attempts to get basic data about these missing resources were, unfortunately, extraordinarily unsuccessful. Despite multiple individual solicitations by e-mail (i.e., no mass broadcast e-mails) to participants that had placed stickers, including a reminder of how many stickers that participant had placed in certain approximate locations, not a single reply was received to help fill in these gaps.

2.2.4 Discussion/Analysis

2.2.4.1 Patterns and Trends in the Bibliographic Data, and Insights about Knowledge Gaps

Because of the small number of contributions about research programs and data resources that were collected for the GOMWIR Inventory, attempting to draw any valid conclusions about marine ecosystem science in the SGOM based on those small datasets does not make sense. However, with the almost 900 bibliographic data records that were collected some observations can be made. We make the assumption that these almost 900 records provide a generally representative sample of the overall population of marine ecosystem science that has actually been undertaken in the SGOM. Some of the results from the Inventory suggest this is the case, but other results suggest these almost 900 records are probably biased for some obvious reasons as explained below.

One simple observation can be drawn about the distribution of records with respect to the three BOEM thematic areas. There were more than twice the number of contributions in the Baseline Studies thematic area (571 out of 897 total records, or ~64%) than in the Environmental Monitoring Studies thematic area (250 out of 897 total records, or ~28%), and more than three times the number of records than in the Fates and Effects Studies thematic area (166 out of 897 records or ~19%). The same general trend was evident in the interests of the researchers who wanted to attend the actual workshop event. Specifically, there were many more people who declared an interest and expertise in Baseline Studies than the other two thematic areas. Given the concordance of these trends, this suggests the almost 900 bibliographic references are indeed representative of the greater population of marine ecosystem science that exists for the region. Furthermore, the distribution between thematic areas is perhaps not too surprising given the long historical tradition of field surveys and exploration that can easily be related to Baseline Studies, and the potentially lower barrier to entry for this type of research depending on details. In turn, Fates and Effects Studies might be interpreted as more specialized and as having a potentially higher barrier to entrance given the need for specialized analytical equipment and methods. Depending on perspective, the overall distribution might suggest that efforts should be made to increase the numbers of Fates and Effects Studies in the future, if they are valued equally or greater than the other thematic areas.

With respect to resource categories, in all three thematic areas, resources focused on “Pelagic ecology,” “Infauna/Meiofauna,” and “Water quality” categories were always well-represented. In turn, resources focused on “Coral and Hardbottom” and “Physical processes” categories were generally represented at intermediate percentages. Finally, for all three thematic areas, resources focused on “Geology,” “Demersal Fish,” and “Seagrass” categories were never particularly well represented. Depending on

perspective, this might be seen as a strong suggestion that more efforts should be dedicated to work focused on the least-represented categories of "Geology," "Demersal Fish," and "Seagrass."

Some interesting trends can be seen in the Year of Publication results from the three thematic areas. Both Baseline Studies and Environmental Monitoring Studies show strong peaks in activity during the decades of the 1980s and 1990s with decreases from those earlier peak values over the last two decades. In turn, publications focused on Fates and Effects Studies ramp up slowly, but have a strong, increasing growth over the last four decades, with the numbers of studies peaking in the last two decades. Currently, there is not a good explanation about why this is the case.

Several interesting observations can be made about the geographic footprint results. With respect to qualifications by Mexican states, Veracruz is by far the most well-represented state in all three thematic area groups, with two to five times more representation than any other Mexican state. In turn, Quintana Roo is uniformly the least well-represented in all of the thematic areas. The very strong representation for Veracruz is probably related to the massive amounts of effort that have been dedicated to marine ecosystem science in and around the Veracruz Reef system through the years. In turn, poor representation of Quintana Roo is probably due to the fact that only a small corner of the state actually falls within the SGOM with the majority of the state facing the Caribbean Sea. Likewise, Tamaulipas and Tabasco have received less attention than Campeche and Yucatán, likely also reflecting the focus on coral reefs and the exploitation of oil reserves on Campeche Bank.

The top 10 geographic place names for each thematic area based on frequency analysis suggest that the Inventory has a geographic bias in some cases. For example, certain place names like Laguna de Tamiahua and Laguna de Tampamachoco have bubbled to the top of these frequency analyses, but other well-known places like Laguna de Términos have not. However, there is a clear and straightforward explanation for this. In particular, part of the local effort at TAMUCC involved an attempt at mining the voluminous SGOM literature collection amassed by Wes Tunnell over his career. The collection is organized on a thematic basis, for example, a folder of material is focused on Laguna de Tamiahua, another on Laguna de Tampamachoco, and so on and so forth. Given the expedited schedule for this exploratory effort, only a subset of these thematic folders was examined within the timeline for this project. The thematic folders for these two lagoons did make it into this exploratory effort, hence, their elevated frequency counts. This bias, therefore, simply results from the short amount of time that was available for developing the Inventory.

Some other observations from these frequency analyses emerge as well, in particular, the existence of place names that indicate very broad, generic areas like "Gulf of Mexico," or "Southern Gulf of Mexico," or "Campeche Bank." These results, together with the map figures, point out one of the weaknesses in the GOMWIR Inventory—the collection of geographic footprints—and this is further explained in the following section.

2.2.4.2 Geographic Footprints—An Opportunity for the Future

Besides the comprehensive textual metadata that composed the main body of each BOEM ESID record, we noted that GIS polygons delineating the geospatial extent of study areas were also included. However, given the realities of the GOMWIR Inventory data collection effort, the collection of GIS polygons for each record could not be accomplished for a couple of reasons.

First, because of the extremely short time frame for collecting the data, the effort was only meant to be exploratory. To define a detailed GIS polygon for any individual study or publication, it is imperative to carefully read the text to find the geographic footprint. In some cases, it is obvious because a publication includes a map indicating the specific region of interest. However, this is not the case for many other publications so a detailed reading is necessary. Also, much of the actual data entry was performed by

students who did not have had the necessary skill set or equipment/software to produce GIS polygons. Thus, our geographic footprints were provided as textual descriptions by geographic coordinates, physiographic or bathymetric place names or descriptions, and political descriptions.

A significant issue with collecting only textual information is that if it is not easy to determine a detailed geographic footprint for any particular record, substituting in a broad, generic place name like "Gulf of Mexico" or "Campeche Bank" is often the only solution. Thus, many database records end up sharing the same, broad generic place names for their geographic footprint. This has major detrimental effects on maps that attempt to plot this data. For example, this can be seen with the three BOEM thematic area maps (see Figure 45, Figure 47, and Figure 49). In those maps, it looks like there are very few data points compared to the actual ~900 database records that exist. This is because many of those records were entered into the Inventory with just a broad, generic place name as their geographic footprint, so many records share the exact same data point.

Thus, a future opportunity for expansion of the GOMWIR Inventory, if resources were to become available, would be to include the addition of the geographic footprints for the studies of interest. Such inclusion would provide more specific metadata describing each study, which would improve searching capabilities and gap analyses, by identifying the specific regions of interest for each study.

2.2.4.3 Completeness of the GOMWIR Inventory—Another Opportunity for the Future

At the beginning of the process to develop and populate the GOMWIR Inventory, it was unclear just how many records with complete metadata could be collected in the time allotted especially given its nature as a collaborative, international effort. The nearly 900 bibliographic references that were compiled are a satisfactory start, but the current Inventory just scratches the surface of what is available. For example, beyond these ~900 references, simple bibliographic and publication lists were also collected and received, and they hold hundreds of additional resources that could potentially be added to the Inventory after gathering the supporting metadata. Similarly, Wes Tunnell's SGOM literature collection contains many hundreds more resources that simply could not be added to the Inventory give time constraints. And furthermore, several important, well-known compilation volumes focused on SGOM ecosystem science were only sparsely referenced in the Inventory or simply did not make it at all. A few select volumes in this category include Yáñez-Arancibia and Day (1988), Salazar-Vallejo and González (1993), Caso et al. (2005), Botello et al. (2011) and Sanchez et al. (2012). Just these five example compilations comprise almost 200 publications, which collectively include some 8,600 bibliographic references of potential interest.

Finally, we also note that a vast amount of ecosystem science also exists for the Cuban part of the Gulf, but the Inventory contains almost nothing from Cuba at this point. This unfortunate fact is simply due to the short amount of time that was available for this exploratory effort, and the general lack of good Internet connectivity available on the island. Fortunately, efforts are underway to change this. For example, HRI is currently involved with Cuban colleagues from the Center for Fisheries Research (CIP) in Havana to scan and OCR their historic collection of the *Revista Cubana de Investigaciones Pesqueras* (Cuban Journal of Fisheries Research). This effort involves about 60 physical volumes published between 1975-2006 (later versions are already online) that together comprise about 5,800 journal pages. This effort will unlock a huge amount of historic baseline data for the region, and is just one of many resources focused on Cuban coastal and marine ecosystem science that could be added to the GOMWIR Inventory as time and resources permit.

In sum, the ~900 bibliographic references compiled for the first version of the GOMWIR Inventory represent a satisfactory start given the short time frame and collaborative, international nature of the effort. But a vast amount of additional information is already known about, and could be added to the future revisions of the Inventory.

2.2.4.4 Other Observations and Comments

We have a couple of observations and comments about the GOMWIR Inventory process that, in general, may benefit future efforts.

First, volunteer data contributions were essentially neglible despite the large number of scientists and researchers that we reached with our workshop and inventory announcement messages. This is perhaps understandable given the many generic requests for information that scientists receive on a frequent basis. But even in the case of the follow-up to the map sticker exercise, when we wrote individual, personalized e-mails to participants who indicated via their stickers that they knew about studies or publications that were missing from the Inventory, not a single followup response was received. Thus, efforts to solicit data contributions from scientists were not very effective in this case.

Second, an improved process for contributing geographic footprints with data records is needed. Supposing that only textual contributions for footprints are possible, much stronger guidance about what constitutes a good option—and what will actually be accepted—is recommended. But ideally, a system with simple and easy-to-use geospatial tools (like Google Earth) might allow for direct collection of GIS polygons from volunteer data contributors.

2.2.4.5 Disposition of the GOMWIR Inventory

As discussed at the start of this paper, one of the main design goals during the planning and development stages of the GOMWIR Inventory was to match it with the structure of the database behind the BOEM ESID web application as closely as possible. The aspiration behind this effort was that the GOMWIR Inventory could one day, with sufficient polishing and effort, possibly serve as a pluggable data module for the ESID providing the missing coverage for the SGOM. Unfortunately, while the data collection process for the GOMWIR Inventory was underway, we learned that BOEM had made a strategic decision to stop pursuing the ESID due to increasing costs and IT requirements (personal communication, J. Blythe).

This unexpected change provides some flexibility regarding the final format and disposition of the GOMWIR Inventory. Given the time and effort that were expended to pull it together, and the very nature of the GOMWIR event to encourage trinational cooperation and coordination, the hope is that the Inventory can be made freely available to the greater GOM community. At the very minimum, it could be made available to end users as a simple, static, downloadable data table or GIS data file. But ideally it would be deployed as a web application with similar, but simplified, functionality as the former ESID, especially the easy-to-use mapping interface for visual/geospatial searching. An even better outcome is that the Inventory would not be a static database, but a living database that the community could continue to polish, revise, and extend as needed. However, our experience that voluntary data contributions are generally negligible suggests that our own efforts would probably be more key. Different potential solutions that would meet these goals are currently being investigated as are the necessary time, energy, and financial commitment that the various solutions would entail. Thus, for the moment, we will simply leave the GOMWIR Inventory as extensive, spreadsheet-based data tables until a decision is made.

2.2.5 Summary

The GOMWIR Inventory was developed as a foundational document to support the first Gulf of Mexico Workshop on International Research (GOMWIR). The Inventory was conceived as a comprehensive, cross-disciplinary inventory of LME research that would include:

1. An annotated bibliography of peer-reviewed literature, reports and other publications,
2. An annotated listing of Mexican research programs, and
3. An annotated listing of Mexican data sources.

Each of these annotated resources would address one or more of the three BOEM thematic areas of interest, i.e. Baseline Studies, Fates and Effects Studies, or Environmental Monitoring Studies (see 1.2.1 Introduction, in this volume) and be focused on the waters of the SGOM.

To produce the Inventory, a stand-alone web form system was developed using Google Apps Script to electronically collect data into a backend Google Sheets spreadsheet. A team of local and international partners was assembled to help gather information about resources of interest and enter it into the online web forms. Using this mechanism, 897 bibliographic references, 33 records about research programs, and five records about data sources were captured and entered into the database.

Several geographic trends, research themes, and temporal patterns were recognized in the collected data, and they were used to make loose suggestions about certain priorities for future investigation of marine ecosystem science in the SGOM.

2.2.6 Reference

Botello AV, Villanueva-Fragos S, Gutiérrez J, Rojas Galaviz JL (editors). 2011. Vulnerabilidad de las zonas costeras Mexicana ante el cambio climático, segunda edición. Campeche: Instituto EPOMEX, Universidad Autónoma de Campeche. 2 volumes.

Caso M, Pisanty I, Ezcurra E (editors). 2005. Diagnóstico ambiental del Golfo de México. Mexico, DF: Instituto Nacional de Ecología (INE-SEMARNAT). 2 volumes.

Madsen K, Zimmer B, Fritze B, Manzello L, Metzger K. 2014. EcoSpatial Information Database: US Atlantic Region. OCS Study BOEM. 2014-058. Available at: https://www.boem.gov/ESPIS/5/5379.pdf.

Salazar-Vallejo SI, González NE (editors). 1993. Biodiversidad marina y costera de México. Chetumal: Centro de Investigaciones de Quintana Roo.

Sánchez AJ, Chiappa-Carrara X, Brito Pérez R (editors). 2012. Recurso acuáticos costeros del sureste. Mérida, Yucatán: Red para el Conocimiento de los Recursos Costeros del Sureste. 2 volumes.

Yáñez-Arancibia A, Day JW Jr. (editors). 1988. Ecología de los ecosistemas costeros en el sur del Golfo de México: la region de la Laguna de Términos. Mexico, DF: Universidad Nacional Autónoma de México.