A METHOD FOR COMBINING INDICES OF ABUNDANCE ACROSS FLEETS THAT ALLOWS FOR PRECISION IN THE ASSIGNMENT OF ENVIRONMENTAL COVARIATES WHILE MAINTAINING CONFIDENTIALITY OF SPATIAL AND TEMPORAL INFORMATION PROVIDED BY CPCs

Matthew V. Lauretta¹, John F. Walter¹, Alex Hanke², Craig Brown¹, Irene Andrushchenko² and Ai Kimoto³

SUMMARY

Recent collaborative efforts in standardizing swordfish longline catch and effort information across U.S. and Canadian fleets allowed for hypothesis testing of stock distribution shift effects on fleet catchabilities. Scientists and managers agreed that a similar effort is worth pursuing for highly migratory stocks in which a changing oceanic environment may result in a shift in stock distribution and availability to fleets. The logistics of implementing standardized methods in a manner that does not violate data confidentiality concerns has not been outlined explicitly. We present methods for combining catch and effort information that allows scientists from individual CPCs to assign key environmental covariates to observations, and then assign coarser spatial (e.g. 5 by 5 degree cell, or larger areas) and temporal (e.g. month and season) categories. The resulting datasets preserve the links between testable factors and catch rates, and the observational level variability required for statistical hypothesis testing, while meeting confidentiality requirements. A common data structure allows the resulting datasets to be combined to develop a single standardized index that is more robust to changes in individual fleet catchability over time.

RÉSUMÉ

Les efforts de collaboration déployés récemment pour standardiser les informations sur la prise et l'effort ciblant l'espadon des flottilles palangrières des États-Unis et du Canada ont permis de tester les hypothèses des effets du changement de distribution des stocks sur les capturabilités de la flottille. Les scientifiques et les gestionnaires ont convenu qu'un effort similaire devrait être poursuivi pour les stocks de grands migrateurs pour lesquels un environnement océanique en évolution peut se traduire par un changement de la distribution des stocks et de la disponibilité pour les flottilles. La logistique de la mise en œuvre de méthodes standardisées d'une manière qui n'enfreint pas la confidentialité des données n'a pas été abordée explicitement. Le présent document décrit des méthodes servant à combiner les informations de prise et d'effort permettant aux scientifiques des CPC individuelles d'attribuer des covariables environnementales clés aux observations, et de leur allouer ensuite une résolution spatiale (p.ex. carrés de $5^{\circ}x5^{\circ}$ ou zones plus grandes) et temporelle (p.ex. par mois et saison) plus grossière. Les jeux de données obtenus préservent les liens entre les facteurs testables et les taux de capture et la variabilité du niveau d'observation requis pour les tests d'hypothèses statistiques, tout en respectant les exigences de confidentialité. Une structure de données commune permet de combiner les jeux de données obtenus pour développer un indice standardisé unique plus résistant aux changements de la capturabilité individuelle de la flottille au fil du temps.

RESUMEN

Recientes esfuerzos de colaboración en la estandarización de la información sobre la captura y el esfuerzo de la pesquería de palangre de pez espada en las flotas de Estados Unidos y Canadá ha permitido probar la hipótesis de los efectos de los cambios en la distribución del stock en las capturabilidades de la flota. Científicos y gestores acordaron que vale la pena seguir realizando un esfuerzo para stocks altamente migratorios en los que un entorno oceánico cambiante pueda generar un cambio en la distribución del stock y en su

¹ U.S. National Marine Fisheries Service, Southeast Fisheries Center, Sustainable Fisheries Division, 75 Virginia Beach Drive, Miami, FL, 33149-1099, USA. E-mail: matthew.lauretta@noaa.gov

disponibilidad para las flotas. La logística para la aplicación de métodos estandarizados de modo que no se infrinja el carácter confidencial no ha sido expuesta explícitamente. Se presentan métodos para combinar información de captura y esfuerzo que permiten a los científicos de las CPC individuales asignar covariables medioambientales clave a observaciones, y asignar categorías espaciales (por ejemplo cuadrículas de 5°x5°, o áreas más grandes) y temporales (por ejemplo, mes y temporada) con una escala más amplia. Los conjuntos de datos preservan los vínculos entre factores comprobables y tasas de captura, y la variabilidad del nivel de observación requerida para la comprobación de hipótesis estadísticas, cumpliendo al mismo tiempo los requisitos de confidencialidad. Una estructura de datos común permite que los conjuntos de datos resultantes puedan combinarse para elaborar un único índice estandarizado que es más robusto frente a los cambios en la capturabilidad de la flota individual en el tiempo.

KEYWORDS

Catch/Effort, Multivariate Analysis, Collaborative Research, Surface Temperature, Bottom Topography

1. Introduction

During the 2013 stock assessment for North Atlantic Swordfish, scientists from many CPCs collaborated to generate a combined index of relative abundance from commercial longline data aggregated at a trip level that covered almost the entire spatial extent of the fishing area (Ortiz et al. 2014). The resulting index was the primary index used in the surplus production modeling and also was useful in elucidating divergent patterns in individual fleet CPUEs that appeared to be related to changing oceanographic conditions (Schirripa 2014). The collaboration produced several beneficial results, including improved information on stock distribution patterns over time, formal hypothesis tests of the effect of stock redistribution on individual fleet catchability, incorporation of the Atlantic Multidecadal Oscillation as an environmental covariate that explained heterogeneity in catchability between different fleets in the stock assessment (Schirripa 2014), and potential estimation of the effect of gear restrictions on specific fleet catch rates. A similar attempt was made to combine Canadian and U. S. data at a finer set by set level of resolution (Walter et al. 2014); however, several hurdles were identified which included maintaining the confidentiality of data, and data sharing concerns specific to individual CPCs.

The sharing of the exact location and date information associated with an observation is often considered a potential violation of confidentiality, as it is argued that it might be possible to determine the identity of the individual vessel if the location and date of the fishing operation were known. Past practices for collaborative work conducted by the SCRS have included the aggregation of individual set observations into a dataset of catch and effort summarized by five by five degree squares and month (or quarter). Summaries which incorporate data from fewer than three separate vessels are then excluded, as vessels might still be identifiable in such cases through a process of elimination if the data from one vessel is known. Unfortunately, this approach may violate the assumptions of the standardization models intended to be applied to observational level data. The lack of fine-scale data and location information precludes the incorporation of data from auxiliary datasets (e.g., sea surface temperature values and gradients, bathymetry, depth of thermocline or minimum dissolved oxygen thresholds) which could provide information on factors which influence catch rates. Information on other potentially influential factors which may vary from set to set, such as bait and gear configuration, is also generally lost through this aggregated data approach.

In this paper, we present a methodology to handle set by set level CPUE information that assigns useful environmental (or other) covariates, removes information which may cause confidentiality concerns (including vessel identifier, exact fishing location, and date) and then assigns the observations to coarser spatial (e.g., five by five degree cell or larger areas) and temporal (e.g., month and season) categories. The resulting datasets preserve the links between testable factors and catch rates, as well as the observational level variability required for statistical hypothesis testing, while meeting confidentiality requirements. A common data structure would then allow the resulting datasets to be combined to develop a single standardized index that is more robust to changes in individual fleet catchability over time.

These methods are proposed to create a unified dataset spanning multiple fleets for CPUE standardization for Swordfish but could be applied to other species. Similarly, the methodology can be utilized to incorporate additional covariates beyond the examples described here.

2. Material and methods

The steps required for combining indices of abundance across fleets involve individual CPCs assigning environmental covariate data to their own data based on fine-scale set-by-set (or trip-by-trip) information of sample location and date, then assigning coarse-scale spatial (the examples described here include five by five degree geographic cell and ICCAT species area) and temporal (month, season, and year) categories to the data, and finally dropping the detailed information in favor of the coarse-scale fleet, spatial, temporal, catch, and effort data paired with the fine-scale environmental and gear covariates. The datasets are then merged into a single dataset and can be used in standard CPUE modeling. Below we outline each step in detail and provide R codes to accomplish them.

2.1. Assignment of fine-scale environmental covariates, accomplished by individual CPCs.

2.1.1. Sea Surface Temperatures.

Sea surface temperature (SST) data are collected by U.S. National Oceanic and Atmospheric Administration (NOAA) meteorological satellites are available online and in situ data and at http://www.esrl.noaa.gov/psd/data/gridded/data.noaa.oisst.v2.html. Weekly mean sea surface temperature estimates are available for the period November 1981 to present and assignment requires data fields on sample latitude (one degree resolution minimum), longitude (one degree resolution minimum), and date (Reynolds et al. 2002). Figure 1 provides an example of the ST data summarized for one date, January 1, 2000. The R code to assign these data follows:

```
#requires R package "ncdf"
setwd('R:/Sea_Surface_Temperatures/')
library(ncdf)
sst1=open.ncdf('sst.wkmean.1981-1989.nc') #Data accessed here:
http://www.esrl.noaa.gov/psd/data/gridded/data.noaa.oisst.v2.html
sst2=open.ncdf('sst.wkmean.1990-present.nc')
                                                      #ss1 downloaded on November 23, 2014 and ss2 downloaded
on November 24, 2014.
x1=get.var.ncdf(sst1,"lon")
y1=get.var.ncdf(sst1,"lat")
z1=get.var.ncdf(sst1,"time")
temp1=get.var.ncdf(sst1,"sst")
x2=get.var.ncdf(sst2,"lon")
y2=get.var.ncdf(sst2,"lat")
z2=get.var.ncdf(sst2,"time")
temp2=get.var.ncdf(sst2,"sst")
assign_sst=function(lon,lat,date)
           if(date<'1981-10-29')
                      'Data Not Available'
           }
           else
                      if(date<'1990-01-01')
                                 week=which(z1>julian(as.Date(date),origin=-62091)[1])[1]-1
                                 temp1[which(trunc(x1,0)==trunc(lon,0)),which(trunc(y1,0)==trunc(lat,0)),week]
                                 }
                      else
                                 week=which(z2>julian(as.Date(date),origin=-62091)[1])[1]-1
                                 temp2[which(trunc(x2,0)==trunc(lon,0)),which(trunc(y2,0)==trunc(lat,0)),week]
                                 }
                      }
           }
```

DATA FORMAT NOTES: Longitude (lon) is 360 degree grid from the Prime Meridian east, Latitude (lat) is 180 degree grid from -90 to +90 centered at the equator,

Date format is "YYYY-MM-DD". Excel can assign dates to this format by selecting the English(U.K.) locale under the Format Cells-Date menu

Example for FL Gulf coast (-84 lon, 28 lat, on November 1, 2013):

assign_sst(lon=360-84,lat=28,date='2013-11-01')

Example appending weekly mean SST to a dataset, titled "mydata" in the following example, that contains the fields "lon", "lat", and "date"

mydata=read.csv("example.csv",header=TRUE) mydata\$date2=as.Date(mydata\$date) mydata\$sst=sapply(1:length(mydata[,1]),function(i)assign_sst(mydata\$lon[i],mydata\$lat[i],mydata\$date2[i

]))

write.csv(mydata,"example_with_SST.csv")

CREDITS: NOAA_OI_SST_V2 data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.esrl.noaa.gov/psd/

NOAA/OAR/ESRL PSD, Boulder, Colorado, would appreciate receiving a copy of any relevant publications and reports, which will assist in keeping the NOAA_OI_SST_V2 data set freely available online in the future. Thank you!

2.1.2. Oceanic Depth.

Ocean bathymetry is available from NOAA using the ETOPO1 global relief model (Amante and Eakins 2009), which has land elevation and ocean bathymetry on a resolution of one arc-minute grid. The data can be downloaded at http://www.ngdc.noaa.gov/mgg/global/relief/ETOPO1/data/bedrock/grid_registered/xyz/. The R package 'marmap' (Pante and Simon-Bouhet 2013) has several built in tools that are very useful for obtaining, plotting and assigning depth data to specific latitude and longitude locations. In particular, the package has functions for retrieving the data at a user-specified resolution and for assigning the ocean depth to a set of latitude and longitude coordinates. We explored whether to use five minute (~five nautical mile resolution) or a courser 10 minute grid and found very little difference when compared with U.S. Pelagic longline data where the depth is recorded at sea. As longlines are spread over several to 10s of nautical miles and cover a fairly wide area during their drift, the five minute resolution is likely sufficient. Depth data in meters were downloaded at five minute resolution for an area from the equator to 50°N latitude and 0 to 98°W longitude (Figure 2). Code for downloading and assigning depths to observations is provided here:

```
#requires R package "marmap"
library(marmap)
NATL <- getNOAA.bathy(lon1 = -98, lon2 = -0, lat1 = 0, lat2 = 60, resolution = 5)
dim(NATL)
XY= data.frame(mydata$lon, mydata$lat)
names(XY)=c("x","y")
system.time( {XYZ=get.depth(NATL, x=XY[,1], y=XY[,2], locator=FALSE ) })
mydata2= cbind( mydata2,XYZ)</pre>
```

2.1.3. Depth Gradient.

Fish often aggregate to underwater structure. Depth gradient (slope of the seafloor) may be an indicator of the vertical structure and a potential useful variable in CPUE standardization. Depth gradient was calculated from the above depth data and obtained by finding the first difference in a depth value in either the x (longitude) or the y (latitude) coordinates. Some of the basic ideas were borrowed from the following online resource, http://stackoverflow.com/questions/25704711/how-do-i-calculate-the-gradient-of-a-matrix-to-draw-a-vector-field-in-r, based on an R discussion forum and the solution provided by Ben Bolker. The estimates of depth gradient of the Atlantic Ocean are shown in **Figure 3**. The R code for assigning depth gradient to a given latitude-longitude is:

bdiff <- function(x) c(NA,diff(x))

dx <- t(apply(NATL,1,bdiff)) #calculate differences in at 5 min (grid cell resolution) in x direction dy <- apply(NATL,2,bdiff) #calculate differences in at 5 min (grid cell resolution) in y direction $grad= sqrt(dx^{2} + dy^{2})$ #gradient is slope but this a rudimentary function, in reality distances in the x direction are shorter further north. The XY data needs to be converted to an equidistant set of units so this is an approximation

dimnames(grad)=dimnames(NATL) #replace the "NA" in dimnames XYZgrad=getGradient(grad, x=XY[,1], y=XY[,2], locator=FALSE)

#COMMENT: getGradient function is a slightly altered version of get.depth() with the requirement that the first argument be an object of class(bathy) commented out so as to not have to convert grad to a 'bathy' class

2.2. Assignment of coarse-scale spatial and temporal categories, accomplished by individual CPCs.

The assignment of coarse-scale spatial and temporal categories is essential to maintain fisher confidentiality and meet the regulatory requirements and concerns for data sharing amongst CPCs. The following R code assigns two coarse-scale spatial cell categories to the fine-scale geographic coordinates (decimal degrees), namely a) five by five geographic cell and b) ICCAT species area. It also assigns coarse-scale temporal categories of month, season, and year using the date of the observation.

```
mydata2=read.csv("US_Observer_Data_SWO_withSST.csv",header=TRUE)
mydata2$5x5_lat=trunc(mydata2$lat/5)*5+2.5
mydata2$5x5_lon=trunc(mydata2$lon/5)*5+2.5
#requires R package "sp"
library(sp)
SWO1=list(x=c(-30,20,20,-30), y=c(30,30,5,5))
SWO2=list(x=c(-50,-30,-30,-50), y=c(30,30,5,5))
SWO3=list(x=c(-100,-50,-50,-100), y=c(20,20,5,5))
SWO4=list(x=c(-100,-80,-80,-100), y=c(30,30,20,20))
SWO5=list(x=c(-80,-70,-70,-80), y=c(35,35,20,20))
SWO6=list(x=c(-70,-50,-50,-70), y=c(35,35,20,20))
SWO7=list(x=c(-85,-65,-65,-85), y=c(50,50,35,35))
SWO8=list(x=c(-65,-55,-55,-65), y=c(50,50,35,35))
SWO9=list(x=c(-55,-35,-35,-55), y=c(50,50,40,40))
SWO10=list(x=c(-35,-20,-20,-35), y=c(50,50,40,40))
SWO11=list(x=c(-20,0,0,-20), y=c(50,50,40,40))
SWO12=list(x=c(-20,0,0,-20), y=c(40,40,30,30))
SWO13=list(x=c(-40,-20,-20,-40), y=c(40,40,30,30))
SWO14=list(x=c(-55,-40,-40,-50,-50,-55), y=c(40,40,30,30,35,35))
SWO_area=c("SWO1", "SWO2", "SWO3", "SWO4", "SWO5", "SWO6", "SWO7", "SWO8", "SWO9", "SWO10",
"SWO11", "SWO12", "SWO13", "SWO14")
lat=mydata2$lat
lon=mydata2$lon
mydata 2\$ area = sapply (1: length(mydata 2[, 1]), function(i) SWO\_area [which(c(i) SWO\_are
                     point.in.polygon(lon[i],lat[i],SWO1$x,SWO1$y),
                     point.in.polygon(lon[i],lat[i],SWO2$x,SWO2$y),
                     point.in.polygon(lon[i],lat[i],SWO3$x,SWO3$y),
                     point.in.polygon(lon[i],lat[i],SWO4$x,SWO4$y),
                     point.in.polygon(lon[i],lat[i],SWO5$x,SWO5$y),
                     point.in.polygon(lon[i],lat[i],SWO6$x,SWO6$y),
                     point.in.polygon(lon[i],lat[i],SWO7$x,SWO7$y),
                     point.in.polygon(lon[i],lat[i],SWO8$x,SWO8$y),
                     point.in.polygon(lon[i],lat[i],SWO9$x,SWO9$y),
                     point.in.polygon(lon[i],lat[i],SWO10$x,SWO10$y),
                     point.in.polygon(lon[i],lat[i],SWO11$x,SWO11$y),
                     point.in.polygon(lon[i],lat[i],SWO12$x,SWO12$y),
                    point.in.polygon(lon[i],lat[i],SWO13$x,SWO13$y),
                     point.in.polygon(lon[i],lat[i],SWO14$x,SWO14$y))==1)])
mydata2$year= format(as.Date(mydata$date,format="%m/%d/%y"),"%Y")
mydata2$month=format(as.Date(mydata$date,format="%m/%d/%y"),"%m")
mydata2$season= ifelse(month%in%c(3:5),'spring', ifelse(month%in%c(6:8),'summer',
                     ifelse(month%in%c(9:11),'fall', ifelse(month%in%c(1,2,12),'winter','NA'))))
```

write.csv(mydata2,"US_Observer_Data_SWO_coarse_Spatial.csv")

The coarse-scale spatial categories assigned here (5 by 5 degree squares and ICCAT species-specific area codes) are intended as examples. In fact, it is ultimately preferable that the area definitions be determined based on the data. In the CPUE standardization procedure, the effect of each area on the catch rates is estimated. The standardized CPUE index can therefore be improved if the area delineations correspond to actual spatial differences in catch rates. Ultimately, it may be preferable for CPC scientists to define areas such that catch rates are more similar within each area than between areas, perhaps using cluster analysis. In cases where fisheries overlap, scientists from each CPC can compare separately developed area delineations and reach agreement on final area definitions. Alternatively, it may be possible for scientists from all involved CPCs to work with combined data assigned to 5 by 5 degree resolution squares, and assign data to larger areas as warranted by differences in catch rates.

2.3. Data filter to remove detailed confidential fishery information, accomplished by individual CPCs.

The removal or relabeling of detailed fishery information, including vessel identification, precise set location, and sample date is the next step taken to maintain the confidentiality of the data. The following code generates an example of a final dataset that only includes catch, fleet identifiers, and the coarse-scale spatial and temporal information paired with the fine-scale environmental covariate data required for hypothesis testing.

This basic example can be modified to maintain in the analysis dataset additional factors which may potentially influence catch rates, such as set by set information on bait, gear configuration, time and duration of set. These factors can be maintained at the most detailed level or assigned to larger categories common across fleets.

Although not currently included in the code shown above, at this stage additional filters can be imposed, such as eliminating observations from area-month strata in which there are fewer than three separate vessels. The analysis data set does not require inclusion of all available data, so long as the exclusion of the eliminated data does not introduce a bias.

2.4. Combining filtered datasets to generate a single dataset for index standardization, accomplished by the collaborative CPC analytical team.

The final step is to combine the filtered datasets produced in Step 3 and provided by the individual CPCs into a single dataset that contains only the necessary information required for index standardization. Each dataset in this example includes the flag, gear, coarse spatial and temporal information, environmental covariates, catch, and effort information. The following code combines the individual datasets:

USA=read.csv("USA_LL_SWO_Filtered.csv", header=T) CAN=read.csv("CAN_LL_SWO_Filtered.csv", header=T) NAM_LL=rbind(USA, CAN) write.csv(NAM_LL, "N_America_SWO_LL.csv")

3. Results and discussion

The key contribution of this paper is a methodology that provides the basis for creating a single index of abundance from many fleets rather than relying upon multiple, potentially divergent, CPUE indices. Scientists from individual CPCs process their own data, assign relevant variables necessary for CPUE standardization (e.g., SST, water depth, gear configuration, etc.), and then assign the data to spatial and temporal scales where the data is no longer considered confidential while removing vessel name and other identifying information.

It is noteworthy that almost all of the CPUE model factors (catch rate, year, area, season, SST, water depth, bait, hooks set, hook type, time of day of set, etc.) commonly used in statistical modeling are not, by themselves, confidential. For instance, while the fine-scale latitude, longitude and date are needed to assign a SST, as outlined above, the precise spatial location (along with identifying vessel characteristics) are not needed to create an index. This key finding allows for the development of multi-CPC datasets that meet data confidentiality requirements while supporting the SCRS need for improved CPUE indices.

Above, we outlined a method that matches catch data with high resolution environmental covariate information while preserving the confidentiality of fishers. Confidentiality was the primary concern raised by parties during discussions of future applications, and domestic regulations of individual CPCs often require that the data confidentiality be preserved. This paper represents only an outline of methods and statistical tools and is not expected to be a comprehensive solution to combining all available catch rate information. No doubt, additional obstacles and possible code bugs will be encountered when attempting to apply these methods to other species and incorporating other fleets. However, this paper does demonstrate that broad spatial scale, multi CPC analyses of catch data are technically feasible and it is our hope that this initiative creates a forum for further collaboration, discussion of potential hurdles, and innovative solutions. These and future efforts will allow for the incorporation of high resolution environmental data required for future stock assessments, while preserving the proprietary and confidential information of individual fleets and fishers.

References

- Amante C, Eakins BW .2009. ETOPO1 1 arc-minute global relief model: Procedures, data sources and analysis. NOAA Technical Memorandum NESDIS NGDC-24: 1-19.
- Ortiz M., J. Mejuto, I. Andrushchenko, K. Yokawa, J. Walter, M. Santos, & N. Abid.2014. An updated combined biomass index of abundance for the North Atlantic Swordfish Stock 1963-2012. Collect. Vol. Sci. Pap. ICCAT, 70(4): 1912-1926.
- Pante E, Simon-Bouhet B (2013) marmap: A Package for Importing, Plotting and Analyzing Bathymetric and Topographic Data in R. PLoS ONE 8(9): e73051.
- Reynolds, R.W., N.A. Rayner, T.M. Smith, D.C. Stokes, and W. Wang. 2002. An improved in situ and satellite SST analysis for climate. J. Climate, 15, 1609-1625.

Schirripa, M. 2014. A hypothesis of a recent poleward shift in the distribution of North Atlantic Swordfish. Collect. Vol. Sci. Pap. ICCAT, 70(4): 1998-2008 (2014)

Walter, J., M. Lauretta, A. Hanke, I. Andrushenko & C. Brown. 2014. Standardized catch rate of Swordfish from the U.S. dealer landing system with a preliminary consideration of a combined U.S.-Canada pelagic longline fleet dataset. Collect. Vol. Sci. Pap. ICCAT, 70(4): 1885-1911.

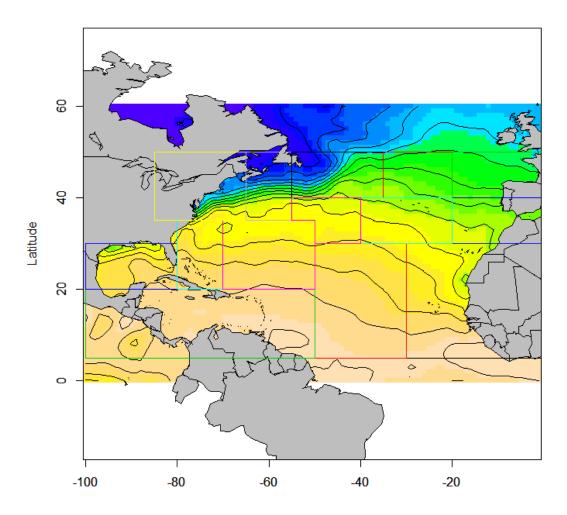


Figure 1. Example of sea surface temperature data obtained from NOAA in situ and satellite observations, data shown for January 1, 2000.

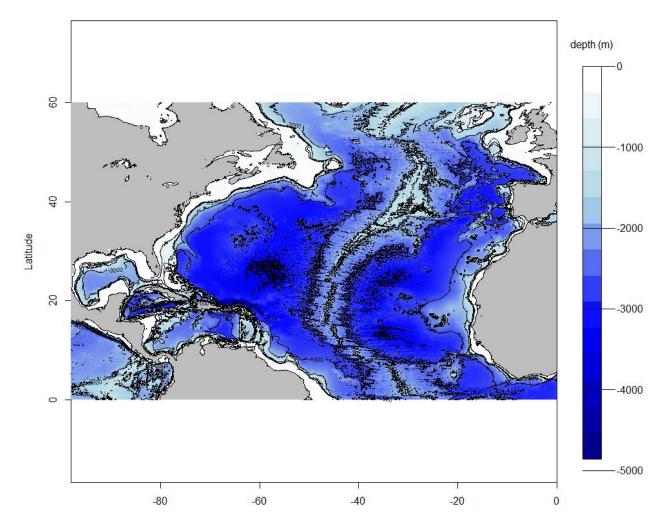


Figure 2. Distribution of seafloor depth from http://www.ngdc.noaa.gov/mgg/global/relief/ETOPO1/.

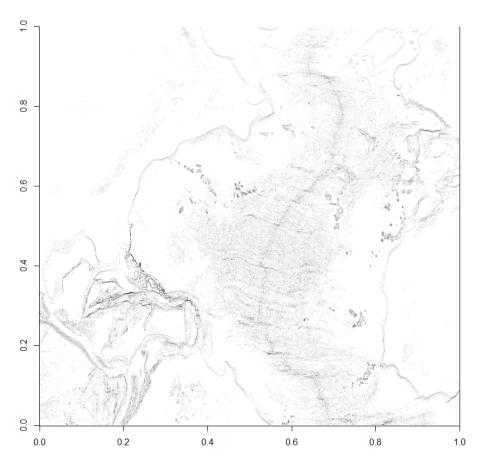


Figure 3. Distribution of seafloor depth gradient. Darker areas indicate steep slopes.