

# A Principal Component Analysis of Vertical Temperature Profiles for Tracking Movements of Large Pelagic Fishes

Kathryn G Carmody

Halmos College of Natural Sciences and Oceanography, Nova Southeastern University,  
United States

E-mail: kc1296@nova.edu

Arthur J Mariano

University of Miami, United States

E-mail: professorarthurmariano@gmail.com

David William Kerstetter

Halmos College of Natural Sciences and Oceanography, Nova Southeastern University,  
United States

E-mail: kerstett@nova.edu

Received: November 30, 2016 Accepted: January 18, 2017 Published: January 24, 2017

doi:10.5296/ast.v5i1.10649 URL: <http://dx.doi.org/10.5296/ast.v5i1.10649>

## Abstract

Pop-up satellite archival tag (PSAT) technology that records depth, temperature, and light-level data has expanded the understanding of free-swimming behavior for numerous pelagic animals. Astronomical algorithms using these light-level data have allowed geolocation estimates of daily longitude and latitude. However, many pelagic animals have a crepuscular behavior pattern in which individuals are at depths below the photic layer during the day, thus precluding the use of traditional light-based movement algorithms for geolocation in such species as swordfish. A principal component analysis (PCA) of temperature profiles is described herein that utilizes depth and temperature data rather than light to estimate the horizontal movement between the initial location of tag release and transmission. PSAT data from swordfish (n=4), blue marlin (n=14), white marlin (n=2), and

black marlin (n=1) were used to generate daily coordinate estimates. The marlin data provided sufficient light information to derive geolocation estimates using two light-based state space models, while the hydrographic PCA model was used to derive comparison estimates. Comparisons of the two model types show an average root mean square difference of 175.4 km demonstrating that the PCA model can be used to extract the movement of tagged swordfish and other pelagic species demonstrating crepuscular behavior. Integration of this PCA-based geolocation methods with both the best available estimates of the ocean temperature at the time of tag deployment and the existing light-based geolocation models would provide additional information on fine-scale movement of tagged fish.

**Keywords:** Geolocation, Model, Satellite tag, Swordfish, Tracking

## 1. Introduction

Swordfish *Xiphias gladius* Linnaeus, 1758 is a monogeneric species found circumglobally in tropical, subtropical, and temperate pelagic waters, where it supports numerous commercial fisheries (Sepulveda et al., 2010). The movement of swordfish is one of the most complex among pelagic fishes (Dewar et al., 2011). Previous swordfish tagging studies in the Atlantic, Pacific, and Mediterranean Sea have revealed large horizontal and vertical distribution patterns that vary considerably with season, gender, and ocean basin (Sedberry and Loefer, 2001; Neilson and Smith, 2010; Sepulveda et al., 2010; Dewar et al., 2011). Pop-up satellite archival tags (PSAT) are integrated with the combined technology of satellite telemetry and archival data storage tags. PSATs are equipped with three sensors that record ambient variables including temperature, pressure (converted to depth), and light intensity at specified sampling frequencies. Once the tag reaches the sea surface following release from the animal, the stored data is automatically transmitted to an Argos satellite. The data retrieved has been critical to understanding the physical and physiological variables that relate fish behavior to its surrounding environment (Arnold and Dewar, 2001). Using electronic pop-up satellite tag (PSAT) technology, prior studies have shown swordfish moving against light so that during daylight hours they are found predominantly below the thermocline (300-1000 m), while night hours are spent above the thermocline in near-surface waters (Sepulveda et al., 2010; Lerner et al., 2013). These tagging studies also show that swordfish usually descend to depth one hour before sunrise and ascend after sunset. This is referred to as crepuscular diving behavior, and is a typical foraging strategy strongly associated with movement of the deep sound scattering layer (Sepulveda et al., 2010; Dewar et al., 2011).

PSATs archive ambient light-level data, which are often used to estimate geolocation. However, raw geolocations derived from light-based algorithms are noisy with geolocation errors as large as hundreds of kilometers (see review in Nielsen et al., 2006). To reduce the geolocation estimation error, various mathematical methods have been coupled with the recorded light data to produce a relatively more accurate track. In particular, Hill and Braun (2001) described astronomical equations to minimize the variability of latitude estimates from light level-based geolocation methods. Under optimal conditions, light-based estimation error in latitude and longitude can be reduced to  $0.7^\circ$  (ca. 78 km) and  $0.3^\circ$  (ca. 33 km) respectively (Hill and Braun, 2001; Sibert et al., 2003; Nielsen and Sibert, 2007). However, due to the systematic and random errors associated with practical sampling of the ambient light signal, these light-based methods produce longitude and latitude estimates with errors that can still be hundreds of km (Nielsen and Sibert, 2007; Musyl et al., 2003).

Light-based methods have also attempted to minimize error through additional ambient environmental data from PSATs, including sea surface temperature (SST) and bathymetry. State-space models have expanded to include supplementary data, such as coastline, bathymetry, and sea surface temperatures to constrain the model (Musyl et al., 2003; Teo et al., 2004; Nielsen and Sibert, 2007; Lam et al., 2010). In particular, the incorporation of satellite-derived sea surface temperature (SST) data has further refined light-based geolocation estimations. However, geolocation estimates can still have large errors (Hill and Braun, 2001; Lisovski, 2012). For example, Lam et al. (2011) incorporated sea-surface temperature (SST) into the state-space model *TrackIt*, which also uses light information and a Kalman filter. Similar to all light-based geolocation methods, *TrackIt* produces coordinate estimates usually with errors of  $0.5$  to  $1^\circ$  (ca. 60 km to 110 km) for longitude and  $1^\circ$  to  $2^\circ$  (ca. 110 km to 220 km) for latitude (Nielsen and Sibert, 2007). Although double-tagging comparisons by Lam et al. (2011) showed an increase in overall accuracy, there were larger

geolocation errors when the SST did not exhibit a strong thermal gradient. The most sophisticated models rely on distinct spatial variability of these environmental parameters to increase the accuracy of estimated positions.

SST-enhanced methods do not incorporate all of the hydrographic information in the data transmitted from satellite tags. Data such as temperatures recorded at depths below the surface help describe the water column where distinct thermoclines and thermal patterns are revealed. For example, the thermal variability in SST can be as low as 1-2°C across the Florida Straits, while there is an 8-10°C change in temperatures at depth below the main thermocline (Leaman et al., 1989). The use of light geolocation-based models assimilated with such hydrographic data has been demonstrated in previous archival tagging studies. A study by Skomal et al. (2004) using PSAT data from basking sharks deployed in the North Atlantic described how the combined use of geolocation estimates and a vertical temperature profile analysis revealed a more extensive track than what the pop-up locations implied.

Temperature versus depth profiles vary over geographical area and time due to ocean dynamics and variations in ocean forcing. As the tagged fish make their crepuscular vertical migrations, the tags sample the vertical distribution of temperature over the water column. In addition, the temperature profiles will significantly change by the fish going across a horizontal oceanic front. Thus, variations in daily temperature profiles can be used to track the movement of a fish if some information about the distribution of temperature profiles with latitude and longitude is known. A *principal component analysis* (PCA) of such thermal profiles will take advantage of these larger subsurface changes in temperature. By means of this analysis, rather than rely on the limited light records and just SST, a new model is proposed herein that uses temperature and depth data alone to estimate geolocation. Our proposed PCA-based geolocation method is sufficient as a statistical estimator since it uses all of the temperature versus depth data. Our method is also objective compared to the subjective approach of Skomal et al. (2004).

As a statistical tool, PCA (and its various extensions) is particularly useful for analyses of large data sets, such as archival tag records, as it greatly reduces data volume and noise in the process. PCA calculates the eigenvectors of the covariance matrix of the temperature versus depth profiles. These eigenvectors, known as empirical orthogonal functions (EOFs), are the most efficient basis for representing the temperature profile data (Mariano et al., 1996). The daily temperature profiles are regressed onto the set of EOFs and the resulting coefficients are known as principal components (PCs). A simple bilinear model for the spatial distribution of PCs is assumed and is used to calculate daily estimates of longitude and latitude.

## **2. Materials and Methods**

### *2.1 Satellite Tagging*

Four Microwave Telemetry high rate (HR) PTT-100 tags were deployed on swordfish caught and released at night during the annual Cayman Swordfish Challenge fishing tournament held during the spring of 2012, 2013, and 2014. Tagging was performed using methods described in Fenton (2012). Data from PSATs deployed on marlin in 2008, 2009, 2011, 2012 and 2013 were reanalyzed in the comparison portion of this study. Microwave Telemetry (MWT) PTT-100 HR tags were rigged and deployed on blue marlin in the Caribbean as described in Graves and Horodysky (2005).

Platform Transmitter Terminals (PTTs) send signals to satellites once the antennae breaks the surface at pop off. The data is relayed to ground stations where it is electronically transferred

to processing centers to be made available to researchers. These tags were programmed to record measurements every 90-120 seconds. The temperature sensor ranges from -4 to 40 °C with a resolution of 0.16 to 0.23 °C. Rated to 3000 psi, depth is recorded 0 to 1296 m with a resolution of 5.4 m. Unfortunately, the high sampling rate programming, including the increased depth resolution, results in decreased resolution in light level data due to on-board data storage limitations. The light sensitivity of these tags is therefore less than  $4 \times 10^{-5}$  lux (1 lumen) at 555 nm and does not allow for raw geolocation estimates to be generated by the tag itself. To generate estimates, the light data received was run through a light-based model. These tags can be programmed to release for up to thirty days following activation. PTT-100 HR tag deployments for this study ranged from 9 to 29 days.

Mk10 and Mini-PAT popup archival tags (PATs) manufactured by Wildlife Computers (Redmond, Washington, USA) were deployed on blue marlin *Makaira nigricans*, black marlin *Istiompax indica*, and white marlin *Kajikia albida* (formerly *Tetrapturus albidus*) as part of the International Game and Fishing Association (IGFA) International Great Marlin Race (IGMR). Both tags models are pressure rated to 2000 m, and are equipped with sensors to record and store temperature, depth, and light data. The temperature sensors have a resolution of 0.05°C with a range of -40° to 60°C and -5° to 45°C for the Mk10 and MiniPAT, respectively. The depth sensor of both tag models is rated to 1700 m with 0.5 m resolution. Light sensitivity ranges from  $5 \times 10^{-12}$  W cm<sup>-2</sup> to  $5 \times 10^{-2}$  W cm<sup>-2</sup>. The MiniPAT also includes two light sensors to reduce noise, each measuring at an optimum 440 nm wavelength. Both PAT models are optimized for post-deployment geolocation estimations. Wildlife Computers offers software (Wildlife Computers Global Position Estimator Version 2) that processes the high-resolution light records to produce global positioning estimates (GPEs). A light-based algorithm uses the light data collected by the tag with dawn and dusk light curves to generate the GPEs (Wildlife Computers 2014). The maximum length of deployment for both Wildlife Computers tag models is two years. The Mk10 and MiniPAT data used for this study came from deployments ranging from 28 to 180 days.

## 2.2 Preparation of Tag Data

Archived data are relayed in hexadecimal code by the floating PSAT through the ARGOS satellite system. Once the PTT-100 HR completed its transmission, data reports containing time-series depth, and time-series temperature are decoded, and compiled into Excel files (Microwave Telemetry 2013). For our analyses, the archived depth and temperature data are imported from .csv files into R, where the daily temperature profiles are extracted. Temperature and depth data from WC PAT tags are presented in histograms. Unlike the Microwave Telemetry PTTs, raw temperature and depth records are not given in the processed files. Rather, the archived temperature and depth data are compressed into data bins set at pre-determined sampling intervals. PAT-style depth-temperature profiles (PDTs) are given that consists of all temperature and depth data records for that day divided into eight bins of equal size. For each bin depth, a minimum, maximum, and mean temperature is given (Wildlife Computers 2014). In order to create a more specific temperature profile attributed to each day, the eight mean temperature values were interpolated in our analyses to generate measurements, on average, every 5 m depth.

## 2.3 Geolocation Modeling

The temperature profiles (i.e., temperature at depth values) are the input data for the PCA-based geolocation methods. The typical behavior of swordfish is primarily diurnal as they make one large vertical dive each day. Therefore, it takes one full day to get enough data

to estimate a vertical temperature profile. Although marlin species exhibit similar diel behavior, tag studies reveal much more time is spent the photic zone (0-80m) allowing sufficient light measurements to be recorded as well (Horodysky et al., 2007).

A cubic spline was fit to all of the daily temperature data and for each PSAT. The spline fits for each day were then averaged in 5 m (marlin) or 20 m depth (swordfish data) intervals for each day to produce a mean  $T(z)$  profile (Figure 1). Residual temperature profiles were calculated by removing this mean profile from each daily profile. The covariance function of the temperature profiles is then calculated, with the eigenvectors of this covariance function known as the *empirical orthogonal functions* (EOFs). The corresponding eigenvalue is the variance explained by that mode. EOFs are the most efficient set of basis functions for representing the variability in the tag's temperature profiles. The original data are regressed onto the EOFs and the resulting amplitudes are the principal components; a simple model is assumed for how the principal components vary with longitude and latitude.

The spatial distribution of the PCs is used for horizontal track estimation. The first EOF (and its PC) represents the most dominant pattern of variability among the vectors described by the vertical temperature profiles. The temporal distribution of the first PC contains most of the daily changes in the profiles that result from a change in location. Thus, this PCA method should provide an accurate description of daily movement via the changes in the first PC. The spatial coordinates of the PCs are known at the first and last days of tagging and the PCA-based geolocation method assumes a relationship between the PCs and both longitude and latitude. The method is detailed in next section.

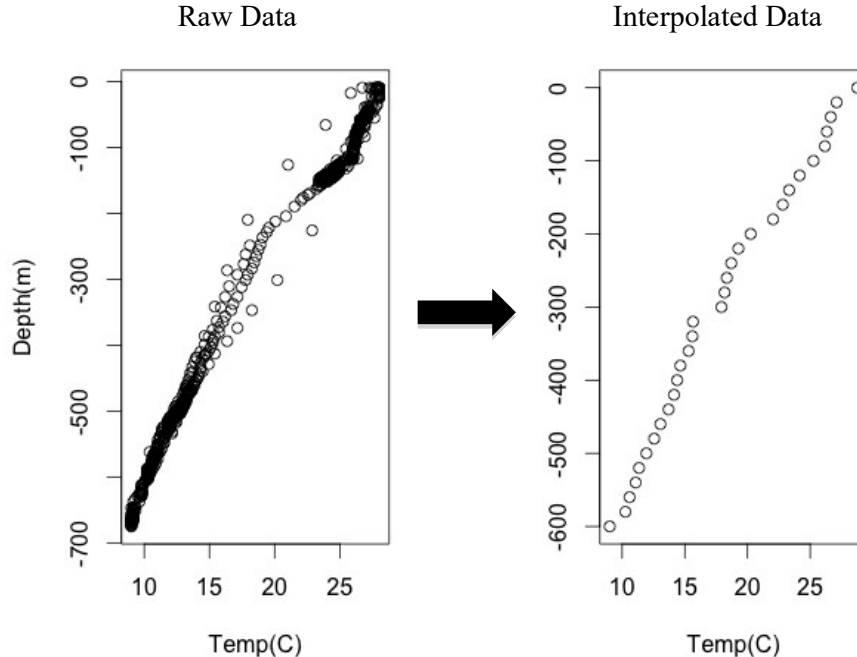


Figure 1. Temperature and depth data recorded by a PTT-100 HR pop-up satellite archival tag (Microwave Telemetry, Columbia, MD) attached to a swordfish (PTT 61665). The raw data were interpolated and averaged at every 20 m depth to a maximum depth of 660 m.

### 2.4 Principal Component Analysis

A matrix  $X_{ij}$  is created from residual temperature at the  $j^{\text{th}}$  depth from the  $i^{\text{th}}$  day (profile) specified at  $n$  depths. The vertical covariance function is calculated by multiplying the  $X_{ij}$  by its transpose  $X_{ij}^t$ . EOFs are the normalized eigenvectors of the vertical covariance function. The EOFs associated with the largest eigenvalues are the most efficient basis for representing the data (see Mariano et al., 1996). A *principal component* (PC) modulates the EOFs for each data realization; these PCs are calculated by regressing the residual temperature profiles onto the EOFs. The first EOF and its PCs explain more of the data variance than any other mode, while the second EOF and its PCs explain more of the remaining data variance than any other mode. After calculating the EOFs and PCs, the PCs corresponding to the first two eigenvectors  $PC_{i,1}$  and  $PC_{i,2}$  for each daily analysis  $i$  are examined (Fig. 2). It is shown below that the second EOF and its PCs explained an insignificant amount of variability, such that their inclusion in the analysis generally did not aid in geolocating the fish. However, there were a few cases where using the second PC did improve the geolocation.

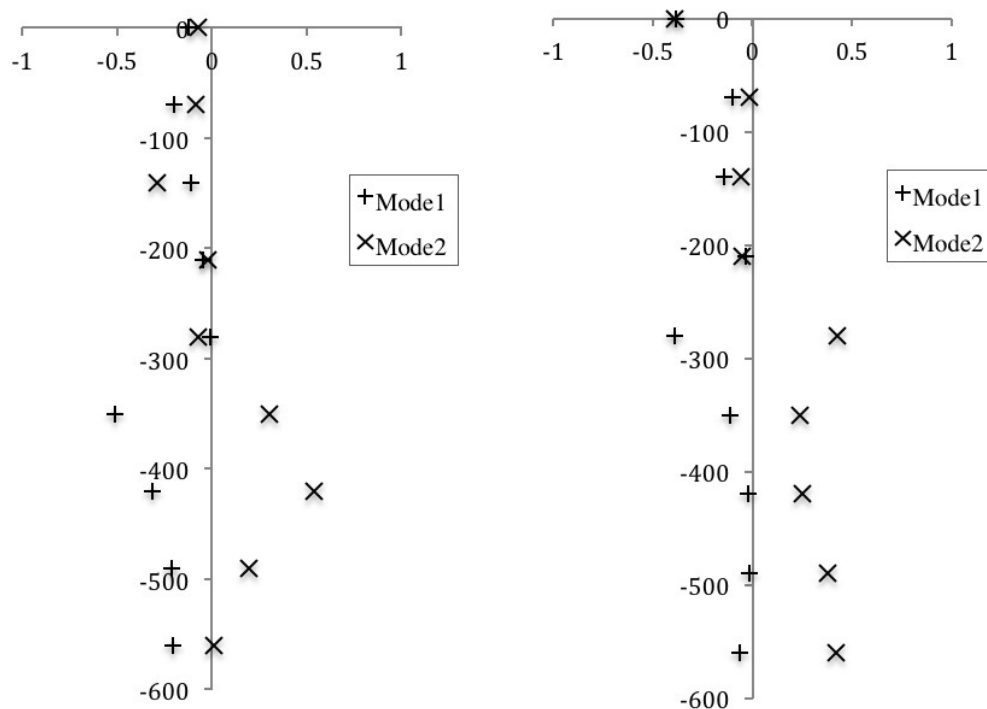


Figure 2. The first two vertical *empirical orthogonal functions* (EOFs) calculated from swordfish (PTT 61665 left and PTT 86995 right) archival data plotted as a function of depth (in meters). The first EOFs have low amplitude in the mixed layer and largest amplitude in the thermocline. This result is expected since the vertical variability in temperature at depth is a strong function of the vertical derivative of temperature, which is a minimum in the maximum layer and is a maximum in the thermocline.

### 2.5 Bilinear Algorithm

An algorithm incorporating the PCs and the initial and final positions from the coordinates of tag deployment and first transmission was written to calculate the spatial distribution. This equation calculates the estimated coordinates for each respective day during the tag deployment. The vectors of known longitude and latitude are given by:

$$\begin{bmatrix} X_1 \\ X_n \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_n \end{bmatrix}$$

where  $X$  is longitude and  $Y$  is latitude and  $i = \text{day } 1, 2, \dots, n$ . Linearity is assumed in PCA due to our reliance on the original data to interpolate between individual data points. A bilinear system is applied to calculate the resulting coordinates as simply as possible. Both longitude and latitude were assumed to change linearly with PC value, so that:

$$\begin{aligned} X_i &= aPC_{i,1} + b \\ Y_i &= cPC_{i,1} + d \end{aligned}$$

Each set of coefficients ( $a, b$ ) and ( $c, d$ ) were found by using the initial and final location. These coefficient sets represent the change in longitude ( $a, b$ ) and latitude ( $c, d$ ) over time as determined by the dominant vector pattern (PC1) of days 1 and  $n$ .

$$\begin{aligned} a &= \begin{bmatrix} PC_{1,1} & 1 \\ PC_{n,1} & 1 \end{bmatrix}^{-1} \begin{bmatrix} X_1 \\ X_n \end{bmatrix} \\ c &= \begin{bmatrix} PC_{1,1} & 1 \\ PC_{n,1} & 1 \end{bmatrix}^{-1} \begin{bmatrix} Y_1 \\ Y_n \end{bmatrix} \end{aligned}$$

### 2.6 Number of Modes Evaluation

The focus of PCA is deriving few principal components from numerous variables. Therefore, a primary objective is determining the number of PCs to use for the analysis versus which to disregard. To support the use of only the first PC (as opposed to the first two or three PCs), an analysis was performed using seven PTT-100 datasets incorporating both the first and second PCs. The algorithm to derive the spatial functions was adjusted as follows:

$$\begin{aligned} X_i &= aPC_{i,1} + bPC_{i,2} \\ Y_i &= cPC_{i,1} + dPC_{i,2} \end{aligned}$$

so that,

$$\begin{aligned} a &= \begin{bmatrix} PC_{1,1} & PC_{1,2} \\ PC_{n,1} & PC_{n,2} \end{bmatrix}^{-1} \begin{bmatrix} X_1 \\ X_n \end{bmatrix} \\ c &= \begin{bmatrix} PC_{1,1} & PC_{1,2} \\ PC_{n,1} & PC_{n,2} \end{bmatrix}^{-1} \begin{bmatrix} Y_1 \\ Y_n \end{bmatrix} \end{aligned}$$

### 2.7 Error Metric

Light-based geolocation coordinate estimates were generated using two separate state-space models. Microwave Telemetry tags were deployed on blue marlin as part of a previous tagging study (Graves and Horodysky, 2010). The light records were formatted in a centralized tag database computer program, *Tagbase* (Lam and Tsonos, 2011) for input into the *TrackIt* model to generate light-based geolocations and an optimized horizontal



movement track.

For the Wildlife Computers Mk10 and MiniPAT tags, a “best fit” track was created courtesy of Stanford University (R. Kochevar, Stanford University, pers. comm.). First, an initial track was generated using the Wildlife Computers WC-GPE2 software and light data to derive longitude, and then SST data were added to derive latitude. The subsequent estimated geolocations were then put through a subsequent, proprietary state-space model that uses a separate algorithm to establish the “best fit” estimates. A combined track then was generated from these estimated geolocations. The accuracy of these geolocation estimates was determined in previous studies, which reported an average error of 145 km (roughly 1.3 degrees) longitude and 210 km (roughly 1.9 degrees latitude) (Stokesbury et al., 2004, Teo et al., 2004).

A summary of the data used for evaluating the PCA-based geolocation method is given in Table 1. A root mean square (RMS) metric is calculated by finding the difference between the daily latitude and longitude estimates from our geolocation method and the methods that are presently used that are described above. It should be noted that the truth data for this evaluation also contains large errors that contribute to the RMS metric. For each day  $j$ , RMS is calculated as:

$$\Delta lat = X_{geo} - X_{PCA}$$

$$\Delta lon = Y_{geo} - Y_{PCA}$$

$$^{\circ}Latitude(C_x) = 111.12 km$$

$$^{\circ}Longitude(C_y) = 111.12 \cos(lat)$$

$$RMSE = \sqrt{(\Delta lat \times C_x)^2 + (\Delta lon \times C_y)^2}$$

Error calculations within 1 degree (*ca.* 111 km) were considered “good,” and those within 2 degrees (*ca.* 222 km) considered “reasonable.” For each tag deployment, error calculated each day was plotted over time to determine any trends. These RMS values were then averaged over the course of the tagging duration and plotted over total days at large to determine any correlation (Figure 3 and Table 1).

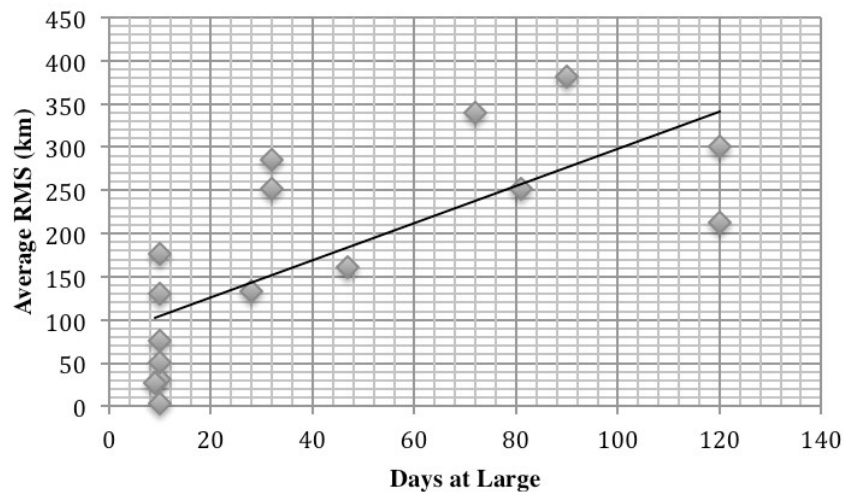


Figure 3. Root mean square (RMS) differences plotted according to the length of tag deployment. Days-at-large (DAL) ranged from 9 to 120 days. A correlation can be seen in the longer tag deployments yielding larger error calculations.

Table 1. Summary of results by species for the pop-up satellite archival tags used in the development and evaluation of the PCA geolocation model. The metric used to evaluate the model output for each tag dataset is root mean square (RMS). For species code, BUM = blue marlin, BLM = black marlin, WHM = white marlin, SWO = swordfish.

Species Code	Tag ID	Tag Model	RMS (km)	Max. Depth (m)	Avg. Depth (m)
BUM	41333	PTT100	75.88	-356	-15
BUM	24523	PTT100	3.52	-156	-14
BUM	34233	PTT100	31.47	-87	-7
BUM	84349	PTT100	193.19	-178	-19
BUM	59080	PTT100	27.69	-237	-22
BUM	84351	PTT100	50.27	-254	-30
BUM	84363	PTT100	176.14	-326	-21
BUM	35687	PTT100	130.02	-323	-28
BUM	111212	Mk10	370.38	-216	-38
BUM	112321	Mk10	284.74	-192	-38
BUM	111213	Mk10	245.36	-208	-39
BLM	111218	Mk10	213.47	-216	-38
BUM	112322	Mk10	133.50	-224	-54
BUM	112320	Mk10	381.01	-336	-62
BUM	112323	Mk10	251.90	-336	-60
WHM	126323	MiniPAT	251.72	-208	-55
WHM	116288	Mk10	161.20	-248	-59
SWO	61669	PTT100	N/A	-506	-218
SWO	86995	PTT100	N/A	-651	-357
SWO	61665	PTT100	N/A	-756	-308
SWO	88095	PTT100	N/A	-597	-295

### 3. Results

The PDT data received from the PAT style tags provided a total of eight temperature and depth records each day. The maximum depths, determined by looking over all data given by the MT and PDT records, varied between 80-200 m for all marlins and between 460-600 m for swordfish. Our supplementary analysis showed the use of both PCs reduced error only 43% of the time (Table 2). The total variance described by the first PC ranged from 87% to 94%, with the second PC explaining 2% to 12% of the total variance (Table 2). This is typical of PCA results, which often find large variances associated with the first PC and then an abrupt drop-off in second and subsequent PCs. Given just the PSAT data, these results support our original use of only the first PC as the most efficient method for representing this spatial data in order to estimate fish movement. Nevertheless, preliminary assessments of second and subsequent PCs should be conducted for future analyses of PSAT data and temperature profiles from high-resolution, data-assimilative models. The reduced error in three of the seven trials incorporating both PCs demonstrates the potential for more than one PC to capture the dynamically significant temperature variance among the original data and subsequently provide more accurate location estimates.

Table 2. Root mean square (RMS) differences between two PCA geolocation model variations and seven datasets from pop-up satellite archival tags deployed on blue marlin in the western North Atlantic that were used to determine the optimal number of principal components used in subsequent analyses.

<b>PTT ID</b>	<b>1 PC RMS (km)</b>	<b>2 PC RMS (km)</b>	<b>PC1 Var. (%)</b>	<b>PC2 Var. (%)</b>
24523	15.3	2.6	91	7.5
84351	55.5	95.4	87	9
34233	40.8	34.5	90	9
41333	75.9	83.1	88	10
59080	27.7	42.7	94	5.2
35687	156.5	183.9	90	6.6
84363	192.8	181.3	89	10

A total of 16 comparison tracks were generated with data from Microwave Telemetry PSATs (PTT-100 HR) and Wildlife Computers PATs (Mk10, MiniPAT) deployed on blue, black, and white marlin (Figures 4 and 5). These comparisons served to assess how well the PCA model works in relation to current light-based geolocation methods. The average of the daily root mean square (H) differences between the PCA-based geolocation method and the noisy “truth” estimate is 175.4 km with a standard deviation of 115.2 km (Table 1). RMS differences calculated each day were plotted over time for each deployment, although these plots did not reveal any significant trends in our results.

The results were broken down by tag model and corresponding light-based methods to which our PCA model results were compared to. Light data from the seven PTT-100 HRs deployed on blue marlin was run through *TrackIt* to generate a light-based “best fit” track. The comparison to the PCA generated estimates yielded an average RMS difference of 86 km with a standard deviation of 61.9 km. Data from a total of nine Mk10s and MiniPATs deployed on blue marlin (n=6), white marlin (n=2) and black marlin (n=1) were used in a separate comparison. The “best fit” track created from these tags used a two-step process incorporating light geolocation data directly from tag software into a state space model integrating satellite imagery-derived SST values. Geolocation estimates from the SST model and our PCA model had an average RMS difference of 254.8 km. Four tracks were made from swordfish tagged and released in the Caribbean Sea near the Cayman Islands (n=3) and in the Florida Straits (n=1) (Figure 6).

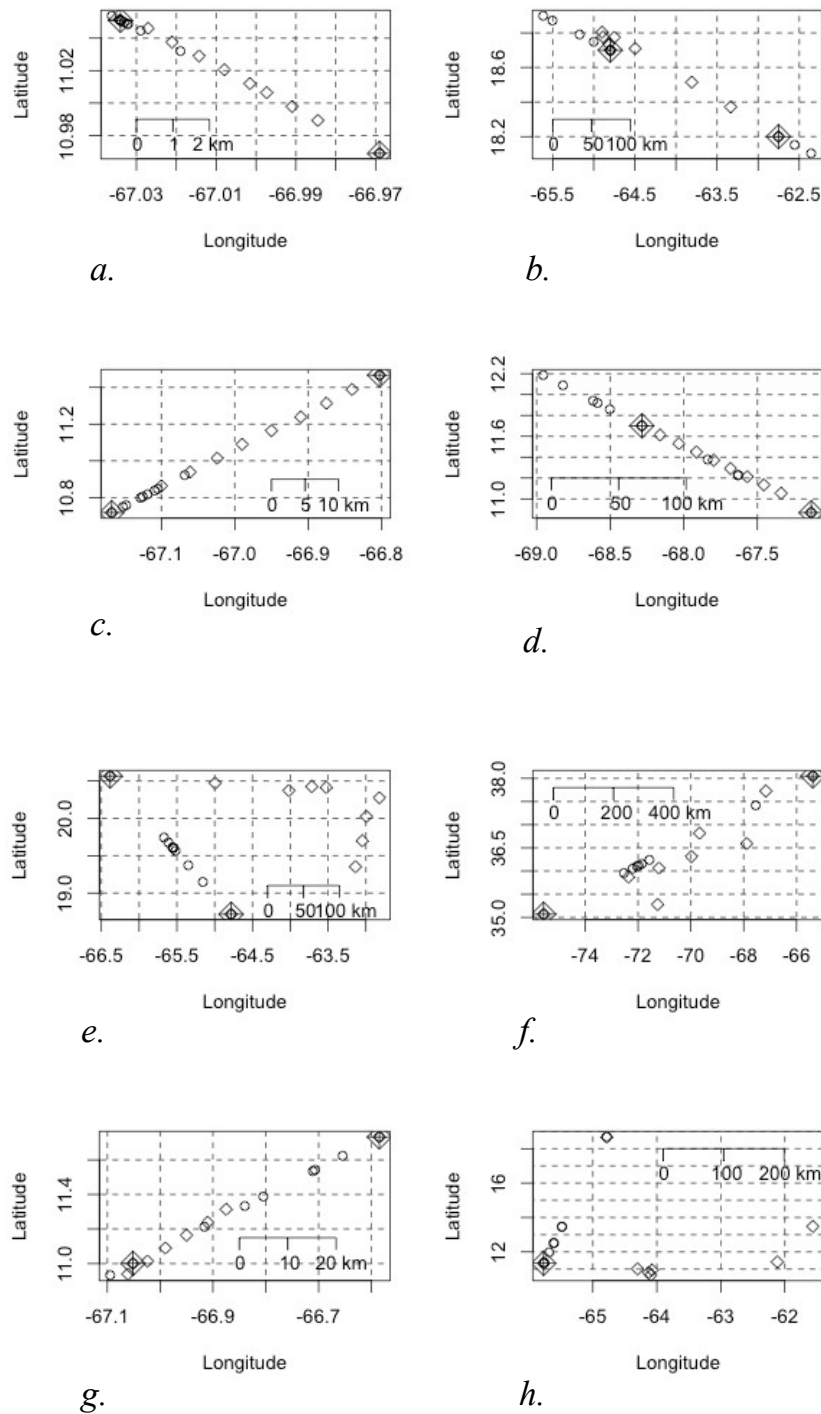


Figure 4. Estimates of blue marlin tagged with Microwave Telemetry, Inc. model PTT-100 tags (n=8) as given by the hydrographic PCA model (open circles) and the light-based model, *TrackIt* (open diamonds): (a) PTT 24523, (b) PTT 84351, (c) PTT 34233, (d) PTT 41333, (e) PTT 84363, (f) PTT 35687, (g) PTT 59080, (h) PTT 84349. The locations of fish release and final tag pop off are highlighted.

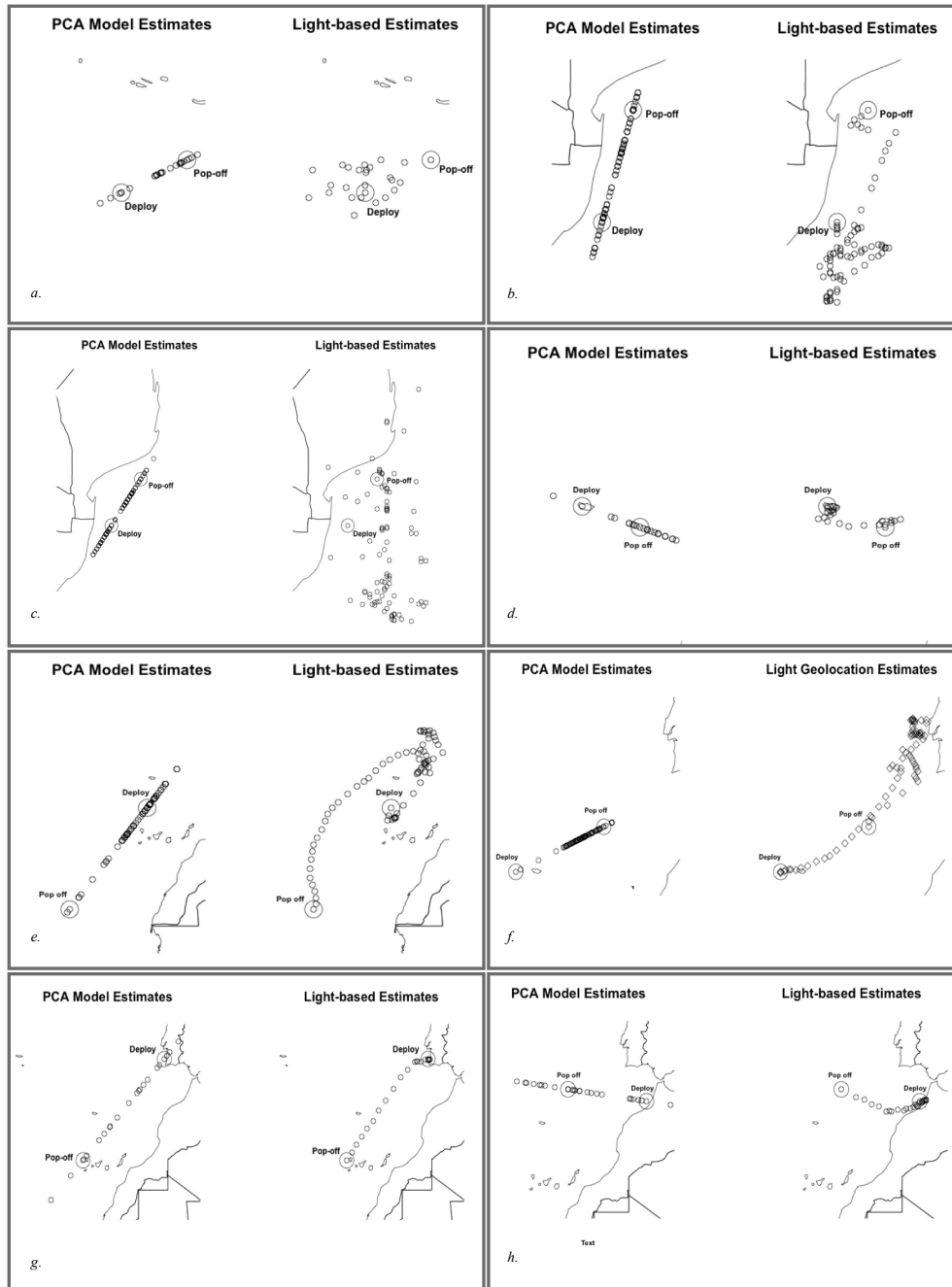


Figure 5. Coordinate estimates of marlin tagged with Wildlife Computers, Inc. model Mk10 PATs: (a) blue marlin 112321, (b) black marlin 111218, (c) blue marlin 111213, (d) blue marlin 112322, (e) blue marlin 112320, (f) blue marlin 112323, (g) white marlin 126323, (h) white marlin 126288. On the left is the track derived from the hydrographic PCA model, while on the right are the estimates from a light-based state-space model that uses the tag's raw global positioning estimates and incorporates satellite sea-surface temperature (SST).

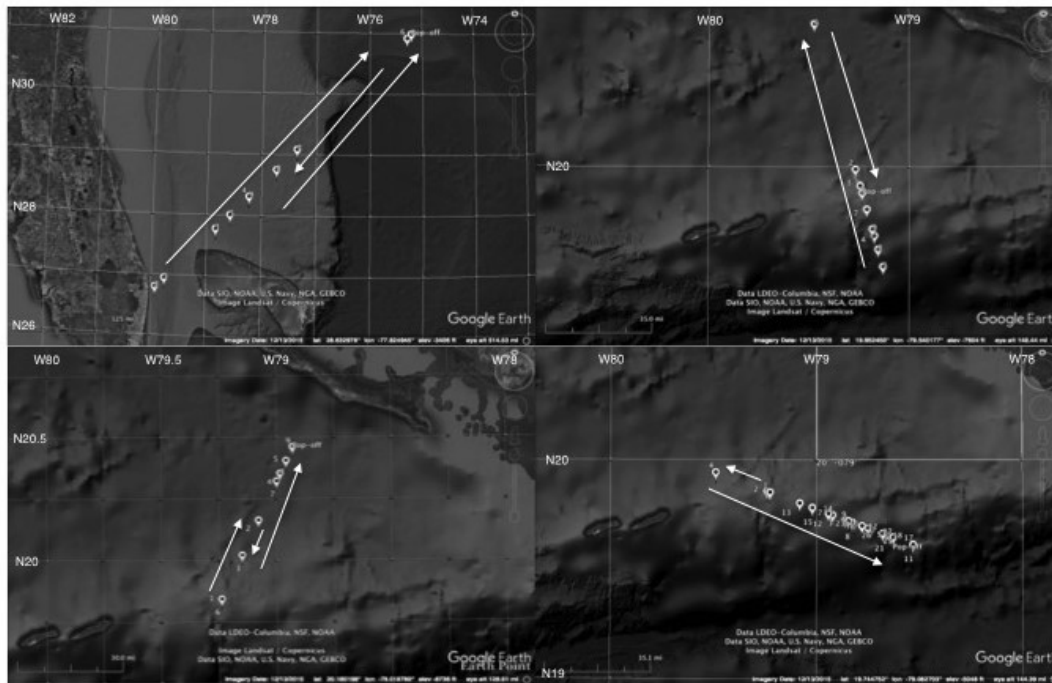


Figure 6. Fish tracks of swordfish tagged with Microwave Telemetry, Inc. model PTT-100 tags ( $n=4$ ) (a) PTT 61699 in the Florida Straits, and (b) PTT 61665, (c) PTT 86995, and (d) PTT 88095 near the Cayman Islands created from the hydrographic PCA model. The locations of fish release and final tag pop off are highlighted. The top image is a latitude and longitude plot with each estimate labeled numerically by day. The bottom image is a Google Earth map of the general track and the directions taken by the fish according to the PCA analysis.

#### 4. Discussion

The PCA model described represents a relatively new approach to tracking methodology using electronic archival tag data. Since the development of light-based methods to generate global positioning of tagged individuals, many studies have aimed at improving the precision of these methods, but very few have deviated from them with a new technique. Those studies that have utilized all environmental information given by archival tags support the use of models such as ours that incorporate the additional hydrographic data to estimate the most probable tracks (Skomal et al., 2004; Neuenfeldt et al., 2004). Such methodologies independent of light-level data are especially valued when tracking marine organisms that exhibit deep-diving behavior similar to that of swordfish.

The comparison analyses showed on average this methodology generates geolocation estimates within 174.3 km of the light-based models. Multiple studies have demonstrated that light-based geolocation can work well within this error margin (Ekstrom, 2004; Hill and Braun, 2001; Lam et al., 2010; Musyl et al., 2003; Nielsen et al., 2006; Nielsen and Sibert, 2007; Shaffer et al., 2005; Sibert et al., 2003; Teo et al., 2004). Such studies were all performed using tags deployed on near-surface fish or with simulated data. Furthermore, their results reported light-based estimations of over 400 km (up to 4 degrees latitude) off the actual position when in temporal proximity to the equinox (Schaefer & Fuller, 2006; Lam et al., 2010).

The PCA model is a unique approach to generating coordinate estimates using temperature at depth data rather than light levels. As the SST-incorporated Kalman filter illustrates, temperature observations improve the accuracy of the light level-based model by analyzing movement in relation to oceanographic variability (Teo et al., 2004; Nielsen et al., 2006). In contrast, the PCA model allows the utilization of the entire water column rather than just the surface layer. Swordfish provide us with valuable data due to their large crepuscular vertical movements. As the tagged swordfish makes their vertical migration recording temperature and pressure, they allow a high rate programmed tag, such as the PTT-100's used in this study, to describe an extensive temperature profile of that water column. These profiles attribute to the pattern of variability derived by EOF analysis needed to predict accurate movement from one geographic point to the next (Figure 2). The error associated with averaging the thermal profiles for data input is low enough to allow the analysis to capture the necessary temperature signals.

Temperature and depth recordings by a tag are functions of local hydrography and movement activity. Much like the SST models, we are able to link the geographical positions with the physical environment described by the tag data (Shaffer et al., 2005). The analysis makes use of distinct differences in the temperature profile structures to assess how far the fish has moved horizontally. These observations are most effective when the temperature signal in a region is spatially stratified (Lam et al., 2011). Therefore the model will work best when tags are exposed to frontal zones where hydrographic properties vary dramatically over a relatively short distance. This allows the temperature profiles to exhibit a stronger signal for the PCA to attribute to distinct locations. Frontal regions include along shelf breaks or within seismic activity where thermal intrusions take place. Studies show a strong relationship between fronts and fish abundance (Podesta et al., 1993).

Hand selecting the data received by the tags is a necessary step in order to ensure the signature of the water column, including the thermocline, is detected in our analysis. For instance, the marlin data selected for the comparison analysis provided deep enough dives to capture thermal profiles that allowed our PCA analysis to work as efficiently as possible. Furthermore, the initial and final thermal profiles for which the known satellite locations are attributed to need to be complete, e.g., not a partial dive. The pattern of variability the PCA uses to "locate" the fish each day is derived from the difference in these profiles. This is evident in our results from the PAT style tags where the binned measurements most likely did not allow the PCA to capture the full thermal stratification of each day.

As the first attempt to apply this model to generate location estimates, there are limitations evident in our results that should be addressed in future studies. The first of which is the assumption that the animal is traveling in a straight path. A simple bilinear model is assumed for the PC distribution in space, because only two locations are definitively known. If more profile information is available from models, climatology, or surveys, then a more complicated model could be formulated for the PC distribution in space. The algorithm used here can be generalized given different data sources. Climatological hydrographic data, or output from a data assimilative, numerical ocean circulation model such as HYCOM (<http://hycom.org>), or World Ocean Database Project ([www.nodc.noaa.gov](http://www.nodc.noaa.gov)) can be used to calculate the EOFs of the temperature profiles. Spatial maps of PCs can be constructed from these large datasets. These maps should therefore be used instead of SST values in the tracking algorithms incorporating the Kalman filter. Estimation methods should be sufficient by incorporating all of the data.

A second limitation is our utilization of the tag data available to us. Ideally, a single tag

model and light-based method used for our comparison analysis would have provided more consistent results. However, the increased number of trials run allowed a more thorough analysis of this model's performance. The two tag models used for this study did provide an opportunity to assess what data works best with our model. The tag models and respective light-based geolocation estimates yielded a difference of 168.8 km in average RMS difference values (Table 2). The PTT-100 HR tags provide the raw temperature and pressure (converted to depth) data at frequent intervals necessary to construct a precise vertical temperature profile. However, the low-resolution light data does not allow light-based models to work to their best ability. The PAT light data and IGFA/Stanford light-based model provided statistically robust geolocation estimates. However, the histogram-style binned data comprised of summarized temperature and depth records limited the ability of the PCA to capture accurate thermal signatures.

Those large errors associated with light-based geolocations did inhibit our ability to fully ground-truth the PCA method in this study. Therefore, double-tagging studies using both satellite telemetry tags and archival tags should be considered for future assessments of this model. For example, the TDR10 tag model (Wildlife Computers) is equipped with FastLoc technology allowing the tag to take sub-second "snapshot" of the GPS signal to take advantage of any brief surfacing events. The high rate PTT-100 model used in this study provides substantial hydrographic data from which our PCA analysis can derive effective coordinates. The ample data provided by archival tags coupled with the geographical precision of the TDR10 (within 20-75 m) would allow for direct comparisons between our model estimates and actual geolocations. An increased effort in sharing tag data – preferably from double-tagged animals – would provide a more objective way of determining the accuracy of this model. In addition to the gathering of more tag data, this model should be expanded to incorporate auxiliary environmental information. Hydrographic data can be accessed from expendable bathythermographs (XBTs), by profiling ARGOS floats deployed throughout most oceanic regions, or through free access to the World Ocean Database (WOD) provided by the National Oceanographic Data Center ([www.nodc.noaa.gov](http://www.nodc.noaa.gov)). A PCA performed on temperature profiles from a high-resolution data-assimilative ocean circulation model would give EOF modes that better describe the pattern of thermal variability among the water masses sampled. The incorporation of this data into the analysis of tag data would generate estimations, in theory, closer to actual animal locations.

Although methods exist to approximate light cycles from diel behavior, thus allowing the use of more traditional light-based geolocation models (e.g., Lam et al., 2014 for bigeye tuna *Thunnus obesus*), our PCA-based model provides an additional use for archived temperature data. Future studies should integrate light-level geolocation methods with our model, which could create an improved tracking system for animals that stay within the photic zone. User-friendly programs, such as *Tagbase* and *TrackIt*, have demonstrated their readiness and ease of incorporating additional data streams such as SST (Lam et al., 2010). The statistical model framework of this analysis merely requires the temperature and depth data recorded by PSATs be organized into daily sets. If the PCA algorithm was to be integrated either as a function or data export option, daily coordinate estimates based on the temperature and depth data rather than light levels could be easily accessed. Such an integration of comparative statistical methods applied to a single data set will advance the utilization of costly tagging operations, especially given recent funding constraints on research applied to fisheries management. Ultimately, an XBT survey of the study area at the time of tagging would provide the best data for obtaining the most accurate estimates of animal locations.



## 5. Conclusion

In conclusion, a new PCA-based geolocation method incorporating all of the temperature-depth data produced reasonable daily geolocation estimates of satellite-tagged fish. The method utilizes all of the subsurface temperature data, which are not distorted by weather conditions, to provide a stronger signal for tracking. The PCs calculated using temperature-profiles from data-assimilative models such as HYCOM could eventually replace SST in the current Kalman filter geolocation methods. Integration of this PCA-based geolocation methods with both the best available estimates of the ocean temperature at the time of tag deployment and the existing light-based geolocation models would provide additional information on fine-scale movement of tagged fish.

## Acknowledgements

The authors thank J.E. Graves (Virginia Institute of Marine Science) and A.Z. Horodysky (Hampton University) for the data from eight PTT-100 HR tags deployed on blue marlin, as well as the International Game and Fishing Association in collaboration with Stanford University for the sharing of PAT data and geolocation tracks of eight tagged blue, black, and white marlin. A.J.M. acknowledges the support of the *Consortium for Advanced Research into the Transport of Hydrocarbons in the Environment* (CARTHE) funded by the BP/Gulf of Mexico Research Initiative.

## References

- Dewar, H., et al. 2011. Movements and behaviors of swordfish in the Atlantic and Pacific Oceans examined using pop-up satellite archival tags. *Fish Oceanogr*, 20, 219-241, [http://dx.doi.org/10.1007/978-1-4020-9640-2\\_9](http://dx.doi.org/10.1007/978-1-4020-9640-2_9).
- Ekstrom, P. A. 2004. An advance in geolocation by light. *Mem Nat Inst Polar Res Special Issue*, 58, 210-226.
- Fenton, J. 2012. Post-release survival and habitat utilization of juvenile swordfish in the Florida Straits. M.S. thesis, Nova Southeastern University Oceanographic Center, 107 pp.
- Graves, J. E., & Horodysky, A. Z. 2010. Asymmetric conservation benefits of circle hooks in multispecies billfish recreational fisheries: a synthesis of hook performance and analysis of blue marlin (*Makaira nigricans*) postrelease survival. *Fish Bull*, 108(4), 433-441.
- Hill, R. D. Braun, M. J. 2001. Geolocation by light level – the next step: latitude, *Electronic Tagging and Tracking in Marine Fisheries*, J.R. Sibert and J. Nielsen, Ed., Kluwer Academic Publishers, p 315-330. [https://doi.org/10.1007/978-94-017-1402-0\\_17](https://doi.org/10.1007/978-94-017-1402-0_17)
- Horodysky, A. Z., et al. 2007. Habitat utilization and vertical movements of white marlin (*Tetrapturus albidus*) released from commercial and recreational fishing gears in the western North Atlantic Ocean: inferences from short duration pop-up archival satellite tags. *Fish Oceanogr*, 16(3), 240-256. <https://doi.org/10.1111/j.1365-2419.2007.00419.x>
- Lam, C. H., & Tsonos, V. M. 2011. Integrated management and visualization of electronic tag data with Tagbase. *PLoS ONE*, 6(7): e21810. <http://dx.doi.org/10.1371/journal.pone.0021810>.
- Lam, C. H., et al. 2008. Improving light and temperature based geolocation by unscented Kalman filtering. *Fish Res*, 91, 15-25. <https://doi.org/10.1016/j.fishres.2007.11.002>
- Lam, C. H., et al. 2010. Incorporating sea-surface temperature to the light based geolocation

- model TrackIt. *Mar Ecol-Prog Ser*, 419, 71-84. <http://dx.doi.org/10.3354/meps08862>.
- Lam, C. H., Galuardi, B., & Lutcavage, M. E. 2014. Movements and oceanographic associations of bigeye tuna (*Thunnus obesus*) in the Northwest Atlantic. *Can J Fish Aquat Sci*, 71, 1-15. <http://dx.doi.org/10.1139/cjfas-2013-0511>.
- Leaman, K. D., et al. 1989. The average distribution of volume transport and potential vorticity with temperature at three sections across the Gulf Stream. *J Phys Oceanogr*, 19, 36-51. [https://doi.org/10.1175/1520-0485\(1989\)019<0036:TADOVT>2.0.CO;2](https://doi.org/10.1175/1520-0485(1989)019<0036:TADOVT>2.0.CO;2)
- Lerner, J. D., et al. 2013. Swordfish (*Xiphias gladius*) vertical distribution and habitat use in relation to diel and lunar cycles in the western North Atlantic. *Trans Am Fish Soc*, 142, 95-104. <https://doi.org/10.1080/00028487.2012.720629>
- Lisovski, S., et al. 2012. Geolocation by light: accuracy and precision affected by environmental factors. *Method Ecol Evol*, 3, 503-612. <https://doi.org/10.1111/j.2041-210X.2012.00185.x>
- Mariano, A. J., et al. 1996. Principal component analysis of biological and physical variability in a Gulf Stream meander crest. *Deep-Sea Res Pt I*, 43, 1531-1565. [https://doi.org/10.1016/S0967-0637\(96\)00021-0](https://doi.org/10.1016/S0967-0637(96)00021-0)
- Musyl, M. K., et al. 2003. Vertical movements of bigeye tuna (*Thunnus obesus*) associated with islands, buoys, and seamounts near the main Hawaiian Islands from archival tagging data. *Fish Oceanogr*, 12(3), 152-169. <https://doi.org/10.1046/j.1365-2419.2003.00229.x>
- Neilson, J. D., & Smith, S. 2010. Update on the Canadian program for pop-up satellite archival tagging of swordfish. *Coll Vol Sci Pap. ICCAT*, 65, 229-240.
- Neuenfedt, S., et al. 2004. A method to geolocate eastern Baltic cod by using Data Storage Tags (DSTs). *ICES C.M. 2004*. L:06, 1-14.
- Nielsen, A., & Sibert, J. R. 2007. State-space model for light-based tracking of marine animals. *Can J Fish Aquat Sci*, 64, 1055-1068. <http://dx.doi.org/10.1139/F07-064>.
- Nielsen, A., et al. 2006. Deriving light-based geolocation by including sea surface temperature. *Fish Oceanogr*, 15, 314-325. <https://doi.org/10.1111/j.1365-2419.2005.00401.x>
- Palko, B. J., et al. 1981. Synopsis of the biology of the swordfish, *Xiphias gladius* Linnaeus, NOAA Tech Rep 441, NMFS circular, U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service.
- Podesta, G. P., et al. 1993. Exploring the association between swordfish catch rates and thermal fronts on United-States longline grounds in the Western North-Atlantic. *Cont Shelf Res*, 13(2-3), 253-27. [https://doi.org/10.1016/0278-4343\(93\)90109-B](https://doi.org/10.1016/0278-4343(93)90109-B)
- Sepulveda, C. A., et al. 2010. Fine-scale movements of the swordfish *Xiphias gladius* in Southern California Bight. *Fish Oceanogr*, 19(4), 279-289. <https://doi.org/10.1111/j.1365-2419.2010.00543.x>
- Shaffer, S. A., et al. 2005: Comparison of light- and SST-based geolocation with satellite telemetry in free-ranging albatrosses. *Mar Biol*, 147, 833-843. <https://doi.org/10.1007/s00227-005-1631-8>
- Sibert, J. R., et al. 2003. Horizontal movements of bigeye tuna (*Thunnus obesus*) near Hawaii determined by Kalman filter analysis of archival tagging data. *Fish Oceanogr*, 12, 141-151.

<https://doi.org/10.1046/j.1365-2419.2003.00228.x>

Skomal, G. B., et al. 2009. Transequatorial migrations by basking sharks in the western Atlantic Ocean. *Curr Biol*, 19, 1019-1022. <https://doi.org/10.1016/j.cub.2009.04.019>

Stokesbury, M. J. W., et al. 2004. Movement of Atlantic Bluefin tuna (*Thunnus thynnus*) as determined by satellite tagging experiments initiated off New England. *Can J Fish Aquat Sci*, 61(10), 1976-1987. <https://doi.org/10.1139/f04-130>

Teo, S. L. H., et al. 2004. Validation of geolocation estimates based on light and sea surface temperature from electronic tags. *Mar Ecol-Prog Ser*, 283, 81-91. <https://doi.org/10.3354/meps283081>

Appendix 1. Deployment details and summary data for individual pop-up satellite archival tag (PSAT) datasets used in the PCA geolocation analysis. For species code, BUM = blue marlin, BLM = black marlin, WHM = white marlin; for geographic region of tag deployment, VZ = Venezuela, VI = Virgin Islands, NC = North Carolina, ZA = South Africa, POR = Portugal, MR = Morocco, US = United States, CZ = Cayman Islands. Also, DAL = days-at-large; ND = net displacement in km.

Species	Tag ID	Tag	Tag	Region	DAL	Track Duration	Initial	Initial	Final	Final	ND
Code		Man.	Model				Lat.	Lon.	Lat.	Lon.	
BUM	41333	MWT	PTT100	VZ	10	3/17/08 - 3/27/08	10.87N	67.14W	11.7N	68.28W	139
BUM	24523	MWT	PTT100	VZ	10	5/15/08 - 5/25/08	10.97N	66.97W	11.05N	67.03W	10
BUM	34233	MWT	PTT100	VZ	10	5/17/08 - 5/27/08	10.72N	67.17W	11.47N	66.8W	87
BUM	84349	MWT	PTT100	VZ	8	9/10/08 - 9/18/08	11.3N	65.77W	18.68N	64.78W	264
BUM	59080	MWT	PTT100	VZ	9	4/18/09 - 4/27/09	11.0N	67.05W	11.73N	66.59W	96
BUM	84351	MWT	PTT100	VI	10	9/11/08 - 9/21/08	18.7N	64.8W	18.2N	62.75W	197
BUM	84363	MWT	PTT100	VI	10	9/30/09 - 10/10/09	18.72N	64.78W	20.57N	66.38W	271
BUM	35687	MWT	PTT100	NC	10	6/22/08 - 7/2/08	35.07N	75.57W	38.05N	65.38W	943
BUM	111212	WC	Mk10	ZA	72	2/25/12 - 4/30/12	30.75S	31.4E	27.31S	34.4E	322
BUM	112321	WC	Mk10	ZA	32	3/30/12 - 4/25/12	29.25S	34.2E	26.12S	35.5E	558
BUM	111213	WC	Mk10	ZA	120	5/6/12 - 8/28/12	27.08S	33.6E	25.47S	34.8E	170
BLM	111218	WC	Mk10	ZA	120	5/3/12 - 8/28/12	28.71S	32.9E	25.99S	33.8E	413
BUM	112322	WC	Mk10	POR	28	8/4/12 - 8/27/12	17.1S	32.76E	15.2S	32.18E	183
BUM	112320	WC	Mk10	POR	90	8/12/12 - 11/4/12	17.4S	30.45E	24.0S	22.62E	2263
BUM	112323	WC	Mk10	POR	81	8/26/13 - 11/7/13	18.2S	32.67E	13.0S	34.68E	676
WHM	126323	WC	MiniPAT	MR	32	10/15/13 - 11/5/13	9.1S	37.56E	18.3S	28.62E	1819
WHM	116288	WC	Mk10	MR	47	10/15/13 - 11/11/13	7.0S	34.4E	14.5S	35.31E	889
SWO	61669	MWT	PTT100	FL	10	12/16/11 - 12/26/11	26.77N	79.76W	30.81N	75.21W	58
SWO	86995	MWT	PTT100	CZ	10	4/1/12 - 4/10/12	19.8N	79.24W	20.42N	78.93W	77
SWO	61665	MWT	PTT100	CZ	10	4/21/13 - 4/30/13	19.8N	79.22W	19.83N	79.23W	63
SWO	88095	MWT	PTT100	CZ	30	4/12/14 - 5/13/14	19.80N	79.23W	19.6N	78.63W	68

### Copyright Disclaimer

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).