

## Enrichment of trajectories with environmental data, and standardisation of tagging data using NetCDF

A.E. Nieblas<sup>1</sup>, J. Barde<sup>2</sup>, S. Bernard<sup>3</sup>, T Imzilen<sup>4</sup>, V. Kerzerho<sup>3</sup>, T. Rouyer<sup>5</sup>, S. Bonhommeau<sup>6</sup>

<sup>1</sup> IRD, UMR MARBEC (IRD/Ifremer/Univ.Montpellier/CNRS), IRD Réunion, 97744 St Denis, La Réunion, France; [anne.elise.nieblas@gmail.com](mailto:anne.elise.nieblas@gmail.com)

<sup>2</sup> IRD, UMR MARBEC (IRD/Ifremer/Univ.Montpellier/CNRS), IOC, Rue de l'Institut, Ebène, Maurice

<sup>3</sup> LIRMM, CNRS/University of Montpellier, 161 rue Ada 34095 Montpellier Cedex 5, France

<sup>4</sup> IRD - UMR MARBEC 248, Av. Jean Monnet, 34200 Sète, France

<sup>5</sup> IFREMER - UMR MARBEC 248, Av. Jean Monnet, 34200 Sète, France

<sup>6</sup> IFREMER- DOI, rue Jean Bertho, 97822 LE PORT CEDEX, La Réunion, France

### Abstract

Geolocalisation and trajectory analysis can aid in understanding the ecological processes driving an organism. By associating satellite-derived environmental data with individual trajectories of electronically-tagged organisms, it could be possible to define environmental characteristics of the tagged species' functional habitats (i.e., reproduction, nutrition). These data can also help identify biotic envelopes or predict the effects of climate change on marine species distributions. The objective of the present work, undertaken as a collaboration between IFREMER and IRD, is to standardize electronic tag data files into network common data format (NetCDF) format, following the standards defined within the POPSTAR project for tag data (doi <http://dx.doi.org/10.13155/34980>), and enrich the positional data with satellite-derived surface environment data (e.g., sea surface temperature, salinity, sea level) and model-derived environment data at observed depths (e.g., temperature, salinity, currents). We accounted for positional uncertainty using 95%, 75%, and 50% uncertainty polygons around the estimated positions of individuals. We summarised environmental conditions within these uncertainty polygons using the mean, minimum, maximum, quantiles, and standard deviation of the selected enrichment parameter. We generated generic codes to enable the automatic enrichment of position data from points and polygons. Furthermore, we developed algorithms to convert the enriched data into NetCDF format for subsequent visualisation and analysis.

## Introduction

The global objective of IFREMER's POPSTAR project is to develop a system of electronic tags for marine species that are capable of collecting information on the individual organism and its environment, which data are then transmitted via satellite. POPSTAR aims to improve upon current tag technology by decreasing the size of the tag (thus reducing its impact on the organism), improving the rate of data transmission, expanding data collection of environmental variables and vastly reducing the price of each tag.

Geolocalisation and trajectory analysis can aid in understanding the ecological processes driving an organism. Therefore, it is important to be able to use the measurements of the sensors embedded in the tag to recalculate *a posteriori* the trajectories followed by the animal during the survey period. Furthermore, by associating satellite-derived environmental data with individual trajectories, it could be possible to define environmental characteristics of the tagged species' functional habitats (i.e., reproduction, nutrition). These data can also help identify biotic envelopes or predict the effects of climate change on marine species distributions.

### Objective 1 : Identification of environmental data sources

POPSTAR field campaigns in 2017 deployed tags on Atlantic bluefin tuna (*Thunnus thynnus*) and marlin (*Makaira indica*). The objective of the present work, undertaken as a collaboration between IFREMER and IRD, is to standardize the resulting data files, and enrich the positional data with satellite-derived surface environment data (e.g., sea surface temperature, salinity, sea level) and model-derived environment data at observed depths (e.g., temperature, salinity, currents).

### Objective 2 : Accounting for uncertainty

As a second step, we accounted for the uncertainty of the estimated positions of individuals. The geolocalisation algorithm gives an envelope of position uncertainty for each position. This uncertainty must be accounted for when estimating the value of environmental conditions at each position (e.g., average over the area, minimum, maximum, quantiles).

### Objective 3: Automatic enrichment

The current collaboration aims to generate generic codes to enable the automatic enrichment of position data. Furthermore, the enriched data will be transformed into a network common data format (NetCDF) for subsequent visualisation and analysis, following the standards defined within the POPSTAR project for tag data (doi <http://dx.doi.org/10.13155/34980>).

## Data

### Tagging data

Tagging data that have been provided for this work are based on tags that were deployed in the context of the POPSTAR project, and were developed by Wildlife Computers. No raw tagging data are currently available for this work, and the data that are available have been post-processed by Wildlife Computers to calculate the maximum likelihood points and their associated positional uncertainties. One data file was provided with geolocation data derived from a bluefin tuna that was released in the vicinity of Malta (35.9375°N, 14.3754°E) in the Mediterranean Sea on 4 May 2018, and whose tag detached on 2 August 2018 (about three months of data).

Wildlife Computers tag file formats include keyhole markup language zipped (.kmz), NetCDF (.nc), and comma separated values (.csv), and each format contains different information. The .kmz files include position of the Maximum Likelihood Points (n=917), as well as polygons indicating the uncertainty in the geolocation at 50%, 75% and 95% (n=46 for each level of uncertainty, i.e., one polygon every 2 days).

The .nc file contains gridded data of the probability of the tuna's position at 12-hourly (n=181) and monthly (n=4) intervals and at a spatial resolution of 0.25° x 0.25°, from which uncertainty polygons can then be calculated.

The .csv file is a data frame of the Maximum Likelihood Points (n=917) that also includes information on the observation type (i.e., either light or temperature, the metrics used to aid the geolocation algorithm), observed *in situ* sea water temperature, satellite sea surface temperature (SST; NOAA OI SST V2 High Resolution) coincident with the observed *in situ* temperature, observed depth, bathymetry depth (from ETOPO1-Bedrock), observation LL (MSS), observation score for each point and the times of sunrise and sunset for each day.

## Conversion to POPSTAR NetCDF format

We followed the standards outlined in the POPSTAR NetCDF format reference manual (doi <http://dx.doi.org/10.13155/34980>) to convert the WildLife data to NetCDF, and created the Tag2NetCDF conversion script. As inputs to this script, we used the Wildlife .csv for point longitudes, latitudes, sea water temperature, depth, sunrise and sunset. We extracted the longitudes and latitudes of the polygon layers for each of the three uncertainty levels (50%, 75%, and 95%) in the Wildlife .kmz file. Maximum likelihood points were converted to the coordinate variables LATITUDE and LONGITUDE, each with a TIME dimension (here: n=917). We altered the original POPSTAR NetCDF standards from one coordinate variable, POSITION\_UNCERTAINTY, to a total of six coordinate variables, i.e., LATITUDE\_UNCERTAINTY\_# and LONGITUDE\_UNCERTAINTY\_# for each of the three different uncertainty levels. These uncertainty variables have the dimensions N\_VALUES\_#, where the '#' represents the longest polygon vector for each uncertainty level; and TIME\_UNCERTAINTY, which is the time values of the polygons (here: n=46).

Further changes included fixing DEPTH, TEMP (*in situ* sea water temperature), SUNRISE, and SUNSET as measurement variables, as they will always be included in tagging data. Each is defined by the TIME dimension.

NetCDF Climate and Forecast (CF) standards were followed for metadata on [global](#) and [variable attributes](#).

## Satellite data and ocean model outputs

The potential environmental variables that have been identified for use are listed in [Table 1](#). In identifying environmental data, all effort has been made to select the highest temporal and spatial resolution, including near real time constraints. Selected data are freely available NetCDF files that can be accessed directly with a URL. All data are sourced from the [OPeNDAP server](#) of the Jet Propulsion Laboratory's Physical Oceanography Distributed Active Archive Center (PO.DAAC).

**Table 1** Potential environmental variables identified for use to enrich trajectory data. All data are sourced from the Jet Propulsion Laboratory's Physical Oceanography Distributed Active Archive Center.

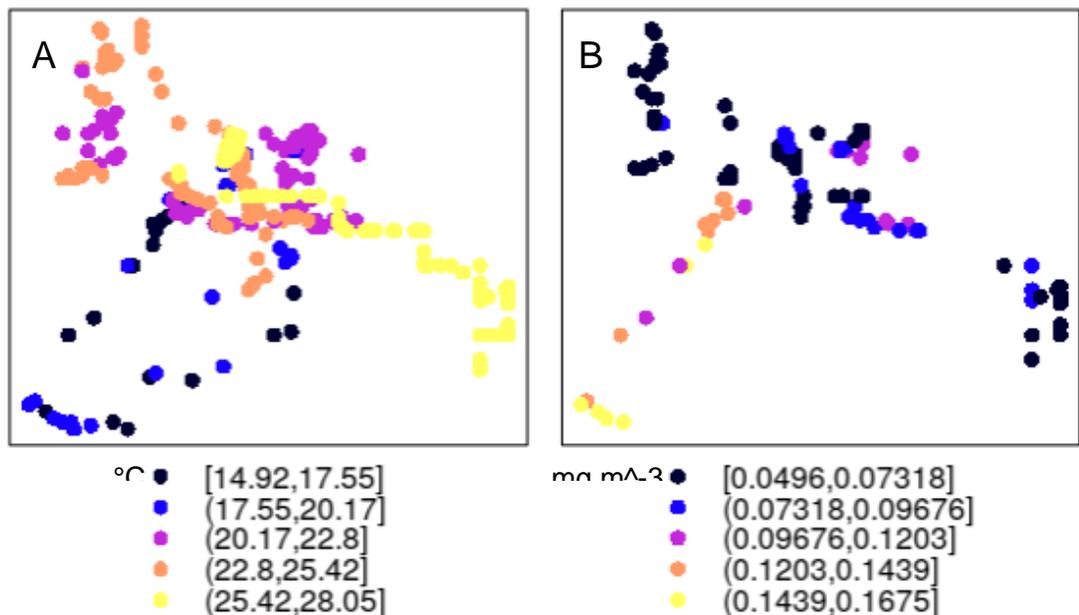
Data type	Instrument/Model	Temporal resolution	Spatial resolution	OPeNDAP directory	Description
Sea surface temperature	Modis Aqua	1 day	4 km	<a href="https://opendap.jpl.nasa.gov/opendap/OceanTemperature/modis/L3/aqua/4um/v2014.0/4km/daily/contents.html">https://opendap.jpl.nasa.gov/opendap/OceanTemperature/modis/L3/aqua/4um/v2014.0/4km/daily/contents.html</a>	Level 3 Thermal IR Nighttime
Chlorophyll a	Modis Aqua	1 day	4 km	<a href="https://opendap.jpl.nasa.gov/opendap/OceanTemperature/modis/L3/aqua/chlA/v2014.0/4km/daily/contents.html">https://opendap.jpl.nasa.gov/opendap/OceanTemperature/modis/L3/aqua/chlA/v2014.0/4km/daily/contents.html</a>	Level 3
Sea surface salinity	SMAP L2B	8-day	0.25°x 0.25°	<a href="https://opendap.jpl.nasa.gov/opendap/SalinityDensity/smap/L3/JPL/V4/8day_running/2018/contents.html">https://opendap.jpl.nasa.gov/opendap/SalinityDensity/smap/L3/JPL/V4/8day_running/2018/contents.html</a>	Level 3, SMAP grid averaged SSS/WSPD v4
Sea Level Anomaly	Kriging technique	1 day	0.25°x 0.25°	<a href="https://opendap.jpl.nasa.gov/opendap/SeaSurfaceTopography/merged_alt/L4/cdr_grid_interim/contents.html">https://opendap.jpl.nasa.gov/opendap/SeaSurfaceTopography/merged_alt/L4/cdr_grid_interim/contents.html</a>	Level 4, gridded altimetry from Kriging technique
Sea Surface Velocity (u,v)	Oscar	5 day	1° x 1°	<a href="https://opendap.jpl.nasa.gov:443/opendap/OceanCirculation/oscar/L4/oscar_1_deg/world_oscar_vel_5d2018.nc.gz">https://opendap.jpl.nasa.gov:443/opendap/OceanCirculation/oscar/L4/oscar_1_deg/world_oscar_vel_5d2018.nc.gz</a>	Estimates of the horizontal near-surface currents
Bathymetry	ETOPO1	-	0.0167°x 0.0167°	<a href="http://new-bock.pmel.noaa.gov:8185/threds/dodsC/data/PMEL/smith_san_dwell_topo_v8_2.nc">http://new-bock.pmel.noaa.gov:8185/threds/dodsC/data/PMEL/smith_san_dwell_topo_v8_2.nc</a>	1/30-degree topography and bathymetry

## Trajectory enrichment

Enrichment of the trajectories of fish aggregating devices (FADs) has previously been explored by Taha Imzilen of IRD (cf [Github repository](#); [Environmental\\_Data\\_Enrichment\\_v2018.R](#)). By reading a shapefile and identifying an environmental data source OPeNDAP URL, these scripts can associate environmental data to polygons, lines or points. Building upon these codes, we introduced the trajectory data from the Wildlife Computers tag. We then enriched the maximum likelihood points and uncertainty polygons with satellite and three-dimensional model data. The Imzilen code requires environmental data to be a single file with a time dimension. Thus, we modified the code to include a function ([EnvList2NetCDF.R](#)) that converts multiple daily environmental files to one NetCDF file that includes all days in the time period of interest. In addition, as the environmental files often have a global spatial scale, this function limits their spatial range to that of the tagging data.

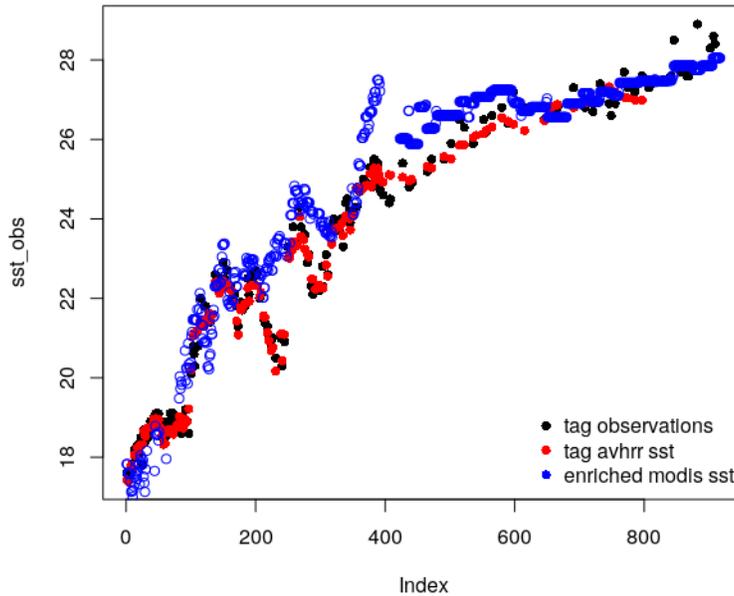
## Maximum likelihood points

In order to enrich the maximum likelihood points of the tag trajectory, first, a function was written to create a list of OPeNDAP URLs that correspond with the environmental variables of interest and the date range of the tagging data. Then, LATITUDE and LONGITUDE coordinate variables of the NetCDF were converted into a shapefile. Finally, the tagging points shapefile and the opendap url list were read into the adapted [Environmental\\_Data\\_Enrichment\\_VPOPSTAR.R](#) script to enrich the tag points (e.g., [Figure 1](#) for an example of enriched tag positions).



**Figure 1:** The maximum likelihood positions as calculated by Wildlife Computers and the A) sea surface temperature and B) surface chlorophyll a at each point. Note that there are several positions where satellite data are missing for a particular date.

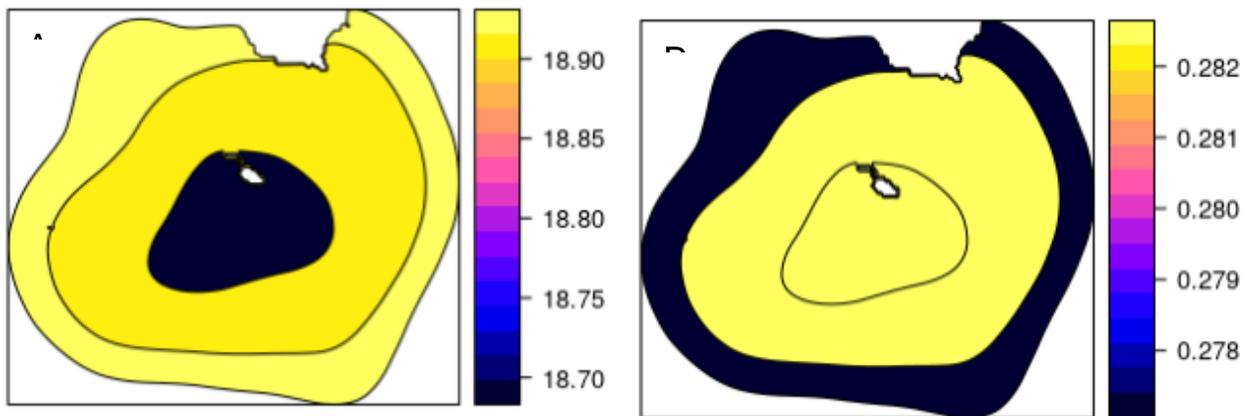
To verify that the enriched data are reasonable, we compared the SST values that were extracted during the present enrichment process from MODIS-Aqua data and compared them to those that were observed with the Wildlife tag, as well as those that were extracted by Wildlife from AVHRR data ([Figure 2](#)).



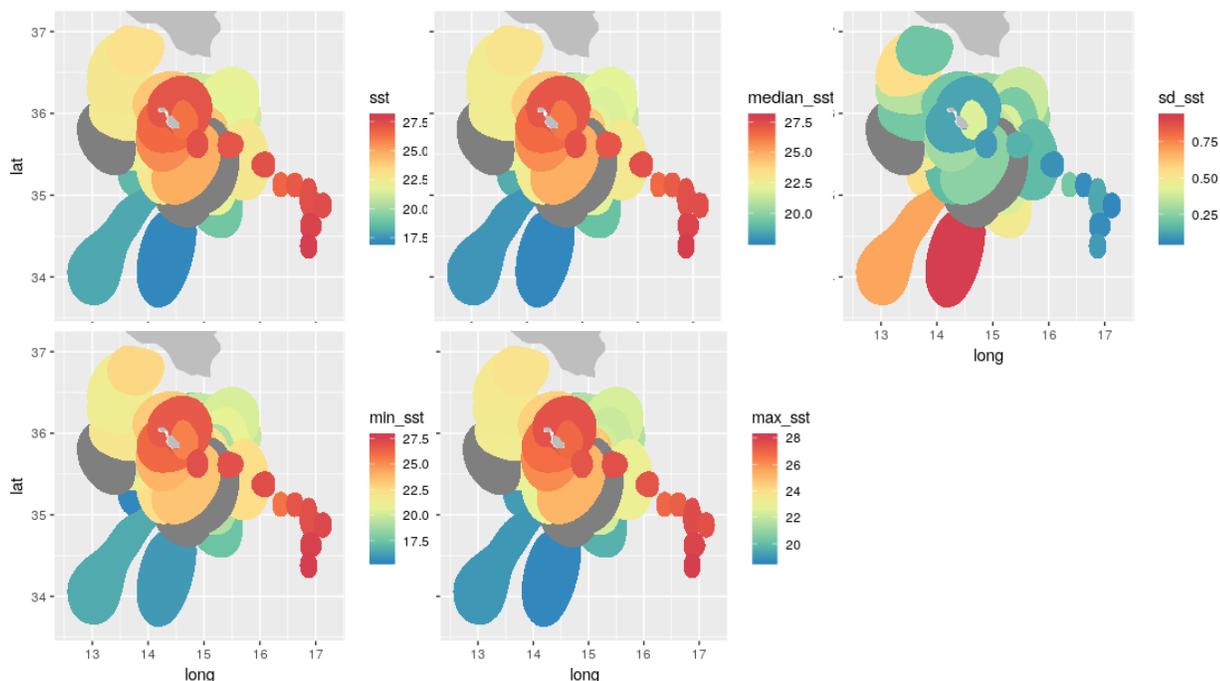
**Figure 2:** Sea water temperature values ( $^{\circ}\text{C}$ ) as recorded by the Wildlife tag (black), as sea surface temperature as enriched in the present work with MODIS-Aqua satellite data.

### Uncertainty Polygons

To enrich the uncertainty polygons, we converted each uncertainty polygon at each uncertainty level, `LATITUDE_UNCERTAINTY_#` and `LONGITUDE_UNCERTAINTY_#` into a shapefile. We identified the environmental data for the appropriate dates and used these to enrich the polygons (see [Figure 3](#) for example of the three levels of enrichment), for each time step and for each environmental variable. The statistics that were calculated for each polygon were the mean, median, minimum, maximum, and standard deviation and 25% and 75% quantiles (e.g., [Figure 4](#)).



**Figure 3:** The uncertainty around one point of a trajectory enriched with the mean of A) sea surface temperature ( $^{\circ}\text{C}$ ) and B) surface chlorophyll a ( $\text{mg m}^{-3}$ ). Outer to inner polygons respectively represent the 95%, 75% and 50% uncertainty envelopes around the likely geolocation as calculated by Wildlife Computers.



**Figure 4:** The 50% likelihood polygons enriched with the A) mean, B) median, C) minimum, D) maximum, and E) standard deviation of sea surface temperature within the area of the polygon.

## NetCDF standards for enrichment variables

### Point enrichment variables

The enriched data were added as enrichment variables to the original NetCDF, continuing to follow the POPSTAR formatting standard and CF standards for global and variable metadata. Enrichment variables for maximum likelihood points are entered with a TIME dimension (here: n=917), may also have a vertical dimension). The TIME variable here indicates the time of the enriched data variable, and the TIME\_DIFF variable indicates the difference in time (days) between the enriched data observation time and the polygon date.

**Table 2** Description of point enrichment variables following POPSTAR NetCDF formatting. More precise descriptions for each enrichment parameter are defined in [Appendix I](#).

Name	Definition	Comment
<ENRICHED_PARAM>	<pre>float &lt;ENRICHED_PARAM&gt;(TIME); &lt;ENRICHED_PARAM&gt;:standard_name = &lt;X&gt;; &lt;ENRICHED_PARAM&gt;:units = &lt;Y&gt;; &lt;ENRICHED_PARAM&gt;:_FillValue = &lt;Y&gt;; &lt;ENRICHED_PARAM&gt;:long_name = &lt;Y&gt;; &lt;ENRICHED_PARAM&gt;:valid_min = &lt;Y&gt;; &lt;ENRICHED_PARAM&gt;:valid_max = &lt;Y&gt;; &lt;ENRICHED_PARAM&gt;:comment = &lt;Y&gt;; &lt;ENRICHED_PARAM&gt;:ancillary_variables = &lt;Y&gt;; &lt;ENRICHED_PARAM&gt;:accuracy = &lt;Y&gt;;</pre>	<p>&lt;ENRICHED_PARAM&gt; contains the values of a enriched parameter, e.g. SST.</p> <p>&lt;ENRICHED_PARAM&gt; names are defined in <a href="#">Appendix I</a>. Examples: SST, CHLA.</p>

	<pre> &lt;ENRICHED_PARAM&gt;:precision = &lt;Y&gt;; &lt;ENRICHED_PARAM&gt;:resolution = &lt;Y&gt;; &lt;ENRICHED_PARAM&gt;:cell_methods = &lt;X&gt;; &lt;ENRICHED_PARAM&gt;:DM_indicator = &lt;X&gt;; &lt;ENRICHED_PARAM&gt;:reference_scale = &lt;Y&gt;; &lt;ENRICHED_PARAM&gt;:coordinates = "JULD LATITUDE LONGITUDE"; &lt;ENRICHED_PARAM&gt;:sdn_parameter_urn = "XXX"; &lt;ENRICHED_PARAM&gt;:sdn_uom_urn = "XXX"; &lt;ENRICHED_PARAM&gt;:sdn_uom_name = "XXX"; </pre>	<p>“units” and “_FillValue” are mandatory. “standard_name” is mandatory if it exists for this variable.</p>
<pre> &lt;ENRICHED_PARAM&gt; _TIME </pre>	<pre> double &lt;ENRICHED_PARAM&gt;_TIME(TIME); &lt;ENRICHED_PARAM&gt;_TIME:long_name = "&lt;ENRICHED_PARAM&gt; enriched for points, time"; &lt;ENRICHED_PARAM&gt;_TIME:standard_name = "time"; &lt;ENRICHED_PARAM&gt;_TIME:units = "days since 1950-01-01T00:00:00Z"; &lt;ENRICHED_PARAM&gt;_TIME:_FillValue = 999999. ; &lt;ENRICHED_PARAM&gt;_TIME:valid_min = 0.0; &lt;ENRICHED_PARAM&gt;_TIME:valid_max = 90000.0; &lt;ENRICHED_PARAM&gt;_TIME:comment = &lt;Y&gt;; &lt;ENRICHED_PARAM&gt;_TIME:sdn_parameter_urn = "SDN:P01::XXX"; &lt;ENRICHED_PARAM&gt;_TIME:sdn_uom_urn = "SDN:P061::UTAA"; </pre>	<p>Date and time (UTC) of the enriched parameter in days since midnight, 1950-01-01.</p> <p>Example: July 25, 2001, 19:14:00 is stored as 18833.8013889885.</p> <p>&lt;Y&gt;: Any optional comment.</p>
<pre> &lt;ENRICHED_PARAM&gt; _TIME_DIFF </pre>	<pre> double &lt;ENRICHED_PARAM&gt;_TIME_DIFF(TIME); &lt;ENRICHED_PARAM&gt;_TIME_DIFF:long_name = "&lt;ENRICHED_PARAM&gt; time difference between enriched data and point data"; &lt;ENRICHED_PARAM&gt;_TIME_DIFF:standard_name = "time_difference"; &lt;ENRICHED_PARAM&gt;_TIME_DIFF:units = "days"; &lt;ENRICHED_PARAM&gt;_TIME_DIFF:_FillValue = 999999. ; &lt;ENRICHED_PARAM&gt;_TIME_DIFF:valid_min = 0.0; &lt;ENRICHED_PARAM&gt;_TIME_DIFF:valid_max = 90000.0; &lt;ENRICHED_PARAM&gt;_TIME_DIFF:comment = &lt;Y&gt;; &lt;ENRICHED_PARAM&gt;_TIME_DIFF:sdn_parameter_urn = "SDN:P01::XXX"; &lt;ENRICHED_PARAM&gt;_TIME_DIFF:sdn_uom_urn = "SDN:P061::UTAA"; </pre>	<p>The difference in time (days) between the enriched data observation time and the point data observation time.</p>

### Uncertainty polygon enrichment variables

Enrichment variables for uncertainty polygons are entered with dimensions, N\_VALUES\_# and TIME\_UNCERTAINTY (here: n=46), and include the mean, standard deviation, minimum, 25% quantile, median, 75% quantile, and maximum of each enrichment variable. BUFFER\_TIME is also included, and indicates the date before or after the date of the polygon that is searched for enrichment data. The data can

be searched before or after the date of the polygon, and can be selected by the user in the Environmental\_Data\_Enrichment step.

**Table 3** Description of uncertainty polygon enrichment variables following POPSTAR NetCDF formatting. “#” indicates the uncertainty level (here: 1=95%, 2=75%, 3=50%), and “X” indicates the summary statistic of the parameter value (here: MEAN, SD, MIN, MAX, Q25, Q75). More precise descriptions for each enrichment parameter are defined in [Appendix I](#).

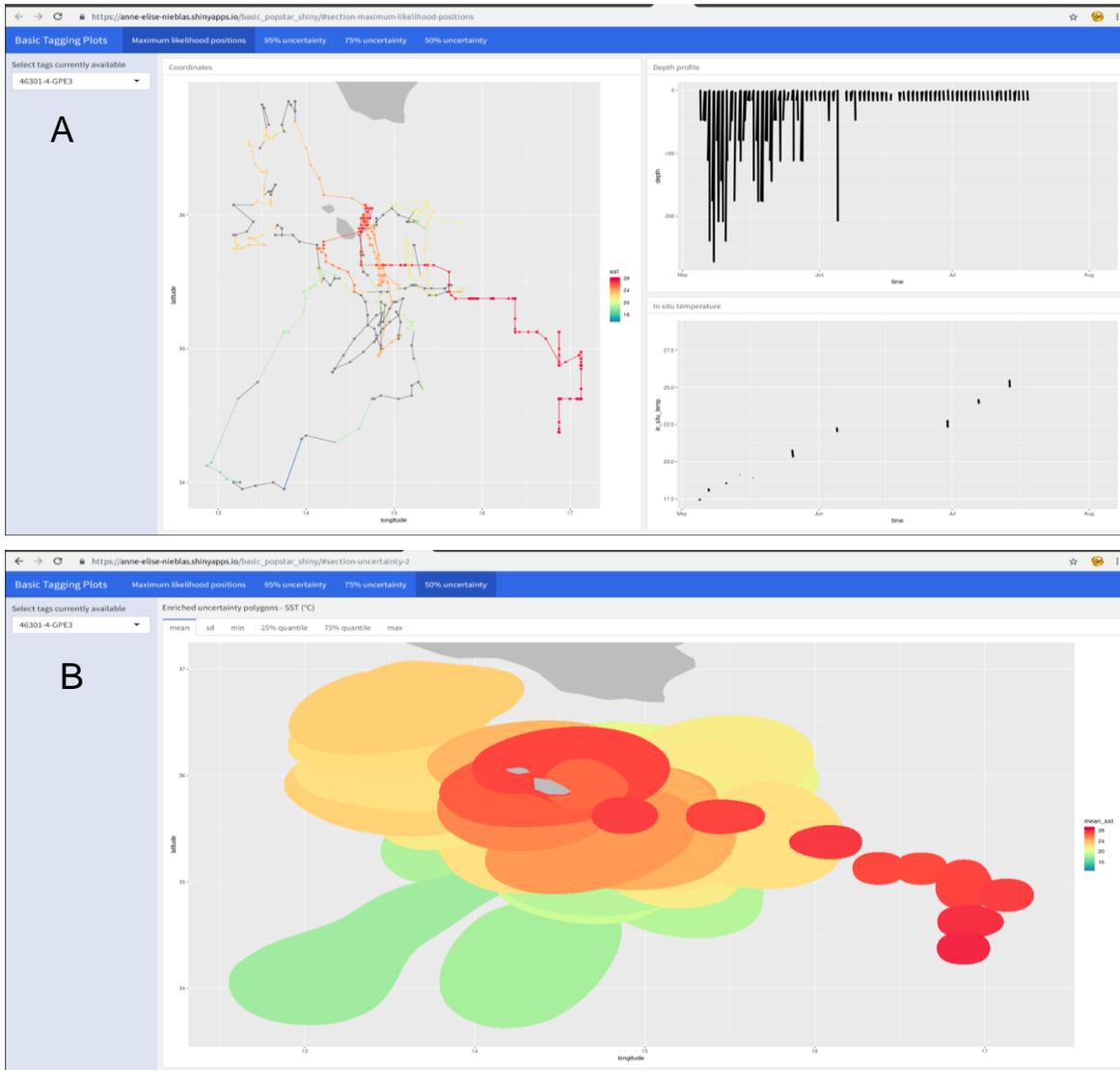
Name	Definition	Comment
<code>&lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_X</code>	<pre>float &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_X (N_VALUES_#,TIME_UNCERTAINTY); &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_X :standard_name = &lt;X&gt;; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_X :units = &lt;Y&gt;; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_X :_FillValue = &lt;Y&gt;; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_X :long_name = &lt;Y&gt;; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_X :valid_min = &lt;Y&gt;; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_X :valid_max = &lt;Y&gt;; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_X :comment = &lt;Y&gt;; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_X:ancillary_variables = &lt;Y&gt;; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_X:accuracy = &lt;Y&gt;; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_X:precision = &lt;Y&gt;; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_X:resolution = &lt;Y&gt;; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_X:cell_methods = &lt;X&gt;; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_X:DM_indicator = &lt;X&gt;; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_X:reference_scale = &lt;Y&gt;; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_X:coordinates = "JULD LATITUDE_UNCERTAINTY_# LONGITUDE_UNCERTAINTY_#"; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_X:sdn_parameter_urn = "XXX"; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_X:sdn_uom_urn = "XXX"; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_X:sdn_uom_name = "XXX";</pre>	<p><code>&lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_X</code> contains the values of the summary statistic (X) of an enriched parameter within the uncertainty polygon (#), e.g. SST_UNCERTAINTY_1_MEAN.</p> <p><code>&lt;ENRICHED_PARAM&gt;</code> names are defined in <a href="#">Appendix I</a>. Examples: SST, CHLA.</p> <p><code>_UNCERTAINTY_#_X</code> are the levels of uncertainty for the uncertainty polygons (#), e.g., 1,2,3 which represent 95%,75%, and 50% uncertainty, respectively, and the summary statistic of the parameter (X), e.g., MEAN, SD, MIN, MAX. See <a href="#">Appendix I</a>.</p> <p>“units” and “_FillValue” are mandatory. “standard_name” is mandatory if it exists for this variable.</p>
<code>&lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_B</code>	<pre>double &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_BUFFER</pre>	<p>Date and time (UTC) of</p>

<p>UFFER_TIME</p>	<pre> _TIME(N_VALUES_#,TIME_UNCERTAINTY); &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_BUFFER _TIME:long_name = "&lt;ENRICHED_PARAM&gt; buffer time to satellite observation, polygon uncertainty level #"; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_BUFFER _TIME:standard_name = "time"; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_BUFFER _TIME:units = "days since 1950-01-01T00:00:00Z"; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_BUFFER _TIME:_FillValue = 999999. ; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_BUFFER _TIME:valid_min = 0.0; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_BUFFER _TIME:valid_max = 90000.0; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_BUFFER _TIME:comment = &lt;Y&gt;; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_BUFFER _TIME:sdn_parameter_urn = "SDN:P01::XXX"; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_BUFFER _TIME:sdn_uom_urn = "SDN:P061::UTAA"; </pre>	<p>the buffer time to remote or model observation, polygon uncertainty level # in days since midnight, 1950-01-01.</p> <p>Example: July 25, 2001, 19:14:00 is stored as 18833.8013889885.</p> <p>&lt;Y&gt;: Any optional comment.</p>
<p>&lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_PCDATA_IN_AREA</p>	<pre> float &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_PCDATA _IN_AREA(N_VALUES_#,TIME_UNCERTAINTY); &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_PCDATA _IN_AREA:long_name = "&lt;ENRICHED_PARAM&gt; percent of data in polygon area, polygon uncertainty level #"; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_PCDATA _IN_AREA:standard_name = "pcdata_in_area"; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_PCDATA _IN_AREA:units = ""; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_PCDATA _IN_AREA:_FillValue = 999999. ; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_PCDATA _IN_AREA:valid_min = 0.0; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_PCDATA _IN_AREA:valid_max = 90000.0; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_PCDATA _IN_AREA:comment = &lt;Y&gt;; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_PCDATA _IN_AREA:sdn_parameter_urn = ""; &lt;ENRICHED_PARAM&gt;_UNCERTAINTY_#_PCDATA _IN_AREA:sdn_uom_urn = ""; </pre>	<p>Percent of &lt;enriched_param&gt; data in polygon, level # area</p>

## Shiny visualisation tool

An R-Shiny visualisation tool ([Basic Tagging Plots](#)) was developed using Rmarkdown to read NetCDF files in the standard POPSTAR format and display basic graphics to aid in quick assessments of the tag outputs, including the map of enriched maximum likelihood points, the depth and temperature profile ([Figure 5a](#)). Summary statistics of the enriched uncertainty polygons of up to three different uncertainty levels can be

visualised, including mean, standard deviation, minimum, maximum, and quantiles at 25% and 75% (e.g., [Figure 5b](#)).



**Figure 5:** The [Basic Tagging Plots](#) shiny using Rmarkdown and hosted on shinyapps.io. A) The first page of the shiny displays the enriched maximum likelihood points (here: enriched with sea surface temperature), the depth profile of the full time series, and the in-situ sea water temperature. B) An example of the uncertainty polygon display (here: 50% uncertainty), enriched for the mean, standard deviation, minimum, 25% quantile, median, 75% quantile, and maximum of the variable of choice (here: sea surface temperature).

## **Acknowledgements**

The Group for High Resolution Sea Surface Temperature (GHR SST) Multi-scale Ultra-high Resolution (MUR) SST data and surface chlorophyll a data were obtained from the NASA EOSDIS Physical Oceanography Distributed Active Archive Center (PO.DAAC) at the Jet Propulsion Laboratory, Pasadena, CA (<http://dx.doi.org/10.5067/GHGMR-4FJ01>).

## **Appendices**

**Appendix I** : [Enriched variable specification table](#). The tables defines the variable attributes to be used when defining enrichment variables in the Enrich2NetCDF.R function.