Pre-processing of sea turtle biologging observations using a clustering algorithm

SynObs Workshop

Aurélien PRAT

LACy (Météo France) & Inria

Nov 2022

STORM-IO Overview

STORM-IO

- Sea Turtles for Ocean Research and Monitoring
- Contribution of **biologging** technology using **sea turtles** equipped with ARGOS transmitters

Collaboration based on la Réunion Island

- **LACy (Météo France)** : Better understanding and modelling of the **tropical cyclones** on the SWIO. Part research, part operational.
- **Kelonia** : **Sea turtle care center** based in la Réunion. Animals caught in fisher nets or hurt by boat are here hosted, healed then released in the ocean $\rightsquigarrow$ 25 turtles/year

STORM-IO Context

South Western Indian Ocean

- **Strong cyclonic activity** from January to May
- Operational forecasting system (AROME-Réunion), 3D-Var DA system
- Very poorly instrumented, need of **in situ observations**

Sea turtles as oceanographic auxiliaries

- Autonomous sampling platforms
- Travel **thousands of km** in several months within the **ocean mixed layers**
- Might be trapped inside **cyclonic structures**

STORM-IO Main objectives

Sea turtle ecology

- Navigation corrected using surface currents
- Diving patterns

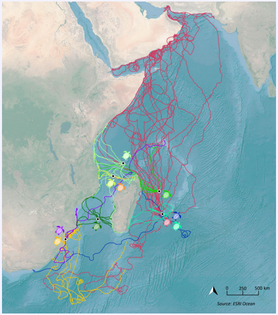
Physical Oceanography & Meteorology

- Sample the **ocean mixed layers** (even deeper layers)
- Verify the **satellite-derived ocean surface products**
- Analyse the AROME-Réunion **forecasting performances**
- **Improve climate forecast** during cyclonic emergence focused on the SWIO

STORM-IO Material

Sea turtles profiles

- **93 sea turtles** equipped : 38 juveniles, 54 nesting females, 1 male
- From **Jan 2019 to Apr 2022**
- **7 initial locations** : Moheli, Aldabra, Seychelles, Réunion Island, Europa, Tromelin, SA
- North global migration for juveniles, 500km/month on average



Tags description

WILDLIFE COMPUTERS		ARGOS Loc.	Temperature sensor (0.05°C res)	Depth sensor (0.5m res)	Internal memory	Perpetual acquisition of temp. series data (5min rate)
Lotek		ARGOS Loc.	Temperature sensor (0.05°C res)	Depth sensor (10m res)	Internal memory	Temp. series profiles (6 points, 5min rate, depth threshold)

Raw data characteristics

ARGOS transmitters

- Transmits **periodic short duration messages** to ARGOS instruments on satellites
- **Polar orbiting satellites** at an altitude of 850km
- Transmission only possible when the turtle is **close to the surface** and a satellite is **passing overhead** (Max. 14 messages per day)

Raw data characteristics

- Location of the ST is computed **during transmissions only**
- Based on **Doppler effect** : received frequency differs from the transmitting one, due to the satellite being mobile

STORM-IO - Oceanography Objectives

Objectives of STORM-IO - Oceanography

- **Elaborate a methodology** allowing the assimilation of ST environmental biologging time series in an operational meteorologic and oceanic forecasting system
- **Produce monitorings** to check data quality, compare to model forecasts, verify satellite-derived ocean surface products
- **Run several reanalysis** experiences for cross-validation

The Variational Approach

- Evaluate the **discrepancy** between the model predicted trajectory and the observation data using a **cost function** :

$$J(\mathbf{x^o}) = \underbrace{\frac{1}{2}(\mathbf{x^o} - \mathbf{x^b})^T \mathbf{P}^{-1}(\mathbf{x^o} - \mathbf{x^b})}_{J^b \text{ : background}} + \underbrace{\frac{1}{2}(H(\mathbf{x^o}) - \mathbf{y})^T \mathbf{R}^{-1}(H(\mathbf{x^o}) - \mathbf{y})}_{J^o \text{ : obs}}$$

  With $\mathbf{x^o}$ the initial state, $\mathbf{x^b}$ the background, **H** the observation operator, **P** the covariance matrix of the background error, **R** the covariance matrix of the observation error.

- Minimize **J** over all the possible $\mathbf{x^o}$ to find the optimal **state** $\mathbf{x^a}$

$$\mathbf{x^a} = \min_{\mathbf{x^o}} J(\mathbf{x^o})$$

Pre-processing of ST biologging observations

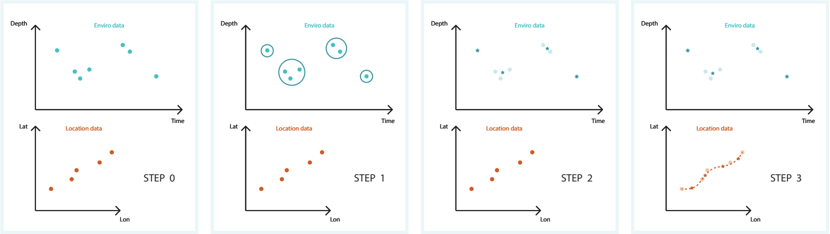
Need to pre-process ST observation time series

- DA requires to compute an *error covariance matrix* **R**, that gathers all the "*inter-dependencies*" of the observations → **Huge matrix** that needs to be inverted !
- Data acquisition **frequency rate** is very high here (every 5 min) → **Large profusion of data**, obs cannot be considered as **non dependant** without pre-processing
- **Two approaches** to reduce the number of information to use as observation data : **data filtering** (pick up a restricted amount of data) or **superobs creation** (average similar data)

Sea turtle observation data pre-processing

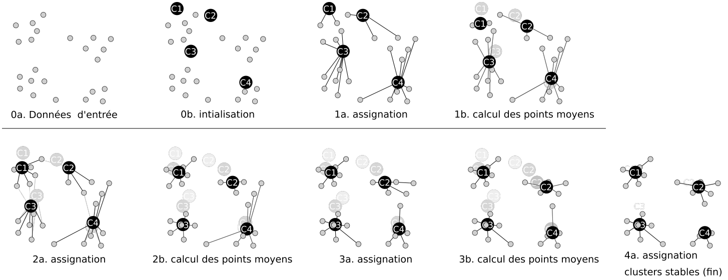
Super-obs creation process

1. **Gather similar environmental data** (temperature and depth) within clusters, using a **k-means method**
2. Compute the **super-obs environmental values and measure errors** for each cluster (arithmetic mean), along with the corresponding date
3. **Assign a location** to each super-obs using the R package *foieGras*
4. Convert the resulting dataset into a **feedback file** (netcdf) to be used as NEMOVar input



K-means clustering (Non-supervised clustering method)

- **Iteration of those steps :**
  1. Assign a cluster to each point, based on the distance to the corresponding cluster node
  2. Compute new nodes as the mean of each cluster
- **Stopping criterion :** Minimizes the intra-class inertia → within-cluster variances (squared Euclidian distances)



Step 1 : K-means clustering

Regarding the super-obs creation

- Parameters : **depth and time only**, as temperature is highly correlated to both of them (turtle trajectory). **Final within-cluster temperature variance** will be used as a verification.
- The final number of clusters should be provided as **input** → Need to define a criterion to determine this value
- Define the final number of clusters **K** such that the final intra-class inertia :

$$\mathbf{ICI} = \frac{1}{N}\sum_{i=1}^{K}\sum_{x\in S_i}\left(||Z_x - Z_i||^2 + ||t_x - t_i||^2\right) < \mathbf{ICI_{max}}$$

With **N** points dispatched into **K** clusters, $\mathbf{Z_i}$ and $\mathbf{t_i}$ being the cluster node depth and time of the $\mathbf{S_i}$ cluster.

Step 1 : K-means clustering

Define $\mathbf{ICI_{max}}$

- We chose $\boxed{\mathbf{ICI_{max}} = \Delta Z_{max}^2 + \Delta t_{max}^2}$
- ΔZ_{max} equals to the size of a **NEMO grid cell** along the Z-axis
- Δt_{max} is determined using the **characteristic autodecorrelation time** of the depth time series $\rightarrow$ The goal is to end up with **non-correlated super-obs**.

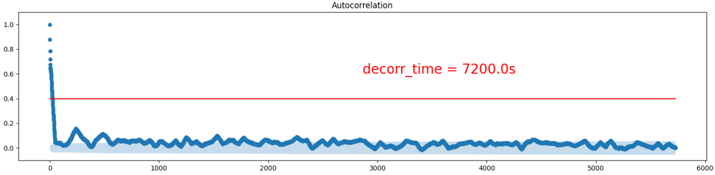


Figure: Example of an autocorrelation graph for a loggerhead turtle (26000 points), with an autocorrelation coefficient set to 0.4

Step 3 : Assign a location to each super-obs

R package *foieGras*

- Animal movement is modelled as a **continuous-time random walk on velocity $\mathbf{v_t}$** in two coordinate axes :
  - $v_t = v_{t-\Delta} + \Sigma_\Delta$ where Δ is the time increment and Σ_Δ is a zero-mean, bi-variate Gaussian random variable with variance $2D\Delta$
  - $x_i = x_{i-1} + v_i \Delta_i$ where x_i is the true location of the animal at time t_i
  - $y_i = x_i + \epsilon_i$, $\epsilon_i \sim N(0, \Omega_i)$ where y_i the location observed at time t_i, Ω_i the measurement error-covariance matrix with elements being derived from the ARGOS error ellipses components
- **Fit the state-space model**, using maximum likelihood to estimate model parameter **D** (*Jonsen et al., 2020* and *Kristensen et al., 2016*)
- Find the **predicted states** corresponding to the super-obs dates using the evaluated model

Obs pre-processing output example

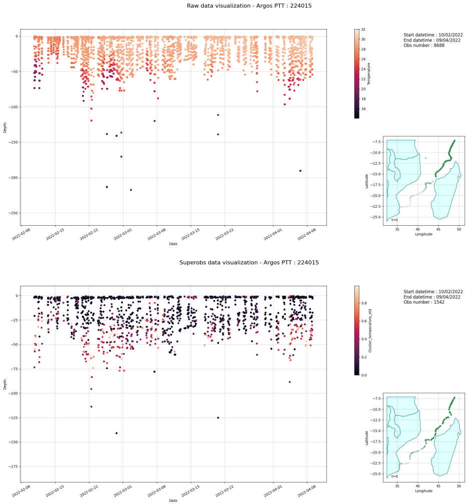


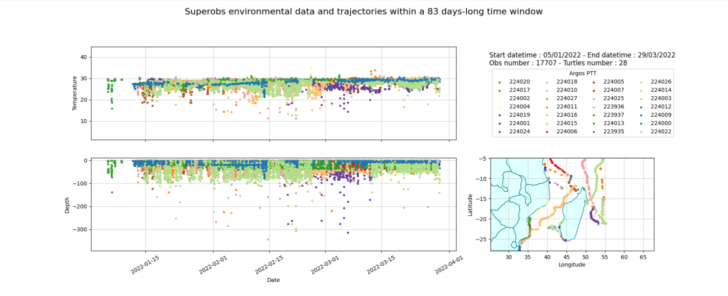
Figure: **Example of obs. pre-processing output** : Loggerhead turtle released from South Africa, equipped with *Wildlife Computers* tag. **Final reduction factor : 78%**

Ocean model monitoring : first results

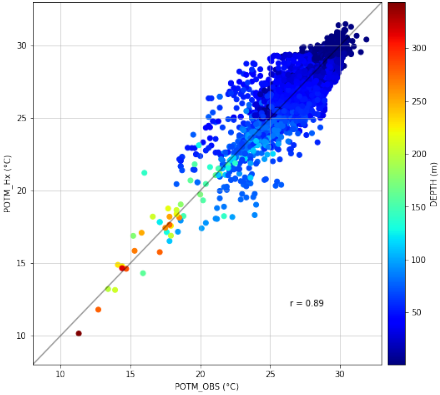
Ocean model characteristics

Model	Config	Spatial res.	Z-levels	Time step	RST	OBC	Forcing
NEMO	REUNION12	1/12°	50	360s	PSY4	PSY4	• IFS (9km res.), forcing rate : 3h • AROME (2.5km res.), forcing rate : 1h

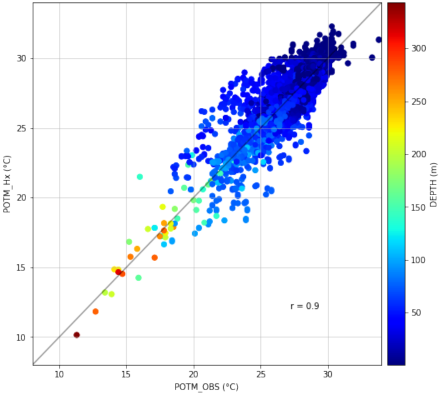
Monitoring period selection



Ocean model monitoring : first results



(a) IFS forcing



(b) AROME coupling

Figure: Obs and forecast temp. correlation graphs for a 3 months-long monitoring, starting from a single PSY4 reanalysis. Only *Wildlife Computers* obs. were processed.

Future developments

Monitoring

- Compare AROME forcing, IFS forcing and AROME coupled forecast temp. to the observed ones.
- Run the same monitorings using the LOTEK datasets.

Reanalysis

- Run a set of **3 different reanalysis** using the NEMOVar (3DFGAT) system : without obs., with conventional obs., with conventional obs. + STORM obs.

Conclusion

- The first results reveal a promise of **quantifying the ocean mixed layers** inside **cyclonic eddies**, along with the use of additional obs. in the NEMO reanalysis.