International Congress on Environmental Modelling and Software

Scalable Big Data Platform, Mining and Analytics Services for Optimized Forecast of Animals Habitats

Zoheir Sabeur Dr

Gianluca Correndo Dr

Fabien Castel Mr

Geoffrey Neumann Dr

Galina Veres Dr

See next page for additional authors

Presenter/Author Information
Zoheir Sabeur Dr, Gianluca Correndo Dr, Fabien Castel Mr, Geoffrey Neumann Dr, Galina Veres Dr, and Banafshe Arbab-Zavar Dr

iEMSs International Congress on Environmental Modelling and Software

9th International Congress on Environmental Modelling and Software
Fort Collins, Colorado, USA, Mazdak Arabi, Olaf David, Jack Carlson, Daniel P. Ames (Eds.)
https://scholarsarchive.byu.edu/iemssconference/2018/

Scalable Big Data Platform, Mining and Analytics Services for Optimized Forecast of Animals Habitats

Zoheir Sabeur[a], Gianluca Correndo[a], Fabien Castel[b], Geoffrey Neumann[a], Galina Veres[a], Banafshe Arbab-Zavar[a]
[a] *University of Southampton, IT Innovation Centre, University of Southampton, Gamma House, Enterprise Road, Southampton, SO16 7NS, UK ([zas, gc, gkn, gvv, baz]@it-innovation.soton.ac.uk)*
[b] *ATOS, Les Espaces St Martin, 6 Impasse Alice Guy, 31300 Toulouse, France (fabien.castel@atos.net)*

Abstract: The effects of climate change have been observed for decades now that we can access to multiple methods of Earth Observation (EO) using in situ, air-borne and space-borne sensing. The generated EO Big Data from these sources is of paramount importance for scientists to understand the effects of climate change and the specific engendered natural (and anthropogenic) processes that are likely to trigger the changing behaviour of species on Earth. In the EO4wildlife project (http://www.copernicus.eu/projects/eo4wildlife), we have access to Copernicus and Argos EO Big Data for investigating the changes of habitats for a variety of marine species. The challenge is to forecast the habitats by identifying the causal relationships between animal presence and Metocean environmental fronts. This is achieved by processing data of animal presence, which are relatively small in size and sparse, and their correlation with environmental datasets, which are large and dense in feature space. This poses big data challenges in terms of optimisation of resources, mining and feature selections. Once overcome, it improves the performance of the forecasting models. The availability of big geospatial information, satellite data and in situ observations enabled us experiment on the scalability of our distributed data storage technologies and analytics services in the cloud. We specifically deployed cluster infrastructure via Spark for a resilient distribution of processing over multiple nodes. The testbed experiments of our big data processing performance are validated under three types of selected habitat forecasting workflows.

Keywords: Big Data, Data Analytics, Features Selection.

1 INTRODUCTION

EO4wildlife brings large number of multidisciplinary scientists such as marine biologists, ecologists and ornithologists around the world to collaborate closely together while using European Sentinel Copernicus Earth Observations more efficiently. In order to reach such important capability, an open service platform and interoperable toolbox is being designed and implemented.
The platform, accessible at http://eo4wildlife.default.eo.sparkindata.com, offers high level data processing services that can be accessed by scientists to perform their respective research. The platform offers dedicated services that will enable scientists' process their geospatial environmental stimulations using Sentinel Earth Observation data and other observation sources. Specifically, the EO4wildlife platform will enable the integration of Sentinel data, ARGOS archive databases and real time thematic databank portals, including Wildlifetracking.org, Seabirdtracking.org, and other Earth Observation and MetOcean databases; locally or remotely, but simultaneously. EO4wildlife research specialises in the intelligent big data processing, advanced analytics and a Knowledge Base for wildlife migratory behaviour and trends forecasting (Sabeur et al, 2017a, b). The research is leading to the development of web-enabled open services using OGC standards for sensor Observation and Measurements and data processing of heterogeneous geospatial observation data with estimated

uncertainties. EO4wildlife designs, implements and validates various scenarios based on real operational use case requirements in the field of marine wildlife migrations, habitats and behaviour.

2 SYSTEM ARCHITECTURE

2.1 Software Architecture

The EO4wildlife system is hosted in a **SparkInData** platform, which offers a set of core services for data discovery, data ingestion, process integration and execution. The **SparkInData** Platform, also known as Smart Elastic Enriched Earth Data (SEEED), is a generic platform which provides an EO data dedicated Cloud platform, infrastructure and services. Furthermore, the platform is organized under three functional zones, as shown in **Figure 1** below. These include: 1) Storage zone for mutualized storage capabilities; 2) Compute zone for mutualized intensive computing; and 3) Service zone for processing services.

Furthermore, the platform infrastructure services are provided by the Big Data Helix Nebula platform. Slipstream is used at a Platform as a Service (PaaS) level. The PaaS is provided under a cloud computing services environment in order to enable developers run, test and manage their own applications while processing Big EO Data while performing analytics for extracting trends on marine species migratory routes with respect to Ocean fronts geospatial and temporal trends. Specifically, PaaS is based on Google's Kubernetes (K8S, see (Bernstein, 2014)) open software with an augmented dedicated **SparkInData** Service Management Layer. The latter is responsible for deploying applications on SaaS mode and managing them with auto-scaling, load balancing, monitoring or decommissioning upon request by application owners.

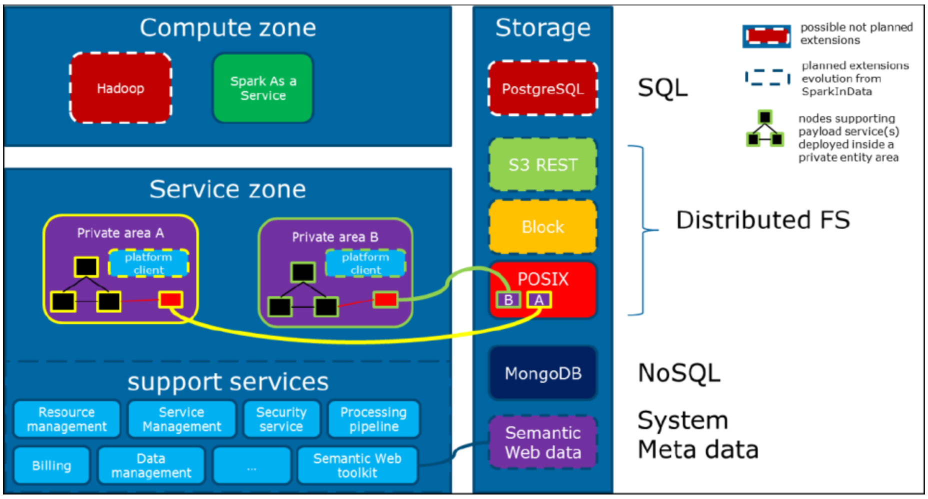


Figure 1. Global Big Data Platform Architecture Overview

In addition to the above, the **SparkInData** platform components consist of the following:

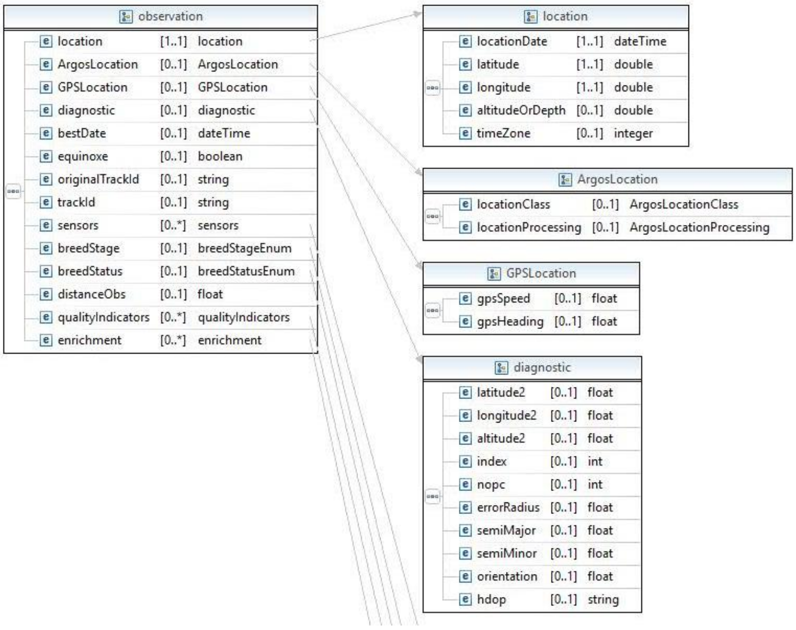
- **Security Service**: It registers users, their roles and rights to ensure their authentication and access control to the platform
- **Processing Pipeline**: It controls and monitors the chaining of the data analytics Web Processing Services (WPS)
- **Data management service**: It ensures the "import" into the system of various data sets which are generated by data providers or new analytics into the system
- **Service management**: This enables the creation and control of applications deploy-ment and their scalability of operations on the platform
- **Resource Management**: It controls resources deployment, their scalability and opera-tions on the platform
- **Data Storage service**: This service ensures and secures persistent data storage into the system.
- **Semantic Web Toolkit**: It ensures linked data storage, access to RDF resources and mechanisms which define access control policies for graph stores
- **Market Place**: It provides a common place for information exchange for publishing application services outputs.

- **Billing service**: It controls how application services can be purchased from the Mar-ket Place where payment can be performed
- **Spark as a Service**: It executes and controls Spark Jobs through a dedicated web service

2.2 Data model

In EO4wildlife, the focus is on the study of animals' behaviour, their preferred habitats and the modelling of such habitats in term of environmental conditions. In order to integrate tracking data sets from different data providers, we established, following the input from the scientific community and the advisory board, a common schema (described as an XML schema document or XSD) for the tracking data. All the services that implement niche modelling workflows (as described in Section 3), involve either processing these tracking data sets or the integration of these data sets with raster time series containing environmental observations. The nature of the information contained in the tracking data sets dictate what type of parallelization we can achieve.

An EO4wildlife XML track file can contain one or more dataset and each dataset has a reference colony which is geo-localised. Moreover, a dataset can contain a collection of animals and for each animal we can have a collection of devices providing observations on the animal. The most important type of information in the context of the habitat modelling addressed by this paper, is the location of the animal. Each location is a point in the geographical space, observed either via ARGOS satellite or GPS, with a time stamp associated to it. The observed position is then supplemented with a position quality attribute ("Argos User's Manual," 2016) and additionally with animals' specific observations and enrichments, which is the way in which the sampled values at the locations are encoded in the track files (see **Figure 2**).



observation		
location	[1..1]	location
ArgosLocation	[0..1]	ArgosLocation
GPSLocation	[0..1]	GPSLocation
diagnostic	[0..1]	diagnostic
bestDate	[0..1]	dateTime
equinoxe	[0..1]	boolean
originalTrackId	[0..1]	string
trackId	[0..1]	string
sensors	[0..*]	sensors
breedStage	[0..1]	breedStageEnum
breedStatus	[0..1]	breedStatusEnum
distanceObs	[0..1]	float
qualityIndicators	[0..*]	qualityIndicators
enrichment	[0..*]	enrichment

location		
locationDate	[1..1]	dateTime
latitude	[1..1]	double
longitude	[1..1]	double
altitudeOrDepth	[0..1]	double
timeZone	[0..1]	integer

ArgosLocation		
locationClass	[0..1]	ArgosLocationClass
locationProcessing	[0..1]	ArgosLocationProcessing

GPSLocation		
gpsSpeed	[0..1]	float
gpsHeading	[0..1]	float

diagnostic		
latitude2	[0..1]	float
longitude2	[0..1]	float
altitude2	[0..1]	float
index	[0..1]	int
nopc	[0..1]	int
errorRadius	[0..1]	float
semiMajor	[0..1]	float
semiMinor	[0..1]	float
orientation	[0..1]	float
hdop	[0..1]	string

Figure 2. XSD Snippet for the **observation** tag

When the information encoded in XML is imported in the system, a data frame is created and the data model is translated from a tree-like structure, to a table-like structure where some of the information is made redundant in order to have a homogeneous data model which can later be distributed over the nodes of the Spark cluster.

3 HABITAT MODELING AND FORECASTING

In this section we will present the habitat modelling workflows developed within the framework of the EO4wildlife project. Each workflow has been developed in function of available literature on the subject and the availability and type of the tracking data available.

3.1 Atlantic Bluefin Tuna Scenario

Atlantic Bluefin tuna (ABFT henceforth) is a highly migratory species which is able to tolerate a wide ranges of environmental conditions (Arrizabalaga et al., 2015) in the Atlantic Ocean and Mediterranean Sea. In this section, we present initial attempts to correlate ABFT tracking data and environmental variables to identify different pattern of ABFT behavior using Ecological Niche Modelling. The methodological steps undertaken to identify ABFT habitat preferences during different types of behavior are inspired by (Druon et al., 2016) and depicted in **Figure 3.**

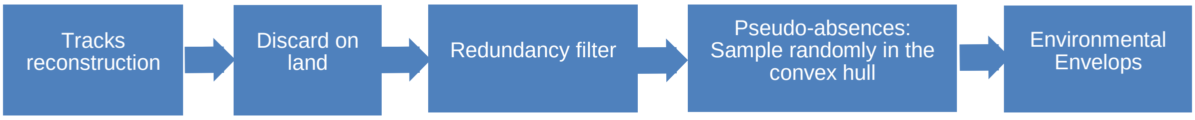


Figure 3. Workflow of Atlantic Bluefin tuna application

The data pre-processing steps included: animal tracks reconstruction, via Track&Lock algorithm by (Royer et al., 2005), discarding the location on the land, a redundancy filtering which removed duplications in the tracks and any relocation points on the same day separated by the less than 2.3 km., smoothing filters to recover spatial missing data (calculating 3-7 days composites for variables with temporal missing data), and calculating gradients for the used environmental variables.

The following environmental variables were identified influencing habitat utilization and preference: daily sea surface temperatures (SST) and chlorophyll a concentration (CHL) which can be used to calculate respective gradients and fronts; bathymetry; CO_2 net primary production (PP); daily sea surface height anomaly (SSH a), eddies, and surface wind speed. Additionally, several previous research papers (such as (Druon et al., 2016) and previous) reported the specific environmental conditions which favoured by ABFT for spawning and feeding. For spawning habitat Bluefin tuna prefers warming water of Mediterranean Sea (SST in the range 20 to 25.5ºC) with high increase in SST over several weeks, relatively low levels of CHL, intermediate levels of Eddy Kinetic Energy (EKE) and preferable range for SSH a. While for feeding habitat ABFT prefer to locate in the vicinity of chlorophyll a frontal features and higher levels of concentrations, wide range of SST and immediate levels of PP. These analysis shows that it should be possible to distinguish between feeding and spawning behaviors of ABTF when environmental variables are added to modelling.

The niche modeling used in this workflow identified environmental envelops in which the animals were observed (Walker and Cocks, 1991). Once the threshold values for environmental variables are set, the specific ecological niche of ABFT is defined for feeding and spawning. Ecological niche model using environmental envelopes are then used to predict the daily suitability of cells within habitat for ABFT feeding and spawning on a scale 0-1. The favorable habitat for each behavior are cells that meet all the suitable ranges of selected variables. This will be the next step of our services development.

3.2 Marine Turtle Scenario

The marine turtle scenario implemented in EO4wildlife is based on the workflow described by (Pikesley et al., 2015). In this workflow 21 female Olive ridley turtles are tracked between 2007 and 2010 in the southeast Atlantic ocean near the west coast of Africa (Pikesley et al., 2013). The aim here is to describe the observed and potential post-nesting habitat for this species in this region. Considering the fisheries catch data for this region, areas which are potential for conflict where fishing activities may result in bycatch of this turtle species are identified. Similar approach is taken in (Pikesley et al., 2015), where 32 adult loggerhead turtles are tracked in the eastern Atlantic. This workflow investigated how the predicted habitat may alter under the influence of climate change. **Figure 4** depicts the workflow that has been implemented.

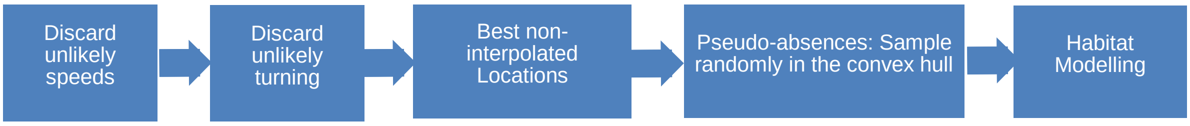

Discard unlikely speeds
Discard unlikely turning
Best non-interpolated Locations
Pseudo-absences: Sample randomly in the convex hull
Habitat Modelling

Figure 4. Marine Turtle workflow

Other approaches describing the marine turtles' behaviour have also been considered and may be part of future work.

Twenty five tracks of adult loggerhead sea turtles are being evaluated during their post-nesting movements near the west coast of Africa. This data includes samples from Aug 2004 to Dec 2009, where in some of the turtles are tracked for a short time and some for longer periods. The considered environmental variables are: Sea Surface Temperature, Bathymetry, Sea surface Height (Absolute Dynamic Topography), Net Primary Production, Current Velocity, and Eddies. These environmental variables were chosen based on previous reported correlations between these variables and marine turtle relocations (Chambault et al., 2016; Pikesley et al., 2013, 2015).

The data pre-processing steps included: discarding locations which produced unlikely (for the marine species) speeds or turning points, a filter which selected the best non-interpolated relocations for each animal, the generation of pseudo-absences from the convex hull. The niche models implemented in this workflow were: GAM (Guisan et al., 2002), GLM (Guisan et al., 2002), Random Forests (Thuiller et al., 2009), Booster Regression Trees (Leathwick et al., 2006), Classification Trees (De'ath and Fabricius, 2000), and Max Entropy (Phillips and Dudík, 2008).

3.3 Data Pre-processing, Aggregation, and Mining

The workflows described above make use of a number of different for the pre-processing, cleaning and aggregation of the data prior to the analytical step.

The **pre-processing** of geospatial data sets is an important step when dealing with potentially imprecise information such as animal positions. These services allow to recognize and eliminate all data elements which are clearly unrealistic given the knowledge of the domain (e.g. the animal is not capable of travelling at such velocities, or of producing such abrupt changes of directions or again to travel inland). Moreover the pre-processing services allow to fill missing data values and accommodate different data grids by interpolating values which were not directly collected and represented in the data sets. Within this category the following services have been implemented and deployed in the platform:

- **Speed filter**: to filter out locations that would give unrealistic speeds
- **Position quality filter**: to filter out location with low ARGOS quality
- **Turning angle filter**: to filter out locations that would give unrealistic trajectories
- **Filter locations in land**: to filter out locations which are in land (unrealistic for some marine species)

Aggregation services allow to reconcile data represented with different spatial or temporal resolutions providing functionalities to sample environmental observations and aggregate them in the right granularity to fuel niche modelling algorithms. In this category are also included services to process animal tracks to provide grouping of tracks in trips or gridding a number of tracks to study the population distribution. This category contains services to process animal tracks and satellite marine observations in order to model animals' use of space and correlate that with available environmental observations. This category is further subdivided in two sub-categories of services: anima track based services and statistical environmental services. Animal tracks based services analyses the tracks alone in order to estimate the animals' home range and the foraging grounds whereas statistical environmental services assess the statistical relevance of environmental observations in modelling animals' presence. Within this category the following services have been implemented and deployed in the platform:

- **Track sampling**: samples multiple raster time series using track locations as sampling points
- **Kernel Density Estimator service**: analyses the animals' tracks to compute the areas utilized by them
- **Metocean statistic service**: provides statistical summaries of multiple raster time series over observed presences over a time period using track locations as input.
- **Pseudo absences service**: generate animals pseudo absences over a tracking dataset convex polygon

Data Mining and Fusion services allows to study animals' habitats on the feature space by using classic classification or regression algorithms. Within this category the following services have been implemented and deployed in the platform:

- **Environmental Envelops**: models the animals' habitat by computing the environmental envelops of the sampled observations.
- **Generalized Linear Models**: models the animals' habitat by applying GLM algorithms.
- **Generalized Additive Models**: models the animals' habitat by applying GAM algorithms.

4 DISTRIBUTED COMPUTING VIA SPARK

Apache Spark is a cloud oriented computing engine which relies on a cluster infrastructure and a distributed file system (usually Hadoop Distributed File System) to distribute data and computation over many nodes. The distributed computing primitives are then offered to the users via a set of APIs available in different languages: Scala, Java, Python, and lately R. The R API to access Spark clusters is named SparkR.

4.1 SparkR API

Within SparkR, alongside the primitive operations to access distributed data frames, there is also a number of Machine Learning algorithms which makes use of Spark's distributed processing features. In particular, SparkR v2.3.0 offers the following ML algorithms which have been used in at least one scenario:

- **spark.glm**: Generalized Linear Model (GLM)
- **spark.gbt**: Gradient Boosted Trees for Regression and Classification
- **spark.randomForest**: Random Forest for Regression and Classification

Moreover, in order to distribute computation over workers, SparkR provides also an API to map a function over a collection:

- **spark.lapply(list, func)**

4.2 Distribution of Computation via Spark

There are many ways in which Spark can be used to support the services described in Section 3.3 and provide therefore a distributed computing environment in this context. By using the Spark API we can either use the Spark data frame APIs to apply filters or compute aggregate functions over the rows and columns on the distributed data frame, or split the initial data frame into smaller parts and then use the **spark.lapply** function to distribute the computation over multiple nodes, or finally using off-the-shelf implementations of ML algorithms.

The above mentioned services provide different constraints in term of data involved in the computation and can therefore be distributed using Spark in different ways. In particular:

- **Speed filter** and **Turning angle filter**: this service computes speeds relevant single animals by using multiple locations to compute average velocities and turning angles at each point of the track. This means that the track data frame cannot be used as it is in a distributed data frame for filtering via Spark. However, the algorithm requires access only to a single animal track at a time and therefore its execution can be split into separate executions per animal using the **spark.lapply** function.
- **Position quality filter**: this service can be trivially implemented as a filter over a column value and can be directly implemented using the Spark data frame APIs.
- **Filter locations in land**: this filter can be trivially distributed using the spark data frame APIs but it requires to have a copy of the world's coastline in each worker node.
- **Track sampling** and **Metocean statistic service**: the distribution over separate worker nodes of large quantities of environmental observations is a serious issue for its distribution over a cluster and it is still under study.
- **Kernel Density Estimator service**: an implementation of a Kernel Density Estimator algorithm has been implemented in geotrellis.io, see (Kini and Emanuele, 2014), and it is under consideration.
- **Pseudo absences service**: this service can be distributed over a Spark cluster provided that the convex hull polygon is computed beforehand.
- **Environmental Envelops**: once the environmental observations have been sampled, the aggregate statistical summaries (i.e. min, max, median, average) can be computed using the Spark data frame APIs.
- **Generalized Linear Models**: Spark ML API has an implementation for GLM which is under study.
- **Generalized Additive Models**: the implementation of a Spark enabled version of GAM is under study.

4.3 Benchmark of the Spark Speed Filter

For testing the performances of Spark-enabled service implementations in supporting the habitat modelling workflows, we implemented a spark version of the speed filter, by splitting the track dataset into sub-data frames; one per each animal; and then running the classic speed filter implementation on the sub-data frame on the worker node.

The benchmark run used Spark worker nodes deployed as pods in the Kubernetes cluster, with no machines dedicated to Spark. The 2 Kubernetes nodes were the Spark pods are run are 4 CPU machines with 8 GB memory. As the machines are multicore, the tasks are still run concurrently as different pods running on the same machine allow using simultaneously all the CPU.

To test the running time of the service, an initial track file with observed locations was processed to generate further synthetic data files with an increasing number of locations. The running times of the service can be seen in **Figure 5**. The classic implementation has better running times than the Spark-enabled service, until the number of locations exhausts the memory capacities of the single Docker container in which the service runs, whereas the Spark-enabled service can process five times more locations without exhausting the cluster's resources showing better robustness than the classic implementation.

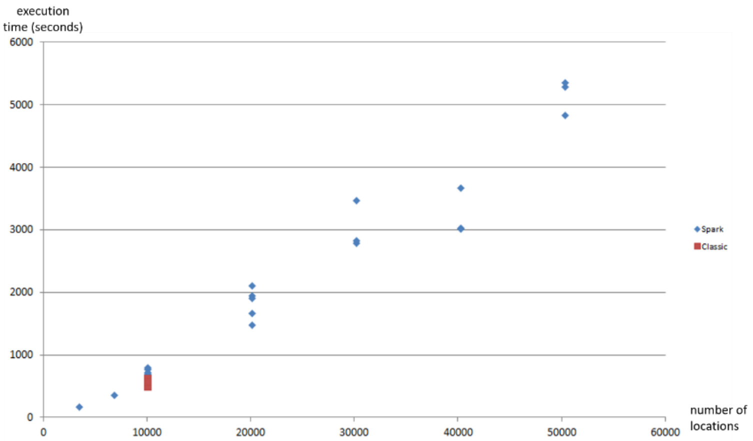


Figure 5. Running time of spark enabled speed filter compared to non-spark implementation.

5 CONCLUSIONS

We have successfully set up an early experiment for testing the scalability of our EO4wildlife big data analytics and mining services. Nevertheless, we have yet to increase the number of processing resources to study how well our implementation scale with larger services loads and clusters in place. This is subject of our ongoing experiments currently for testing the full scalability of our big data services usage by an ever increasing user community communities in need to perform various scenarios of species habitats forecasting under a changing global climate.

ACKNOWLEDGMENTS

This work is partly funded by the European Commission under H2020 Grant Agreement number: 687275. Also, the Authors would specifically like to thank the European Commission for providing access to COPERNICUS data, under the EO4life project, without which this study could not have been conducted.

REFERENCES

Argos User's Manual [WWW Document], 2016. URL http://www.argos-system.org/manual/ (accessed 11.4.16).

Arrizabalaga, H., Dufour, F., Kell, L., Merino, G., Ibaibarriaga, L., Chust, G., Irigoien, X., Santiago, J., Murua, H., Fraile, I., others, 2015. Global habitat preferences of commercially valuable tuna. Deep Sea Research Part II: Topical Studies in Oceanography 113, 102–112.

Bernstein, D., 2014. Containers and cloud: From lxc to docker to kubernetes. IEEE Cloud Computing 1, 81–84.

Chambault, P., De Thoisy, B., Heerah, K., Conchon, A., Barrioz, S., Dos Reis, V., Berzins, R., Kelle, L., Picard, B., Roquet, F., others, 2016. The influence of oceanographic features on the foraging behavior of the olive ridley sea turtle Lepidochelys olivacea along the Guiana coast. Progress in Oceanography 142, 58–71.

De'ath, G., Fabricius, K.E., 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81, 3178–3192.

Druon, J.-N., Fromentin, J.-M., Hanke, A.R., Arrizabalaga, H., Damalas, D., Tičina, V., Quílez-Badia, G., Ramirez, K., Arregui, I., Tserpes, G., others, 2016. Habitat suitability of the Atlantic bluefin tuna by size class: An ecological niche approach. Progress in Oceanography 142, 30–46.

Guisan, A., Edwards Jr, T.C., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological modelling 157, 89–100.

Kini, A., Emanuele, R., 2014. Geotrellis: Adding geospatial capabilities to spark. Spark Summit.

Leathwick, J., Elith, J., Francis, M., Hastie, T., Taylor, P., 2006. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. Marine Ecology Progress Series 321, 267–281.

Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31, 161–175.

Pikesley, S.K., Broderick, A.C., Cejudo, D., Coyne, M.S., Godfrey, M.H., Godley, B.J., Lopez, P., López-Jurado, L.F., Elsy Merino, S., Varo-Cruz, N., others, 2015. Modelling the niche for a marine vertebrate: a case study incorporating behavioural plasticity, proximate threats and climate change. Ecography 38, 803–812.

Pikesley, S.K., Maxwell, S.M., Pendoley, K., Costa, D.P., Coyne, M.S., Formia, A., Godley, B.J., Klein, W., Makanga-Bahouna, J., Maruca, S., others, 2013. On the front line: integrated habitat mapping for olive ridley sea turtles in the southeast Atlantic. Diversity and Distributions 19, 1518–1530.

Royer, F., Fromentin, J.-M., Gaspar, P., 2005. A state–space model to derive bluefin tuna movement and habitat from archival tags. Oikos 109, 473–484. https://doi.org/10.1111/j.0030-1299.2005.13777.x

Sabeur, Z., Correndo, G., Veres, G., Arbab-Zavar, B., Neumann, G., Ivall, T. D., Castel, F, Zigna, J M and Lorenzo, J., (2017a). EO big data analytics for the discovery of new trends of marine species habitats in a changing global climate. Proceedings of the International conference on Big Data from Space (BIDS' 2017) – European Space Agency, 28th-30th November 2017 Toulouse (France). Soille, Pierre and Marchetti, Pier Giorgio (eds.) *European Union Publications 2017.* doi:10.2760/383579.

Sabeur, Z.A., Correndo, G., Veres, G., Arbab-Zavar, B., Lorenzo, J., Habib, T., Haugommard, A., Martin, F., Zigna, J.-M. and Weller, G. (2017b). EO Big Data connectors and analytics for understanding the effects of climate change on migratory trends of marine wildlife. Proceeding of the 12th *International Symposium on Environmental Software Systems*, Zadar, Croatia. 10 - 12 May 2017. 9 pp. IFIP Advances in Information and Communication Technology, Springer Publishing. (PURE UUID: eeb34023-e491-46ea-b380-b71b81905051)

Thuiller, W., Lafourcade, B., Engler, R., Araújo, M.B., 2009. BIOMOD–a platform for ensemble forecasting of species distributions. Ecography 32, 369–373.

Walker, P., Cocks, K., 1991. HABITAT: a procedure for modelling a disjoint environmental envelope for a plant or animal species. Global Ecology and Biogeography Letters 108–118.