

# Ángel Luis Robles Fernández

*Graduate student*

*M.Sc. in Physics*

*Universidad Veracruzana, Mexico*

*Personal data*

*Address: Juarez 83 interior 10 Xalapa Veracruz Mexico*

*Address: 1100 E Lemon St C112, 85281 Tempe AZ USA*

*Phone number: +52 2281318946*

*+1 4807914846*

*e-mail address: a.l.robles.fernandez@gmail.com*

## Education

2021 - Actual **Evolutionary Biology PhD Student**, Arizona State University, School of Life Science

2020 - 2021 **Master of Science in Physics**, Universidad Veracruzana, Faculty of Physics

**Degree in Physics**, Universidad Veracruzana, Faculty of Physics 2017

**Degree in Music**, Universidad Veracruzana, Faculty of Arts 2011

## Honors and awards

05/2023 Grant-Aid-Research American Society of Mammalogists. \$1750.00

10/2022 **2022 Ebbe Nielsen Challenge**. GBIF LACS: GBIF Literature Abstract Classification System. Second Prize (<https://gbif.lacs.info/>). \$5,000.00

03/2022 2022 **Leo S. Rowe Fund Program Beneficiary**. \$7,500.00

10/2020 **2020 Young Research award**. Global Biodiversity Information Facility. \$5,000.00

05/2016 **5th Young Talent award 2016**. Consejo Veracruzano de Investigación Científica y Desarrollo Tecnológico (COVEICYDET). \$1,500.00

11/2016 **Encouragement to Academic and Artistic Recognition**. Universidad Veracruzana. \$250

Total: \$13,500.00

## Peer-reviewed publications

**Robles-Fernández, Á. L.,** Santiago-Alarcon, D., & Lira-Noriega, A. (2022). Wildlife susceptibility to infectious diseases at global scales. *Proceedings of the National Academy of Sciences*, 119(35), e2122851119.

**Robles-Fernández, Á. L.,** Santiago-Alarcon, D., & Lira-Noriega, A. (2021). American Mammals Susceptibility to Dengue According to Geographical, Environmental, and Phylogenetic Distances. *Frontiers in Veterinary Science*, 8

**Robles-Fernández Á. L.** and Lira-Noriega A (2017) Combining Phylogenetic and Occurrence Information for Risk Assessment of Pest and Pathogen Interactions with Host Plants. *Front. Appl. Math. Stat.* 3:17. doi: 10.3389/fams.2017.00017

## Software

**Robles-Fernández ÁL.** mdd

Available: <https://github.com/alrobles/mdd>. An R package to download and manipulate spatial range maps for mammals according with Mammal Diversity Database taxonomy. Currently the map version is related to MDD v1.2.

**Robles-Fernández ÁL.** mammals\_virus\_text\_class

Available: [https://github.com/alrobles/mammals\\_virus\\_text\\_class](https://github.com/alrobles/mammals_virus_text_class). Mammal parasite paper recommender shiny app. This app search and recommend mammal parasite papers from PubMed database.

The app is running in [https://alroble8.shinyapps.io/mammals\\_virus\\_text\\_class/](https://alroble8.shinyapps.io/mammals_virus_text_class/)

**Robles-Fernández ÁL.** abstractsHostParasites

Available: <https://github.com/alrobles/abstractsHostParasites>. The goal of this R package abstractsHostParasites is to hold tools for generate text classification models applied to classify abstracts of scientific papers.. This classification is performed through PU learning models.

**Robles-Fernández ÁL.** gbifliterature

Available: <https://github.com/alrobles/gbifliterature>. The goal of gbifliterature is to connect with GBIF Literature API and retrieve the full information by year.

**Robles-Fernández ÁL.** ecoininteraction

Available: <https://github.com/alrobles/ecoininteraction>. A toolbox R package to model and predict unknown interactions based, but not limited, on phylogenetic, environmental and geographic

distance among the host.

**Robles-Fernández ÁL. fastJaccard**

Available: <https://github.com/alrobes/fastJaccard>. An R package designed to run the Jaccard similarity for binary matrices in parallel using Rcpp and RcppParallel.

**Robles-Fernández ÁL. geotax**

Available: <https://github.com/alrobes/geotax>. An R package to calculate the probability of interaction given phylogenetic information and transfer results to geographical space.

**Robles-Fernández ÁL. maxnetmap**

Available: <https://github.com/alrobes/maxnetmap>. An R package to fit species distributions models from occurrence records and environmental variables, using 'maxnet' packages and manipulate as a raster output.

## International conference - Oral

2017 Robles-Fernández ÁL 8th Biennial Conference of the International Biogeography Society. Tucson, AZ, USA.

2020 Robles-Fernández ÁL 10th Tenth International Conference on Complex Systems. New England Complex Systems Institute.

2022 Robles-Fernández ÁL 10th Biennial Conference of the International Biogeography Society. Vancouver, BC, Canada.

2022 Robles-Fernández ÁL 101st Annual Meeting of the American Society of Mammalogist.

2024 Robles-Fernández ÁL 11th Biennial Conference of the International Biogeography Society. Prague, Czech Republic.

## Specialized Skills

### Computer Skills

General Proficient with PC (Linux and Windows operating systems) computers with knowledge of LATEX and standard office suite.

Virtualization Technologies VirtualBox

Systems Linux (CentOS, Debian / Ubuntu), Windows Data Bases MySQL,

SQLite

Data Warehouse Hive

Programming Languages R, Python, C++

Big Data Technologies Apache Spark, Apache Hadoop, Apache Impala

Domain of the statistical environment R. Through this statistical environment it is possible to link the connection to relational and non-relational databases in different file systems (for example, hdfs, using R as a link for the manipulation of Apache Spark or Apache Hadoop) .

In addition, within the same statistical environment, it is possible to analyze the data using artificial intelligence algorithms, as well as the generation of predictive statistical models. In addition to this it is also possible to generate a web service (REST API) to expose the machine learning models made to be consumed by any application. (Either in a Java framework, C #, etc.)

To a lesser extent I have had experience with the following programming languages:

Python, Mathematica, C ++, JavaScript

## Professional experience

August 2021 - Actual

Arizona State University

Research Assistant

Feb 2020 – December 2021

Universidad Veracruzana

Research Assistant

Conacyt fellowship

CVU 875979

Feb 2020 – December 2021

In my research I model the ecological interaction among mammals and viruses using environmental, geographical and phylogenetic information using machine learning. Mammal species predicted as highly susceptible coincide with sets of species that have been reported infected in field studies, but it also suggests other species that have not been previously considered or that have been captured in low numbers. Also, the environment (i.e., the distance between the species' optima in bioclimatic dimensions) in combination with geographic and phylogenetic distance is highly relevant in predicting susceptibility to DENV in wild mammals. my results agree

with previous modeling efforts indicating that temperature is an important factor determining DENV transmission, and provide novel insights regarding other relevant factors and the importance of considering wild reservoirs. This modeling framework will aid in the identification of potential DENV reservoirs for future surveillance efforts.

Tenaris Tamsa

Feb 2018 – Feb 2020

Senior Automation Analyst

In charge of optimizing processes from mathematical regression and classification models with artificial intelligence algorithms taking large volumes of data from different sources. Responsible for statistical modeling and natural language processing. Maintaining relational and non-relational databases. Conducting data manipulation, cleaning and auditing. Additionally, in charge of data mining via different regression algorithms and classification methods. Implementing machine learning predictive models. Developing web applications with graphic information derived from the machine learning models, such as dashboards with KPIs for process monitoring. Responsible for preparing BI reports from case studies to obtain added value from the data.

Information from a production line was collected from different sources on a distributed computing system. The system was mounted on HDFS and hadoop. Apache Spark controlled with Python through PySpark and with R through sparklyr was used to manipulate the data. The information contained in the HDFS system was information on machine signals, text information from operator comments as well as information from industrial planning systems. Information on other relational database systems was also handled specifically in Microsoft SQL Server

Experience with NLP.

Within the projects carried out with NLP, a comment classification system was generated. Production lines generally have failures that are reported in human language. Many times the problems associated with the failure cannot be analyzed because the comments on the same failures are different, since although they may refer to the same problem, two comments can be written very differently in human language. To solve this problem, a system was generated that:

1. Tokenize comments
2. Calculate the tf-idf
3. Calculate cosine similarity between all comments

4. Generate a graph about the cosine similarities of all comments
5. Use a ranking algorithm on the graph to find similar comments.
6. Finally, group the comments with the result of the classification on the graph and accumulate the interruption time on the faults associated with the comments of the same group.
7. Validate the results.

Technologies used:

- Programming Languages:
  - R
    - R Libraries
      - tidyverse
      - tidymodels
      - caret
      - shiny
      - sparklyr
      - igraph
- DataFlow Technologies
  - Apache Spark
  - Apache Kafka
  - Apache Hadoop

INDEPENDENT EXPERIENCE

Suitability maps of two invasive beetle species in Mexico/Instituto de Ecología A. C.

JAN 2016 – DEC 2016

Role: Project Manager

Models were developed to predict the climatic suitability of two invasive beetle species (*Xyleborus glabratus* and *Euwallacea* sp) from climatic information on locations where records of beetle species are kept. The algorithm used is MAXENT and the information was projected into the geography. Data on bioclimatic presences and coverage and NDVI were obtained from the MODIS satellite at 1 km resolution or 30 arcs of a second and maps of suitability for each species were generated. Everything was developed in both Python and R.

Technologies used: Python/Numpy, Python/SciPy, Python/PyQGIS, R/tidyverse, R/dismo, R/raster.

Evolutionary Risk Assessment Web App/Instituto de Ecología A. C.

JAN 2017 – DIC 2017

Role: Project Manager

In this project, a web application was developed for the generation of risk maps of pest dispersal.

The stages of the project were:

- Compilation of information on plants of commercial interest affected by diseases transmitted by ambrosial beetles.
- Compilation of information on the interaction of ambrosial beetles with their host plants throughout the geography.
- Obtaining the phylogenies of ambrosial beetles and their associated hosts from bioinformatics techniques.
- Mathematical modeling of the beetle-plant interaction system.
- Development of a tool for the phytosanitary evaluation based on the mathematical model of the beetle-plant interaction.
- Production of phytosanitary risk maps from the mathematical model.
- Web application for the deployment of information and evaluation of phytosanitary risk throughout the country from the mathematical model

Technologies used: R/tidyverse, R/dismo, R/Shiny, Python/Numpy, Python/SciPy

**Other experience:**

- Proficient with implementation of Ecological Niche Modeling algorithms in R
- Proficient with Geographic Information Systems in R
- In-depth knowledge of R. (Data visualization, transformation and analysis. Modelling and app development with shiny)

**Research Interests**

- Complex systems
- Quantitative ecology
- Ecological niche theory