Methodology

# Rodent reservoirs for the New World Arenaviruses (NWA)

Existing literature search revealed potential host species of arena viruses that act as reservoirs for spillover of NWA in humans. Since involvement of certain rodent species as reservoirs for rare NWA are ambiguous or underreported in the literature, we selected six well-established rodent species that act as reservoirs for NWA in their zoonotic spillover (Table 1). These six species, namely, *Zygodontomys brevicauda* and *Sigmodon Alstoni* for Guanarito virus, *Calomys callosus* for Machupo virus, *Calomys musculinus, Calomys laucha and Oligoryzomys flavescens* for Junin virus, have established connection to outbreaks of the respective NWA in human populations (Table 1). The distribution of these rodent reservoirs over the geographical bounds of the reported outbreak of NWA in the past were modeled using species distribution models (SDMs).

## Species Distribution Model (SDM)

We followed a similar protocol to the one we used previously for modeling the species distribution of *Calomys musculinus* in Argentina 1. The protocol was adapted to each of the rodent species selected as mentioned above. Here is a summary of the SDM protocol used to model the rodent reservoirs of NWA.

## Data

Occurrence data on all six rodent species were sourced from GBIF database (presence-only) 2. This data was cleaned to remove duplicates and restricted to years 1990 to the last reported record of occurrence in the GBIF database. Presence-only data was preprocessed to be converted to 1:1 presence-absence data using habitat suitability analysis with R-package USE::paSampling() utilizing a uniform approach 3. Environmental and ecological data was sourced from open-source datasets such as Bioclimactic variables from World Clim (Bioclim), Normalized Differential Vegetation Index (NDVI) and Digital Elevation Model (DEM) from MODIS satellite database and Land Use data from LCLUC program maintained jointly by NASA and University of Maryland 4–7. These datasets were resampled to the highest common resolution and clipped to the extent of species occurrence data mentioned above for each individual rodent reservoir. In total three sets of 24 to 26 rasters and 6 occurrence datasets were used for SDM (Table 2).

Similarly, for predicting the species distribution in the future, analogous raster data for Bioclimactic variables 8, NDVI (data from 2020 for NDVI 5), DEM 6 and land use 9 were downloaded for Shared Socioeconomic Pathways 6 (SSP) scenarios of Moderate (SSP 2.45) and Extreme (SSP 5.85) predictions for climate change in the future as established by Coupled Model Intercomparison Project v6 for years 2041 to 2060 10. These datasets were also preprocessed and resampled and averaged over the 20 year time horizon in the future.

## Modeling algorithms

SDMs were developed by fitting four ensemble tree-based classifier algorithms, namely, Random Forest (RF), Extra Trees (ET), XGBoost (XGB) and Light Gradient Boost Model (LGBM). Data was split in train-test ratio of 4:1 (80% train ;20% test) and missing values were imputed using SimpleImputer from sklearn python library 11. Each algorithm was trained with 5-fold cross validation and the predictions for training and the test set were generated to extract cross-validation accuracy, precision, Area under the ROC curve (AUC-ROC), recall and the test-confusion matrix. This process was iterated [1000] times to add to the robustness of the models using 1000 different presence-absence samples from the occurrence data.

Since hypertuning did not yield significant improvements to any of the four classifier algorithms (results not shown), the default hyper parameters were used. Since we expected high level of collinearity between the rasters based on spatial correlation analysis (results not shown), a recursive feature selection (RFE) was performed in each iteration in order to improve the accuracy of the model without having potential issues with multicollinearity. 10 features were selected in each iteration per classifier algorithm based on stratified cross validation of RFE performed before the iterative model training using RFECV() from sklearn library of python 11. Final fitted models were taken as average of all the iterations and of the four classifier algorithms.

Based on the fitted models, the current distribution probabilities for the respective rodent species were imputed in the geographical extent of their occurrence data using pyimpute library in python 12.Similarly, the future distributions of the rodents in response to the two SSP 2.45 and SSP 5.85 scenarios of CMIP6 climate change were imputed based on the fitted models. Changes in the distribution probabilities were mapped on the raster for each of the scenarios by subtracting current probabilities from the future probabilities of species distribution. The process is illustrated in Figure 1.

## Potential hotspots for zoonotic spillover and force of infection

Based on the distribution probabilities of the SDM for each rodent reservoir, a risk profile for human outbreak i.e. zoonotic spillover of NWA was modeled using infection dynamics simulation. The current and CMIP6 scenario-based projection of human population density was overlaid on the SDM probabilities to generate a risk profile based on the force of infection for NWA spillover. Force Of Infection (FOI) was calculated based on contact rate between the humans and the rodent reservoirs and the possibility of rodent testing positive as reservoir for the NWA.

The mechanistic model used for estimating FOI was a density dependent contact rate model with binomial sampling as detailed in Eq 1.

Eq (1)

Where, is the force of infection defined as the contact between susceptible humans and infectious rodents resulting in successful transmission of infection, is the transmission rate parameter derived from review of analogous viral transmission dynamics studies (see Appendix: Force of Infection), is the population of susceptible humans set at 0.95 times that of total human population in the same geospatial coordinates of SDMs (author’s expertise and from study performed on Lassa Fever in Nigeria 13) and is the proportion of infectious rodent population based on the binomial sampling between 1 and 15 rodents per grid cell adjusted with the probability of presence of rodents in the given geospatial coordinates based of the SDMs. The denominator represents the total density of the interacting populations of human hosts and rodent reservoirs. Eq (1) was adapted from generalized formula used in similar transmission studies of vector and rodent borne infections with and without the effect of climate change 14–16.

The map depicting FOI for each of the three viruses under study was further reclassified in 10% quantiles and the geospatial zones which had FOI values in the 10th quantile (>= 90% FOI) were depicted as potential hotspot zones for outbreak of zoonotic arenaviruses. Similar methodology was followed using the projected SDM probabilities and projected human population under the SSP 2.45 and SSP 5.85 scenarios of CMIP6 17.

# Tables

Table 1. New World Arenaviruses (NWA) that were reported to cause zoonotic outbreaks in humans with their reported rodent reservoirs

|  |  |  |  |
| --- | --- | --- | --- |
| **Common name of virus** | **Common name of the disease** | **Reported rodent reservoir** | **Outbreaks reported** |
| **Junin Virus** | Argentine Hemorrhagic fever | *1. Calomys musculinus (Dryland Vesper mouse)* | 1950s, 1990-2018 |
|  |  | *2. Calomys laucha (Small Vesper mouse)* |  |
|  |  | *3. Oligoryzomys flavescens (Yellow pygmy rice rat)* |  |
| **Machupo Virus** | Bolivian Hemorrhagic fever | *1. Calomys callosus(Large Vesper mouse)* | 1960s |
| **Guanarito Virus** | Venezuelan Hemorrhagic fever | *1. Zygodontomys brevicauda (Short-tailed Cane mouse)* | 1989-2010 |
|  |  | *2. Sigmodon alstoni (Alston's Cotton rat)* |  |

Table 2. Datasets and raster variables used in Species Distribution Models (SDMs)

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable/ Raster name** | **Source** | **Year range** | **Notes** |
| **Occurrence data** | GBIF |  |  |
| *Z. brevicauda* |  | 1891-2023 | 6142 reported (688 presence-only - unique) |
| *S.alstoni* |  | 1911-2017 | 326 reported (62 presence-only - unique) |
| *C.musculinus* |  | 1955-2021 | 2287 reported (194 presence-only - unique) |
| *C.laucha* |  | 1900-2024 | 2898 reported (137 presence-only - unique) |
| *O.flavescens* |  | 1900-2024 | 1136 reported (171 presence-only - unique) |
| *C.callosus* |  | 1914-2004 | 2746 reported (165 presence-only - unique) |
| **Bioclimatic Variables** | World Clim Version 2 | |  |
| Annual Mean Temperature |  | 1970-2000 | Bclim 1 |
| Mean Diurnal Range |  | 1970-2000 | Bclim 2 |
| Isothermality |  | 1970-2000 | Bclim 3 |
| Temperature Seasonality |  | 1970-2000 | Bclim 4 |
| Max Temperature of Warmest Month |  | 1970-2000 | Bclim 5 |
| Min Temperature of Coldest Month |  | 1970-2000 | Bclim 6 |
| Temperature Annual Range |  | 1970-2000 | Bclim 7 |
| Mean Temperature of Wettest Quarter |  | 1970-2000 | Bclim 8 |
| Mean Temperature of Driest Quarter |  | 1970-2000 | Bclim 9 |
| Mean Temperature of Warmest Quarter |  | 1970-2000 | Bclim 10 |
| Mean Temperature of Coldest Quarter |  | 1970-2000 | Bclim 11 |
| Annual Precipitation |  | 1970-2000 | Bclim 12 |
| Precipitation of Wettest Month |  | 1970-2000 | Bclim 13 |
| Precipitation of Driest Month |  | 1970-2000 | Bclim 14 |
| Precipitation Seaonality |  | 1970-2000 | Bclim 15 |
| Precipitation of Wettest Quarter |  | 1970-2000 | Bclim 16 |
| Precipitation of Driest Quarter |  | 1970-2000 | Bclim 17 |
| Precipitation of Warmest Quarter |  | 1970-2000 | Bclim 18 |
| Precipitation of Coldest Quarter |  | 1970-2000 | Bclim 19 |
| **Elevation Data** | | | |
| Digital Elevation Model (DEM) | SRTM v4 (NASA, CGIAR) | 2000 | Elevation in meters |
| **Vegetation and Land Use** | | | |
| NDVI | MODIS (NASA - USGS) | 2000 | Normalized Difference Vegetation Index for minimum cloud cover |
| Primary Land | LCLUC (UMA, NASA) | 1990-2005 | lu\_othr |
| Crop Land | LCLUC (UMA, NASA) | 1990-2005 | lu\_crop |
| Urban Land | LCLUC (UMA, NASA) | 1990-2005 | lu\_urbn |
| Secondary Land | LCLUC (UMA, NASA) | 1990-2005 | lu\_secd |
| Pasture Land | LCLUC (UMA, NASA) | 1990-2005 | lu\_past |
| **Future scenarios data from CMIP6** | | | |
| Bioclimatic Variables set 1 | World Clim CMIP6 SSP 2.45 | 2041-2060 |  |
| Bioclimatic Variables set 2 | World Clim CMIP6 SSP 5.85 | 2041-2060 |  |
| Land Use 2.45 | LCLUC CMIP6 SSP245 | 2041-2060 |  |
| Land Use 5.85 | LCLUC CMIP6 SSP585 | 2041-2060 |  |
| NDVI | MODIS (NASA - USGS) | 2020-present |  |
| Digital Elevation Model (DEM) | SRTM v4 (NASA, CGIAR) | 2000 | Same as current data |

# Figures

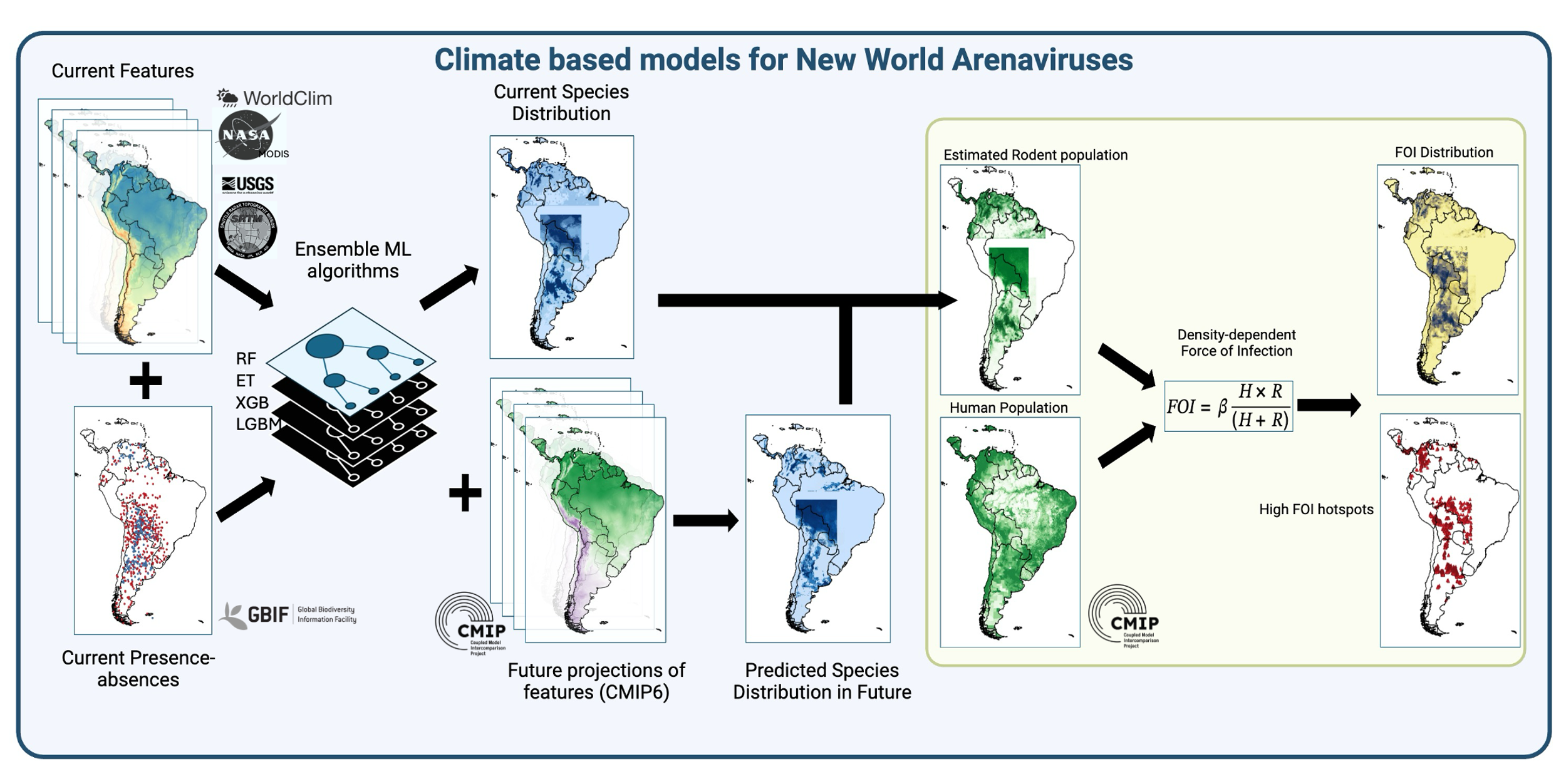


Figure 1. Schematic illustration of the framework for deriving the impact of climate change on the transmission risk for New World Arenaviruses in South America.

# References

1. Flores-Perez, N., Kulkarni, P., Uhart, M. & Pandit, P. Shifting human-rodent interfaces under climate change: modeling the distribution of the reservoir for Junin virus and associated drivers. bioRxiv 2024–06 (2024).

2. GBIF. What is GBIF? https://www.gbif.org/what-is-gbif.

3. USE it: Uniformly sampling pseudo‐absences within the environmental space for applications in habitat suitability models - Da Re - 2023 - Methods in Ecology and Evolution - Wiley Online Library. https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.14209.

4. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. J. R. Meteorol. Soc. **25**, 1965–1978 (2005).

5. Didan, K. MODIS/Terra Vegetation Indices 16-Day L3 Global 500m SIN Grid V061. NASA EOSDIS Land Processes Distributed Active Archive Center https://doi.org/10.5067/MODIS/MOD13A1.061 (2021).

6. Earth Resources Observation And Science (EROS) Center. Shuttle Radar Topography Mission (SRTM) 1 Arc-Second Global. U.S. Geological Survey https://doi.org/10.5066/F7PR7TFT (2017).

7. Potapov, P. et al. The Global 2000-2020 Land Cover and Land Use Change Dataset Derived From the Landsat Archive: First Results. Front. Remote Sens. **3**, 856903 (2022).

8. Presentation and Evaluation of the IPSL‐CM6A‐LR Climate Model - Boucher - 2020 - Journal of Advances in Modeling Earth Systems - Wiley Online Library. https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019MS002010.

9. Hurtt, G. et al. Harmonization of Global Land Use Change and Management for the Period 2015-2300. Earth System Grid Federation https://doi.org/10.22033/ESGF/INPUT4MIPS.10468 (2019).

10. Masson-Delmotte, V., P. Zhai, A. Pirani, S.L., Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K., & Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou. Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC\_AR6\_WGI\_SPM\_final.pdf (2021).

11. Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. Mach. Learn. PYTHON.

12. Matthew Perry. pyimpute: Utilities for applying scikit-learn to spatial datasets. (2022).

13. Barua, S., Dénes, A. & Ibrahim, M. A. A seasonal model to assess intervention strategies for preventing periodic recurrence of Lassa fever. Heliyon **7**, (2021).

14. Begon, M. et al. A clarification of transmission terms in host-microparasite models: numbers, densities and areas. Epidemiol. Infect. **129**, 147–153 (2002).

15. Medone, P., Ceccarelli, S., Parham, P. E., Figuera, A. & Rabinovich, J. E. The impact of climate change on the geographical distribution of two vectors of Chagas disease: implications for the force of infection. Philos. Trans. R. Soc. B Biol. Sci. **370**, 20130560 (2015).

16. Ferrari, M. J., Perkins, S. E., Pomeroy, L. W. & Bjørnstad, O. N. Pathogens, Social Networks, and the Paradox of Transmission Scaling. Interdiscip. Perspect. Infect. Dis. **2011**, 267049 (2011).

17. Jones, K. E. et al. Global trends in emerging infectious diseases. Nature **451**, 990–993 (2008).