

# Estratificación temporal de Aedes Aegypti basada en herramientas geoespaciales y aprendizaje automático

Juan M. Scavuzzo

Facultad de Matemática, Astronomía, Física y Computación  
Universidad Nacional de Córdoba

Diciembre de 2018



- 1 Motivaciones de este trabajo
- 2 Objetivos
- 3 Algunos conceptos importantes
- 4 Modelado de la población del vector de Dengue
- 5 Generalización espacial de modelos epidemiológicos
- 6 Resumen de alto nivel
- 7 Datos de color

# *Motivaciones de este trabajo*

# Motivación de este trabajo: problemática epidemiológica

- Aedes aegypti es el principal vector de Dengue, Chikungunya, Zika y Fiebre Amarilla urbana

# Motivación de este trabajo: problemática epidemiológica

- Aedes aegypti es el principal vector de Dengue, Chikungunya, Zika y Fiebre Amarilla urbana
- Datos de la Organización Mundial de la Salud:

# Motivación de este trabajo: problemática epidemiológica

- Aedes aegypti es el principal vector de Dengue, Chikungunya, Zika y Fiebre Amarilla urbana
- Datos de la Organización Mundial de la Salud:
  - 80 millones de personas se infectan de Dengue anualmente

# Motivación de este trabajo: problemática epidemiológica

- Aedes aegypti es el principal vector de Dengue, Chikungunya, Zika y Fiebre Amarilla urbana
- Datos de la Organización Mundial de la Salud:
  - 80 millones de personas se infectan de Dengue anualmente
  - 550 mil enfermos requieren hospitalización

# Motivación de este trabajo: problemática epidemiológica

- Aedes aegypti es el principal vector de Dengue, Chikungunya, Zika y Fiebre Amarilla urbana
- Datos de la Organización Mundial de la Salud:
  - 80 millones de personas se infectan de Dengue anualmente
  - 550 mil enfermos requieren hospitalización
  - 20 mil personas mueren

# Motivación de este trabajo: problemática epidemiológica

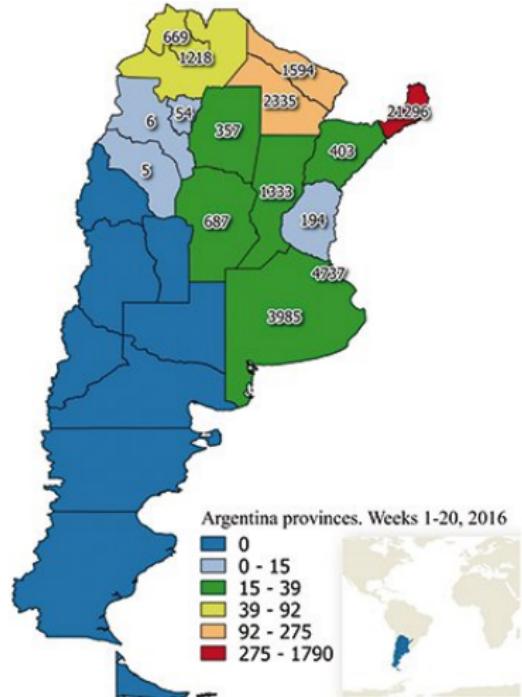
- Aedes aegypti es el principal vector de Dengue, Chikungunya, Zika y Fiebre Amarilla urbana
- Datos de la Organización Mundial de la Salud:
  - 80 millones de personas se infectan de Dengue anualmente
  - 550 mil enfermos requieren hospitalización
  - 20 mil personas mueren
  - 2.500 millones de personas corren riesgo de contraer la enfermedad

# Motivación de este trabajo: problemática epidemiológica

- Aedes aegypti es el principal vector de Dengue, Chikungunya, Zika y Fiebre Amarilla urbana
- Datos de la Organización Mundial de la Salud:
  - 80 millones de personas se infectan de Dengue anualmente
  - 550 mil enfermos requieren hospitalización
  - 20 mil personas mueren
  - 2.500 millones de personas corren riesgo de contraer la enfermedad
  - Más de 100 países con transmisión endémica

# Motivación de este trabajo: problemática epidemiológica

Taza de Dengue a nivel provincia, cada 100mil habitantes



# Motivación de este trabajo: problemática epidemiológica

Características del Aedes aegypti

# Motivación de este trabajo: problemática epidemiológica

## Características del Aedes aegypti

- Gran capacidad adaptativa

# Motivación de este trabajo: problemática epidemiológica

## Características del Aedes aegypti

- Gran capacidad adaptativa
- Resistencia a insecticidas

# Motivación de este trabajo: problemática epidemiológica

## Características del Aedes aegypti

- Gran capacidad adaptativa
- Resistencia a insecticidas
- Resistencia de huevos a la desecación

# Motivación de este trabajo: problemática epidemiológica

## Características del Aedes aegypti

- Gran capacidad adaptativa
- Resistencia a insecticidas
- Resistencia de huevos a la desecación
- Presencia en el medio urbano

# Motivación de este trabajo: problemática epidemiológica

## Características del Aedes aegypti

- Gran capacidad adaptativa
- Resistencia a insecticidas
- Resistencia de huevos a la desecación
- Presencia en el medio urbano
- Preferencia de cría en contenedores artificiales

# Motivación de este trabajo: sistemas de modelado actuales

Modelar utilizando información satelital

# Motivación de este trabajo: sistemas de modelado actuales

Modelar utilizando información satelital

- Información ambiental con alcance regional

# Motivación de este trabajo: sistemas de modelado actuales

Modelar utilizando información satelital

- Información ambiental con alcance regional
- Información espacio-temporal

# Motivación de este trabajo: sistemas de modelado actuales

## Modelar utilizando información satelital

- Información ambiental con alcance regional
- Información espacio-temporal
- Grandes avances en los últimos años

# Motivación de este trabajo: sistemas de modelado actuales

## Modelar utilizando información satelital

- Información ambiental con alcance regional
- Información espacio-temporal
- Grandes avances en los últimos años

# Motivación de este trabajo: sistemas de modelado actuales

Modelar utilizando información satelital

- Información ambiental con alcance regional
- Información espacio-temporal
- Grandes avances en los últimos años

Pero actualmente se utilizan modelos lineales para relacionar las distintas variables!

# Motivación de este trabajo

Algunas cuestiones a tener en cuenta

# Motivación de este trabajo

## Algunas cuestiones a tener en cuenta

- La prevención de las enfermedades en cuestión debe ser a través de control de vectores

# Motivación de este trabajo

## Algunas cuestiones a tener en cuenta

- La prevención de las enfermedades en cuestión debe ser a través de control de vectores
  - Modelos predictivos!

# Motivación de este trabajo

## Algunas cuestiones a tener en cuenta

- La prevención de las enfermedades en cuestión debe ser a través de control de vectores
  - Modelos predictivos!
- Será correcto asumir relaciones lineales entre las variables ambientales y la abundancia del vector?

# Motivación de este trabajo

## Algunas cuestiones a tener en cuenta

- La prevención de las enfermedades en cuestión debe ser a través de control de vectores
  - Modelos predictivos!
- Será correcto asumir relaciones lineales entre las variables ambientales y la abundancia del vector?
  - Modelos no-lineales con... aprendizaje automático!

# Motivación de este trabajo

## Algunas cuestiones a tener en cuenta

- La prevención de las enfermedades en cuestión debe ser a través de control de vectores
  - Modelos predictivos!
- Será correcto asumir relaciones lineales entre las variables ambientales y la abundancia del vector?
  - Modelos no-lineales con... aprendizaje automático!
- Si es un sistema regional, cómo extrapoló los modelos?

# Motivación de este trabajo

## Algunas cuestiones a tener en cuenta

- La prevención de las enfermedades en cuestión debe ser a través de control de vectores
  - Modelos predictivos!
- Será correcto asumir relaciones lineales entre las variables ambientales y la abundancia del vector?
  - Modelos no-lineales con... aprendizaje automático!
- Si es un sistema regional, cómo extrapoló los modelos?
  - A través de relaciones entre características ambientales!

# Objetivos

# Objetivos

- Implementar una herramienta, sencilla, para generar modelos predictivos

# Objetivos

- Implementar una herramienta, sencilla, para generar modelos predictivos
- Validar la hipótesis de que "modelos no-lineales son mejores para predecir la oviposición que los lineales"

# Objetivos

- Implementar una herramienta, sencilla, para generar modelos predictivos
- Validar la hipótesis de que "modelos no-lineales son mejores para predecir la oviposición que los lineales"
- Proponer una solución a la problemática de escases de datos que se evidencia al pensar en sistemas regionales de estimación de riesgo

## ***Algunos conceptos importantes***

# Algunos Conceptos: *Epidemiología Panorámica*

# Algunos Conceptos: *Epidemiología Panorámica*

- La teledetección y su capacidad de adquirir información

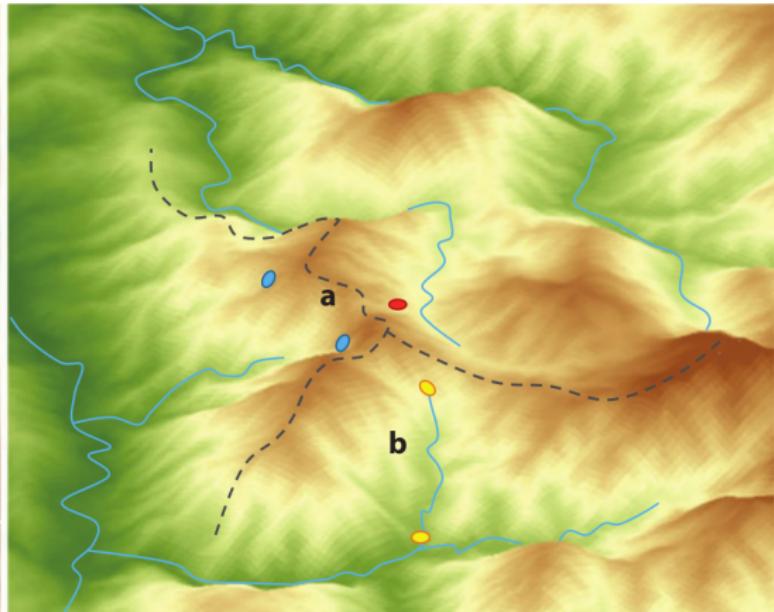
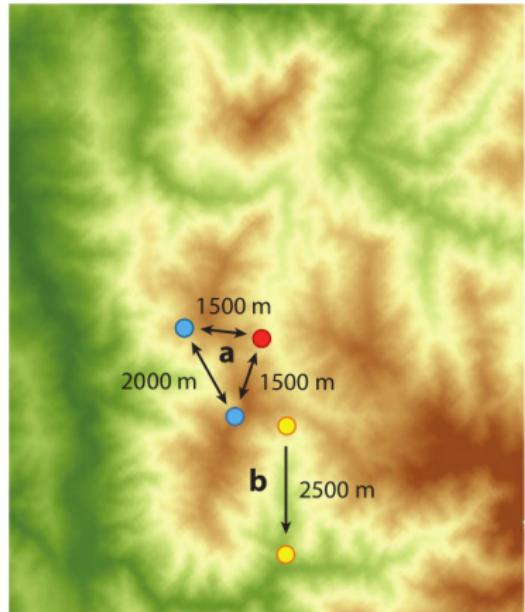
# Algunos Conceptos: *Epidemiología Panorámica*

- La teledetección y su capacidad de adquirir información
- Información sobre hábitat de insectos y artrópodos

# Algunos Conceptos: *Epidemiología Panorámica*

- La teledetección y su capacidad de adquirir información
- Información sobre hábitat de insectos y artrópodos
- Fuente de datos sobre la distribución espacio-temporal de enfermedades transmitidas por vectores (Pavlovsky)

# Algunos Conceptos: *Epidemiología Panorámica*



# Algunos Conceptos: *Epidemiología Panorámica*

- Focalidad
  - **Vectores** con capacidad de transmisión de la infección
  - Vertebrados capaces de funcionar como **reservorio** de la infección
  - **Huéspedes** susceptibles, como humanos o animales domésticos
- Ecología panorámica

# Algunos Conceptos: *Epidemiología Panorámica*

- Focalidad
  - **Vectores** con capacidad de transmisión de la infección
  - Vertebrados capaces de funcionar como **reservorio** de la infección
  - **Huéspedes** susceptibles, como humanos o animales domésticos
- Ecología panorámica

## *Epidemiología Panorámica*

# Algunos Conceptos: *Aprendizaje automático (ML)*

*Se dice que un programa de computadora **aprende** de experiencia  $E$  con respecto a alguna tarea  $T$  y una métrica de rendimiento  $M$ , si con la experiencia  $E$  se incrementa su rendimiento en la tarea  $T$ , medida por  $M$ .*

**Tom Mitchell, 1997**

# Algunos Conceptos: *Aprendizaje automático (ML)*

- Enfoque empírico efectivo para *regresiones* y *clasificaciones*

# Algunos Conceptos: *Aprendizaje automático (ML)*

- Enfoque empírico efectivo para *regresiones y clasificaciones*
- Distintos métodos:

# Algunos Conceptos: *Aprendizaje automático (ML)*

- Enfoque empírico efectivo para *regresiones y clasificaciones*
- Distintos métodos:
  - Supervisados: Regresiones Lineales, SVMs, ANNs, DTRs...

# Algunos Conceptos: *Aprendizaje automático (ML)*

- Enfoque empírico efectivo para *regresiones y clasificaciones*
- Distintos métodos:
  - Supervisados: Regresiones Lineales, SVMs, ANNs, DTRs...
  - No-supervisados: K-NNs, K-means, PCA...

# Algunos Conceptos: *Aprendizaje automático (ML)*

- Enfoque empírico efectivo para *regresiones y clasificaciones*
- Distintos métodos:
  - Supervisados: Regresiones Lineales, SVMs, ANNs, DTRs...
  - No-supervisados: K-NNs, K-means, PCA...
  - Semi-supervisados

# Algunos Conceptos: *Aprendizaje automático (ML)*

- Enfoque empírico efectivo para *regresiones y clasificaciones*
- Distintos métodos:
  - Supervisados: Regresiones Lineales, SVMs, ANNs, DTRs...
  - No-supervisados: K-NNs, K-means, PCA...
  - Semi-supervisados
- Usado en muchos ámbitos:

# Algunos Conceptos: *Aprendizaje automático (ML)*

- Enfoque empírico efectivo para *regresiones y clasificaciones*
- Distintos métodos:
  - Supervisados: Regresiones Lineales, SVMs, ANNs, DTRs...
  - No-supervisados: K-NNs, K-means, PCA...
  - Semi-supervisados
- Usado en muchos ámbitos:
  - Académico

# Algunos Conceptos: *Aprendizaje automático (ML)*

- Enfoque empírico efectivo para *regresiones y clasificaciones*
- Distintos métodos:
  - Supervisados: Regresiones Lineales, SVMs, ANNs, DTRs...
  - No-supervisados: K-NNs, K-means, PCA...
  - Semi-supervisados
- Usado en muchos ámbitos:
  - Académico
  - Industrial

# Algunos Conceptos: *Aprendizaje automático (ML)*

- Enfoque empírico efectivo para *regresiones y clasificaciones*
- Distintos métodos:
  - Supervisados: Regresiones Lineales, SVMs, ANNs, DTRs...
  - No-supervisados: K-NNs, K-means, PCA...
  - Semi-supervisados
- Usado en muchos ámbitos:
  - Académico
  - Industrial
  - Gubernamental

# Algunos Conceptos: *Métodos Supervisados*

- Aprenden a través de pares de ejemplos  $(X, Y_{verd})$

# Algunos Conceptos: *Métodos Supervisados*

- Aprenden a través de pares de ejemplos  $(X, Y_{verd})$
- Conjuntos de entrenamiento y validación

# Algunos Conceptos: *Métodos Supervisados*

- Aprenden a través de pares de ejemplos  $(X, Y_{verd})$
- Conjuntos de entrenamiento y validación
- Evitar *overfitting*

# Algunos Conceptos: *Métodos Supervisados*

- Aprenden a través de pares de ejemplos  $(X, Y_{verd})$
- Conjuntos de entrenamiento y validación
- Evitar *overfitting*
- Ajuste de hiperparámetros...

# Algunos Conceptos: *Parámetros y Hiperparámetros*

- Los modelos de ML poseen *parámetros* e *hiperparámetros*
  - Los hiperparámetros definen el comportamiento durante el proceso de entrenamiento

# Algunos Conceptos: *Parámetros y Hiperparámetros*

- Los modelos de ML poseen *parámetros e hiperparámetros*
  - Los hiperparámetros definen el comportamiento durante el proceso de entrenamiento
  - Los parámetros se ajustan para definir el modelo luego del entrenamiento

# Algunos Conceptos: *Ajuste de hiperparámetros*

Supongamos una **Ridge Regression**:

$$\hat{y}(w, x) = w_0 + w_1x_1 + \dots + w_p x_p$$

$$\min_w \|Xw - y\|_2^2 + \alpha \|w\|_2^2$$

# Algunos Conceptos: Ajuste de hiperparámetros

Supongamos una **Ridge Regression**:

$$\hat{y}(w, x) = w_0 + w_1x_1 + \dots + w_px_p$$

$$\min_w \|Xw - y\|_2^2 + \alpha \|w\|_2^2$$

Un vistazo de sus hiperparámetros...

```
Ridge(alpha=1.0, fit_intercept=True, normalize=False,  
      copy_X=True, max_iter=None, tol=0.001,  
      solver='auto', random_state=None)
```

# Algunos Conceptos: *Ajuste de hiperparámetros*

Supongamos, ahora, un "pequeño" **Perceptron Multicapa**:

# Algunos Conceptos: Ajuste de hiperparámetros

Supongamos, ahora, un "pequeño" **Perceptron Multicapa**:

```
MLPRegressor(activation='relu', alpha=0.0001,  
             batch_size='auto', beta_1=0.9,  
             beta_2=0.999, early_stopping=False,  
             epsilon=1e-08, hidden_layer_sizes=(2, 2),  
             learning_rate='constant',  
             learning_rate_init=0.001,  
             max_iter=200, momentum=0.9,  
             nesterovs_momentum=True, power_t=0.5,  
             random_state=None, shuffle=True,  
             solver='adam', tol=0.0001,  
             validation_fraction=0.1, verbose=False,  
             warm_start=False)
```

# *Modelado de la población del vector de Dengue*

# Obtención y análisis de datos a utilizar



# Obtención y análisis de datos a utilizar

- Oviposición: ovitrampas

# Obtención y análisis de datos a utilizar

- Oviposición: ovitrampas
- De productos satelitales a variables ambientales:

# Obtención y análisis de datos a utilizar

- Oviposición: ovitrampas
- De productos satelitales a variables ambientales:
  - Propiedades de vegetación: **NDVI (MODIS)**

# Obtención y análisis de datos a utilizar

- Oviposición: ovitrampas
- De productos satelitales a variables ambientales:
  - Propiedades de vegetación: **NDVI** (*MODIS*)
  - Humedad: **NDWI** (*MODIS*)

# Obtención y análisis de datos a utilizar

- Oviposición: ovitrampas
- De productos satelitales a variables ambientales:
  - Propiedades de vegetación: **NDVI** (*MODIS*)
  - Humedad: **NDWI** (*MODIS*)
  - Temperatura de la superficie: **LST** (*MODIS*)

# Obtención y análisis de datos a utilizar

- Oviposición: ovitrampas
- De productos satelitales a variables ambientales:
  - Propiedades de vegetación: **NDVI (MODIS)**
  - Humedad: **NDWI (MODIS)**
  - Temperatura de la superficie: **LST (MODIS)**
  - Precipitación: **TRMM/GPM**

# Obtención y análisis de datos a utilizar

- Oviposición: ovitrampas
- De productos satelitales a variables ambientales:
  - Propiedades de vegetación: **NDVI (MODIS)**
  - Humedad: **NDWI (MODIS)**
  - Temperatura de la superficie: **LST (MODIS)**
  - Precipitación: **TRMM/GPM**
- En todos los casos se contempla un *lag*

# Obtención y análisis de datos a utilizar

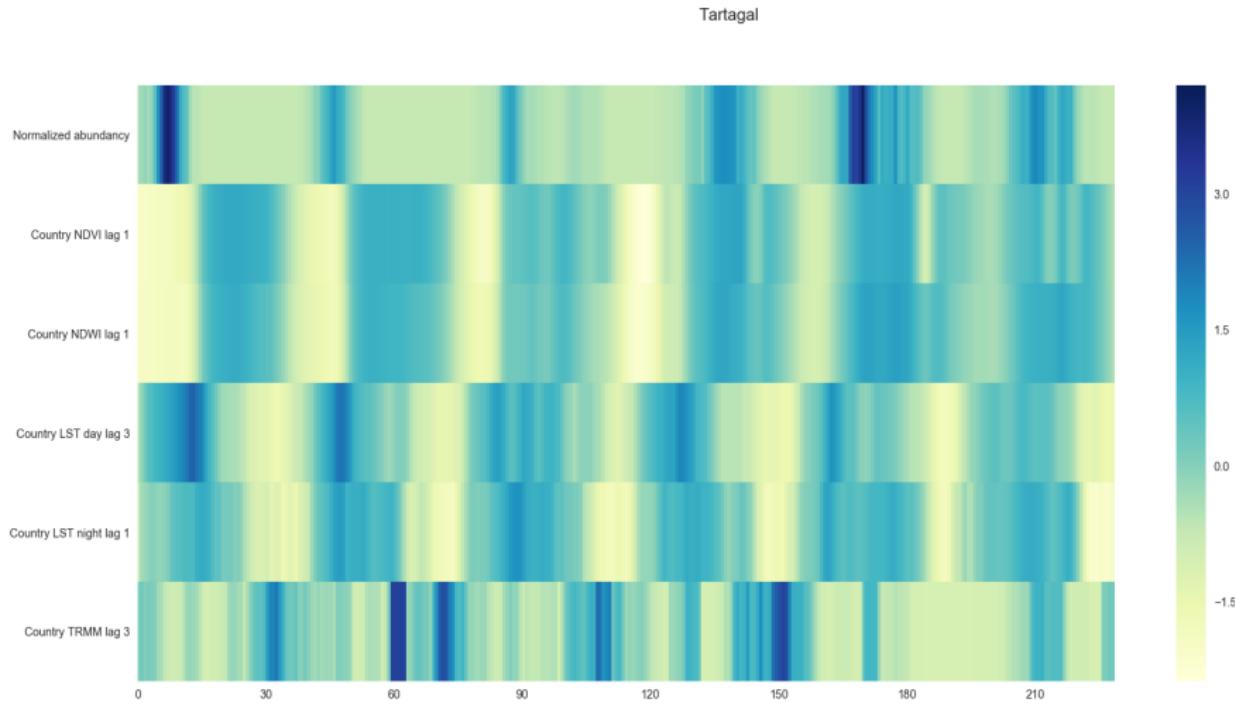
## Datos de área rural y área urbana



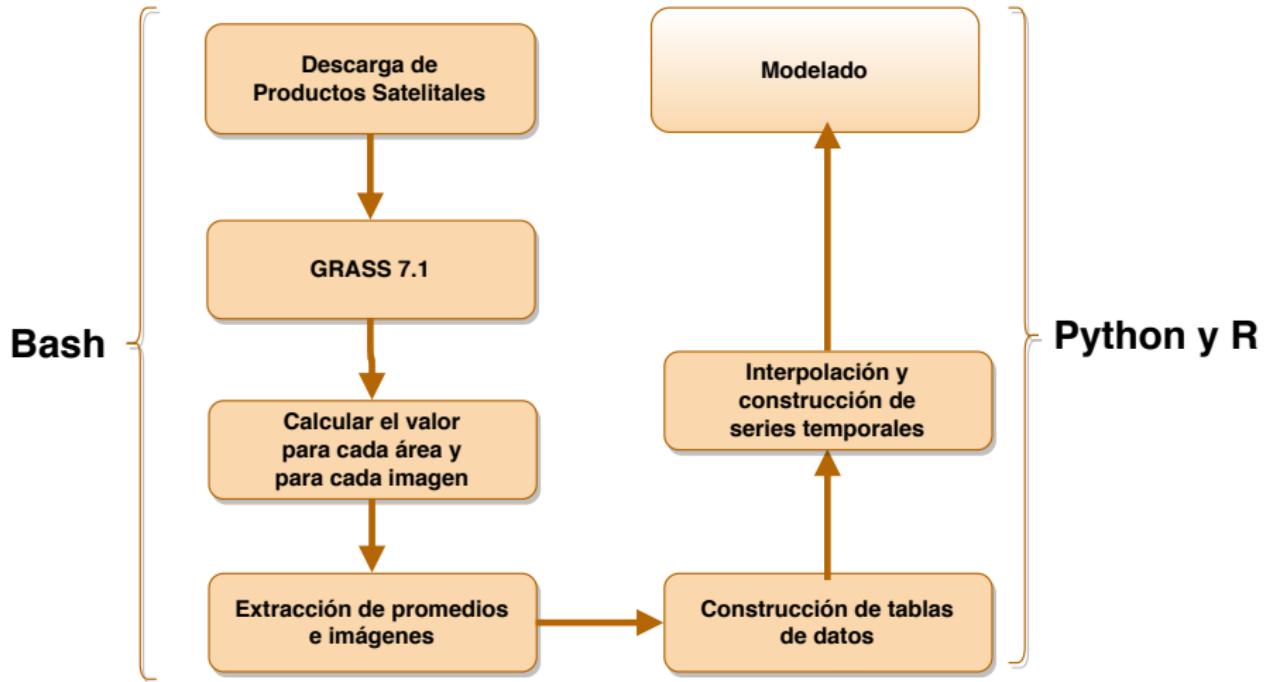
# Análisis y selección de datos a utilizar

- Cinco variables elegidas:
  - NDVI rural lag 1
  - NDWI rural lag 1
  - LST rural dia lag 3
  - LST rural noche lag 1
  - TRMM lag 3
- Normalización con z-score

# Análisis y selección de datos a utilizar



# Sistema completo



# Requerimientos de nuestro sistema

- Facilidad de utilización

# Requerimientos de nuestro sistema

- Facilidad de utilización
- Herramienta de limpieza de datos

# Requerimientos de nuestro sistema

- Facilidad de utilización
- Herramienta de limpieza de datos
- Flexibilidad para cambiar conjuntos de datos

# Requerimientos de nuestro sistema

- Facilidad de utilización
- Herramienta de limpieza de datos
- Flexibilidad para cambiar conjuntos de datos
- Herramienta para ajustar hiperparámetros

# Requerimientos de nuestro sistema

- Facilidad de utilización
- Herramienta de limpieza de datos
- Flexibilidad para cambiar conjuntos de datos
- Herramienta para ajustar hiperparámetros
- Generar modelos en formato utilizable en producción

# Requerimientos de nuestro sistema

- Facilidad de utilización
- Herramienta de limpieza de datos
- Flexibilidad para cambiar conjuntos de datos
- Herramienta para ajustar hiperparámetros
- Generar modelos en formato utilizable en producción
- Flexibilidad para agregar nuevos algoritmos

# Requerimientos de nuestro sistema

- Facilidad de utilización
- Herramienta de limpieza de datos
- Flexibilidad para cambiar conjuntos de datos
- Herramienta para ajustar hiperparámetros
- Generar modelos en formato utilizable en producción
- Flexibilidad para agregar nuevos algoritmos
- Herramienta de visualización de datos

# Módulos del sistema

- Data: *scikit-learn*, *pandas*, *numpy*

# Módulos del sistema

- Data: *scikit-learn, pandas, numpy*
- Models: *scikit-learn*

# Módulos del sistema

- Data: *scikit-learn, pandas, numpy*
- Models: *scikit-learn*
- Tuning: *Irace (Iterated Racing for Automatic Algorithm Configuration)*

# Modelando poblaciones de mosquito

Todos los algoritmos de ML que se utilizaron son de la librería *scikit-learn*

# Modelando poblaciones de mosquito

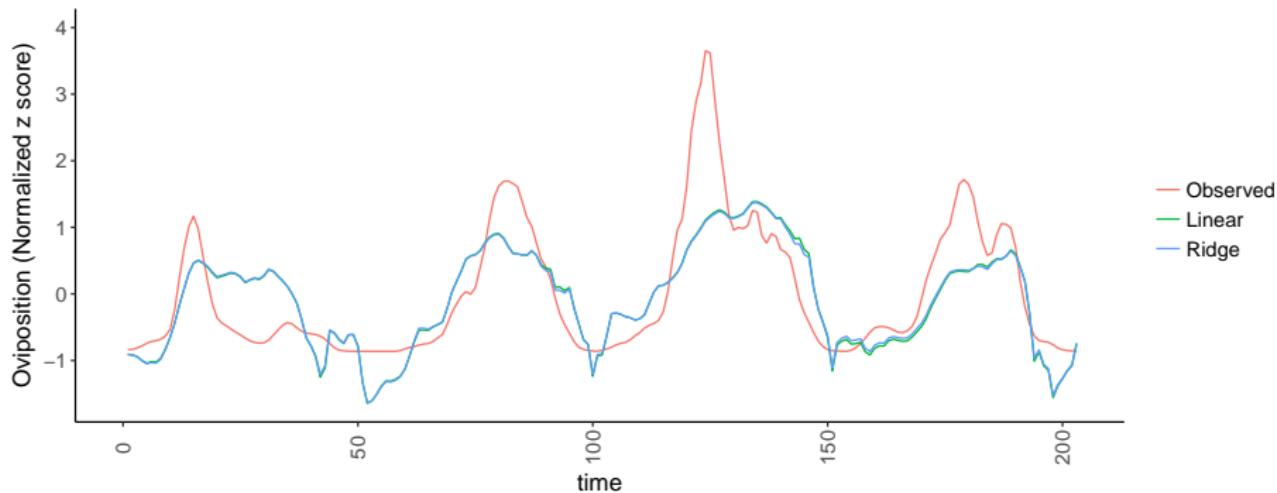
Todos los algoritmos de ML que se utilizaron son de la librería *scikit-learn*

- Lineales:
  - Ridge
  - Tradicional (mínimos cuadrados)
- No-lineales:
  - Regresión por árboles de decisión (DTR)
  - Regresión por K-Vecinos más cercanos (KNNR)
  - Support Vector Regressor (SVR)
  - Perceptron multicapa (MLP)

## *Evaluación y análisis de los modelos generados*

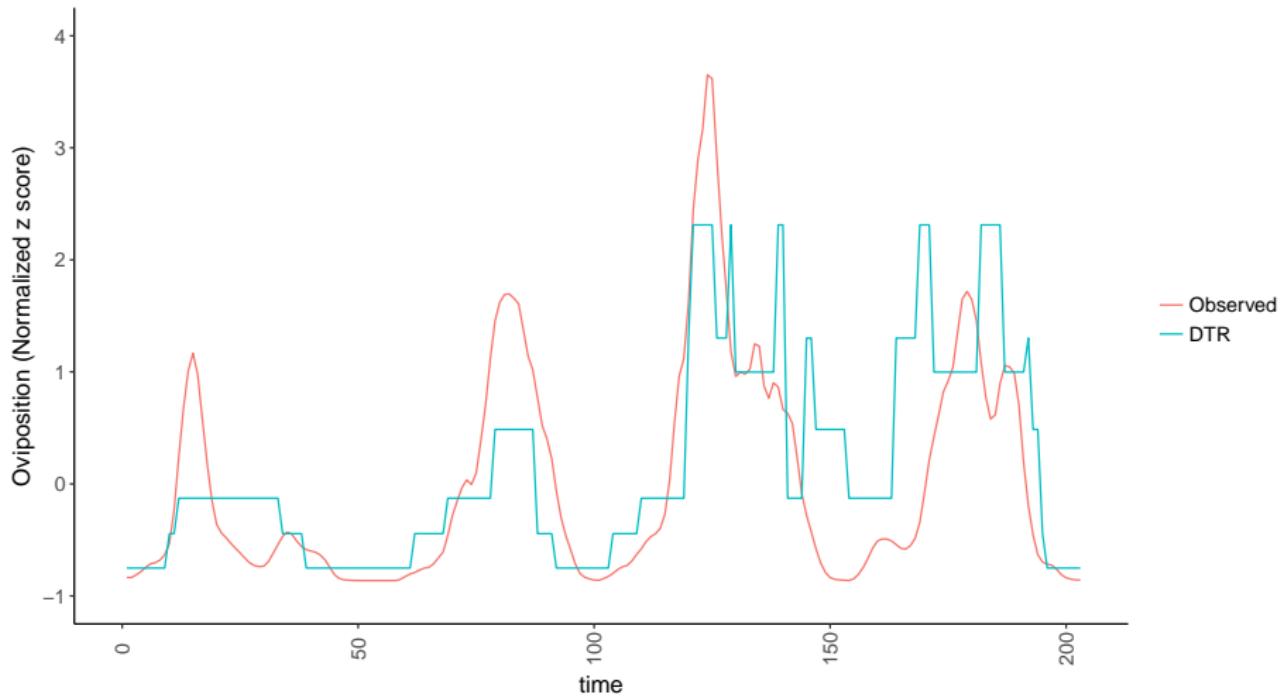
# Evaluación y análisis de los modelos generados

## Métodos Lineales



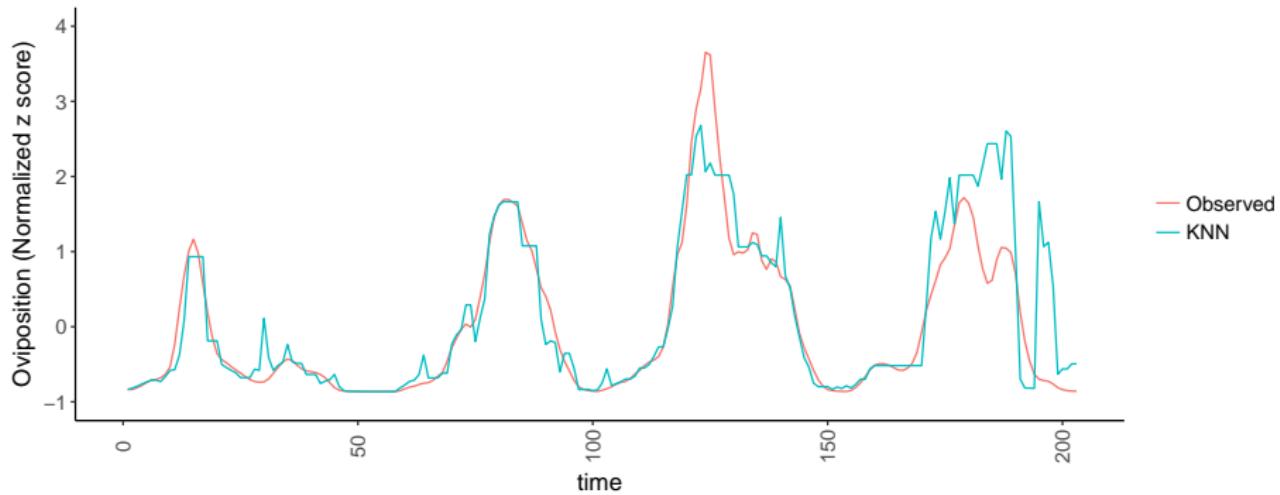
# Evaluación y análisis de los modelos generados

## Regresión de árbol de decisión (DTR)



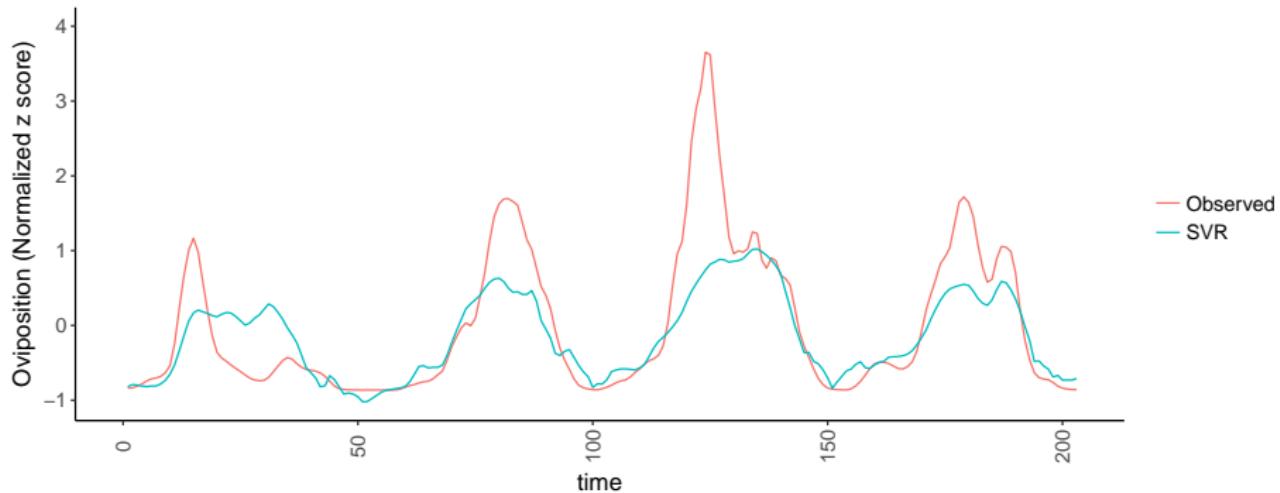
# Evaluación y análisis de los modelos generados

## Regresión de K-Vecinos más cercanos (KNNR)



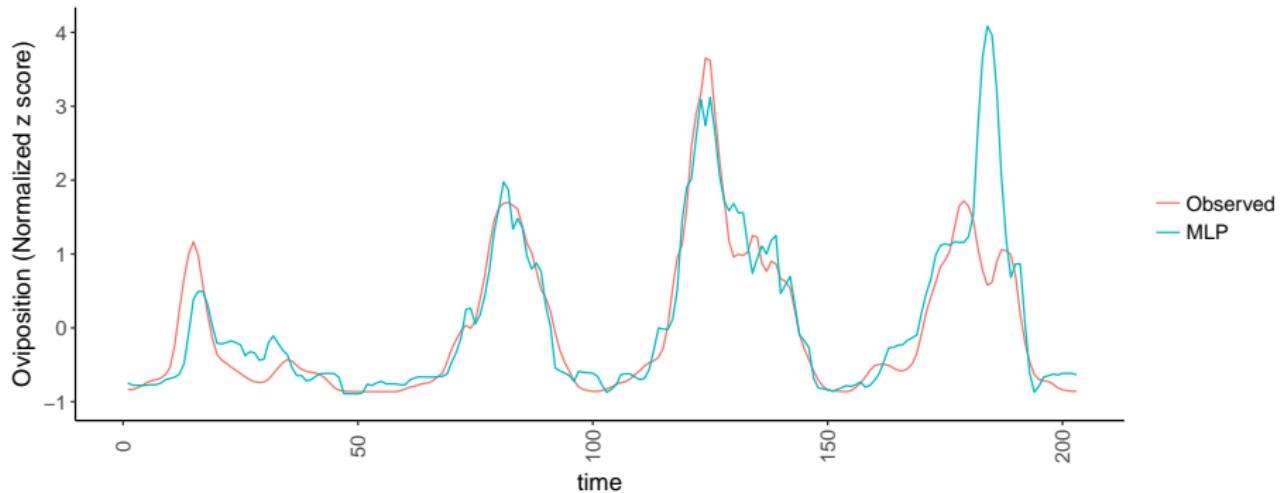
# Evaluación y análisis de los modelos generados

## Support Vector Regressor (SVR)



# Evaluación y análisis de los modelos generados

Perceptron multicapa (MLP)



# Evaluación y análisis de los modelos generados

Resumen de los datos observados y los ajustados

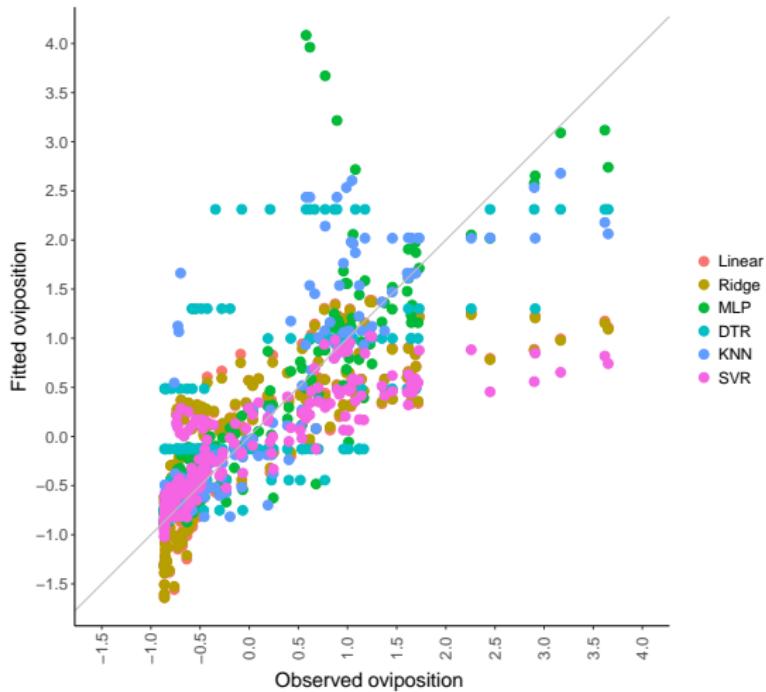
	Mín	$q_{1/4}$	$q_{1/2}$	Media	$q_{3/4}$	Máx
Observado	-0,863	-0,742	-0,487	0,000	0,704	3,652
Lineal	-1,641	-0,716	0,027	-0,087	0,462	1,387
Ridge	-1,638	-0,680	0,028	-0,084	0,459	1,370
MLP	-0,894	-0,677	-0,323	0,093	0,716	4,084
DTR	-0,752	-0,752	-0,128	0,138	0,998	2,312
KNNR	-0,863	-0,699	-0,501	0,099	1,033	2,679
SVR	-1,021	-0,601	-0,232	-0,147	0,309	1,023

### Medidas de calidad de los modelos

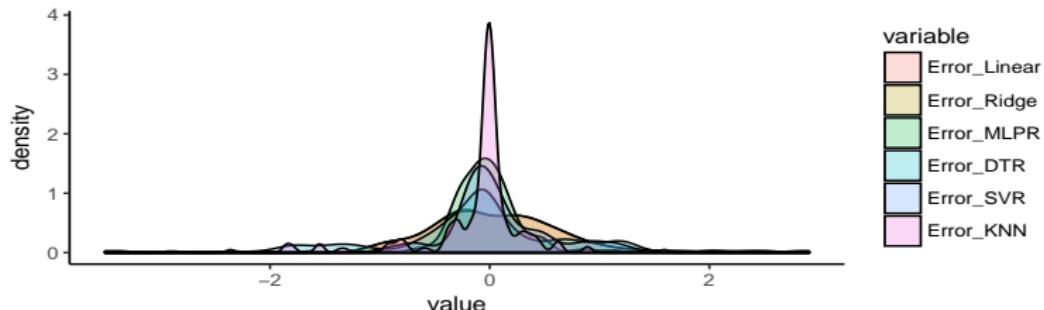
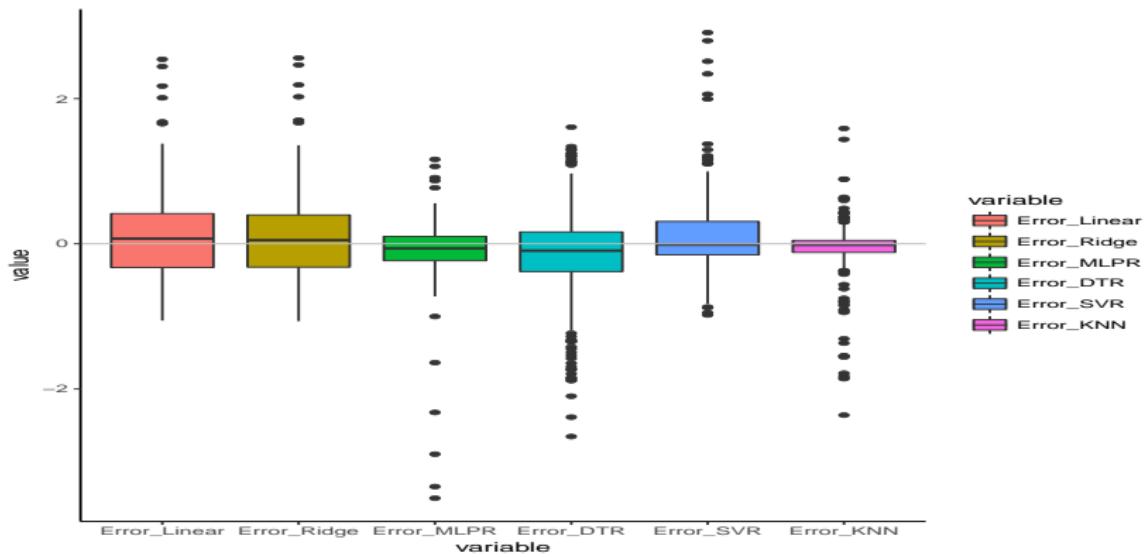
	Corr11	MSE	Media en VC	SD en VC	CorrL20	MSEL20
Lineal	0,774	0,624	1,108	0,278	0,890	0,580
Ridge	0,775	0,621	1,072	0,277	0,896	0,566
SVR	0,837	0,613	<b>0,834</b>	0,490	<b>0,967</b>	<b>0,464</b>
MLP	0,875	0,528	1,086	0,288	0,727	1,023
KNN	<b>0,888</b>	<b>0,494</b>	0,981	0,362	0,797	0,936
DTR	0,679	0,768	1,148	0,544	0,532	1,131

# Evaluación y análisis de los modelos generados

Scatterplot de datos observados vs datos predichos



# Análisis de errores



## *Problemas de un sistema regional de modelado de poblaciones de mosquito*

# Problemas de un sistema regional

- Escases de muestras por cada punto (ciudad)

# Problemas de un sistema regional

- Escases de muestras por cada punto (ciudad)
  - Entrenamiento de modelos que generalicen

# Problemas de un sistema regional

- Escases de muestras por cada punto (ciudad)
  - Entrenamiento de modelos que generalicen
- Escases de puntos (ciudades) con datos

# Problemas de un sistema regional

- Escases de muestras por cada punto (ciudad)
  - Entrenamiento de modelos que generalicen
- Escases de puntos (ciudades) con datos
  - Tener buena capacidad de predicción en otras ciudades

# *Generalización espacial de modelos epidemiológicos*

# Presentación del problema

- Tenemos  $N$  modelos:  $M_1, M_2, \dots, M_N$

# Presentación del problema

- Tenemos  $N$  modelos:  $M_1, M_2, \dots, M_N$
- Queremos saber la abundancia en una nueva ciudad

# Presentación del problema

- Tenemos  $N$  modelos:  $M_1, M_2, \dots, M_N$
- Queremos saber la abundancia en una nueva ciudad

# Presentación del problema

- Tenemos  $N$  modelos:  $M_1, M_2, \dots, M_N$
- Queremos saber la abundancia en una nueva ciudad

Para cada  $Ciudad_j$  con  $j = 1, \dots, N$  sabemos que

$$ovip_j = \beta_j + \sum coef_{ji} \times envVar_i(j)$$

# Aproximación a la solución

## ***Primera Ley de la Geografía - Tobler, 1970***

*"Everything is related to everything else, but near things are more related than distant things"*

# Aproximación a la solución

- El modelo de la nueva ciudad ( $M_{new}$ ) es igual al de la ciudad más cercana espacialmente que posea un modelo entrenado:

$$M_{new} = M_i$$

# Aproximación a la solución

- El modelo de la nueva ciudad ( $M_{new}$ ) es igual al de la ciudad más cercana espacialmente que posea un modelo entrenado:

$$M_{new} = M_i$$

- $M_{new}$  tiene un aporte de similaridad de cada uno de los otros modelos teniendo en cuenta su cercanía espacial:

$$M_{new} = \sum \frac{M_j}{D_j}$$

# Distancia Ambiental Normalizada (NED)

Y si la relación de similaridad de los modelos tiene en cuenta las características ambientales?

# Distancia Ambiental Normalizada (NED)

Y si la relación de similaridad de los modelos tiene en cuenta las características ambientales?

- Definamos la *distancia ambiental* entre dos puntos geográficos como:

$$dist_{x_1 - x_2} = \sqrt{\sum (v_{i_1} - v_{i_2})^2}$$

donde  $v_i$  son las variables ambientales

# Distancia Ambiental Normalizada (NED)

Y si la relación de similaridad de los modelos tiene en cuenta las características ambientales?

- Definamos la *distancia ambiental* entre dos puntos geográficos como:

$$dist_{x_1 - x_2} = \sqrt{\sum (v_{i_1} - v_{i_2})^2}$$

donde  $v_i$  son las variables ambientales

- Luego, si queremos construir el modelo para la ciudad *new*:

$$M_{new} = \sum \frac{M_j}{L_j}$$

donde  $L_j = dist_{x_{new} - x_j}$

# Distancia Ambiental Normalizada (NED)

Y si la relación de similaridad de los modelos tiene en cuenta las características ambientales?

- Definamos la *distancia ambiental* entre dos puntos geográficos como:

$$dist_{x_1 - x_2} = \sqrt{\sum (v_{i_1} - v_{i_2})^2}$$

donde  $v_i$  son las variables ambientales

- Luego, si queremos construir el modelo para la ciudad *new*:

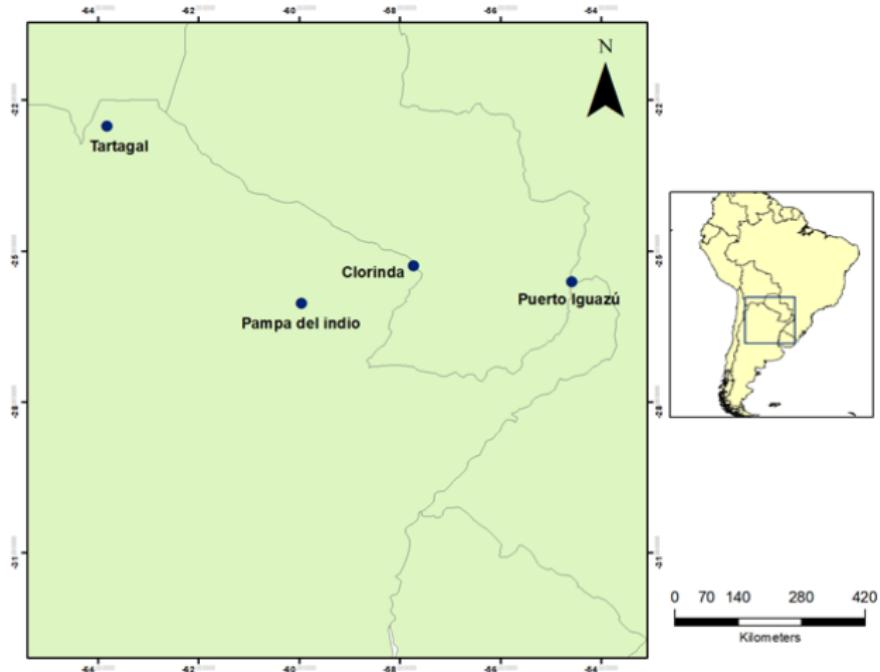
$$M_{new} = \sum \frac{M_j}{L_j}$$

donde  $L_j = dist_{x_{new} - x_j}$

- A su vez, entendemos por distancia *normalizada* a que:

$$1 = \sum \frac{1}{L_j}$$

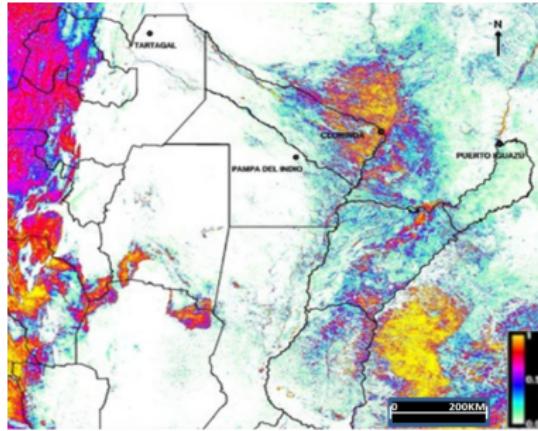
# Evaluación de la propuesta



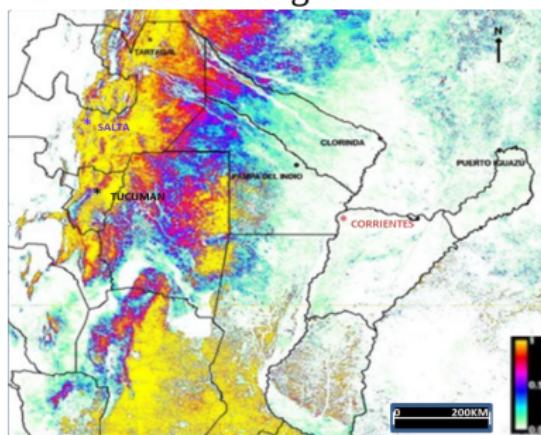
## Evaluación de la propuesta

Distancia Ambiental Normalizada a las cuatro ciudades con datos

Clorinda

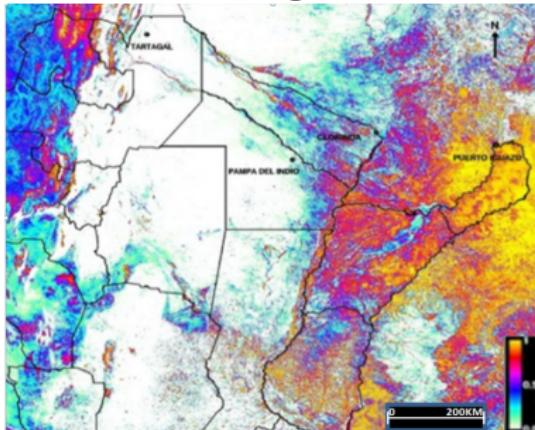


Tartagal

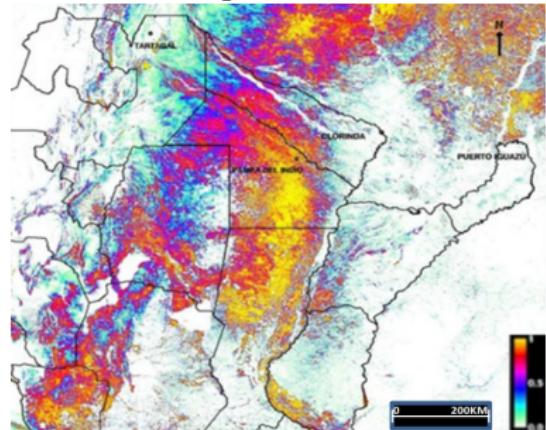


# Evaluación de la propuesta

Puerto Iguazú



Pampa del Indio



# Evaluación de la propuesta

Supongamos que queremos obtener los modelos para Salta, Tucumán y Corrientes...

# Evaluación de la propuesta

Inversa de NED a las 4 ciudades con datos

	Puerto Iguazú	Clorinda	Pampa del Indio	Tartagal
Tucuman	0,197	0,011	0,388	0,402
Corrientes	0,491	0,466	0,039	0,002
Salta	0,112	0,005	0,133	0,749

# Evaluación de la propuesta

Inversa de NED a las 4 ciudades con datos

	Puerto Iguazú	Clorinda	Pampa del Indio	Tartagal
Tucuman	0,197	0,011	0,388	0,402
Corrientes	0,491	0,466	0,039	0,002
Salta	0,112	0,005	0,133	0,749

Notar que la inversa de la NED ( $\frac{1}{NED}$ ) nos dice qué tan similar es una ciudad en comparación con las otras tres

*Todo muy lindo... pero, en resumen, qué hicimos?*

# En conclusión...

- Se generó un *framework* que facilita el entrenamiento y validación de modelos de aprendizaje automático

## En conclusión...

- Se generó un *framework* que facilita el entrenamiento y validación de modelos de aprendizaje automático
- Se generaron modelos con mejor desempeño a la hora de predecir oviposición que los que se utilizan actualmente

## En conclusión...

- Se generó un *framework* que facilita el entrenamiento y validación de modelos de aprendizaje automático
- Se generaron modelos con mejor desempeño a la hora de predecir oviposición que los que se utilizan actualmente
- Se propone una solución novedosa al problema de escases de puntos geográficos que poseen mediciones de campo

## *Algunos datos de color*

# Datos de color!

- Estos trabajos dieron lugar a tres publicaciones indexadas:

# Datos de color!

- Estos trabajos dieron lugar a tres publicaciones indexadas:
  - J. M. Scavuzzo, F. Trucco, M. Espinosa, C. B. Tauro, M. Abril, C. M. Scavuzzo, and A. C. Frery. Modeling dengue vector population using remotely sensed data and machine learning. *Acta tropica*, 185:167–175, 2018.

# Datos de color!

- Estos trabajos dieron lugar a tres publicaciones indexadas:
  - J. M. Scavuzzo, F. Trucco, M. Espinosa, C. B. Tauro, M. Abril, C. M. Scavuzzo, and A. C. Frery. Modeling dengue vector population using remotely sensed data and machine learning. *Acta tropica*, 185:167–175, 2018.
  - J. Scavuzzo, M. Espinosa, E. Di Fino, M. Abril, G. Peralta, and C. Scavuzzo. Generalización espacial de modelos epidemiológicos basada en el concepto de distancia ambiental normalizada ned. 2018.

# Datos de color!

- Estos trabajos dieron lugar a tres publicaciones indexadas:
  - J. M. Scavuzzo, F. Trucco, M. Espinosa, C. B. Tauro, M. Abril, C. M. Scavuzzo, and A. C. Frery. Modeling dengue vector population using remotely sensed data and machine learning. *Acta tropica*, 185:167–175, 2018.
  - J. Scavuzzo, M. Espinosa, E. Di Fino, M. Abril, G. Peralta, and C. Scavuzzo. Generalización espacial de modelos epidemiológicos basada en el concepto de distancia ambiental normalizada ned. 2018.
  - J. Scavuzzo, F. Trucco, C. Tauro, A. German, M. Espinosa, and M. Abril. Modeling the temporal pattern of dengue, chicungunya and zika vector using satellite data and neural networks. volume 2017-January, pages 1–6, 2017

# Datos de color!

Si utilizamos la base de datos de artículos indexados **Scopus**, y buscamos bajo las palabras clave *remote sensing, machine learning, dengue...*

The screenshot shows the Scopus search interface with the following details:

**Search Query:** TITLE-ABS-KEY (remote AND sensing AND machine AND learning AND dengue)

**Document Results:** 2 document results

**Refine results:**

- Access type: Other (2)
- Year: 2018 (1), 2017 (1)
- Author name: Abril, M. (2), Espinosa, M. (2), Scavuzzo, J.M. (2), Tauro, C.B. (2), Frery, A.C. (1)

**Analyze search results:**

Sort on: Cited by (highest)

Document title	Authors	Year	Source	Cited by
Modeling Dengue vector population using remotely sensed data and machine learning	Scavuzzo, J.M., Trucco, F., Espinosa, M., (...), Scavuzzo, C.M., Frery, A.C.	2018	Acta Tropica 185, pp. 167-175	0
Modeling the temporal pattern of Dengue, Chikungunya and Zika vector using satellite data and neural networks	Scavuzzo, J.M., Trucco, F.C., Tauro, C.B., (...), Espinosa, M., Abril, M.	2017	2017 17th Workshop on Information Processing and Control, RPIC 2017 2017-January, pp. 1-6	0

Display: 20 results per page

*Gracias!  
Preguntas?*

# *Agradecimientos*