



TALLERES COMPLEJIDAD EN SALUD Y ECOLOGÍA: RIESGOS Y DATOS

FACTORES





Ecología y Medio Ambiente

11 de septiembre 2023



PROYECTO 42



¿Por qué estamos aquí?



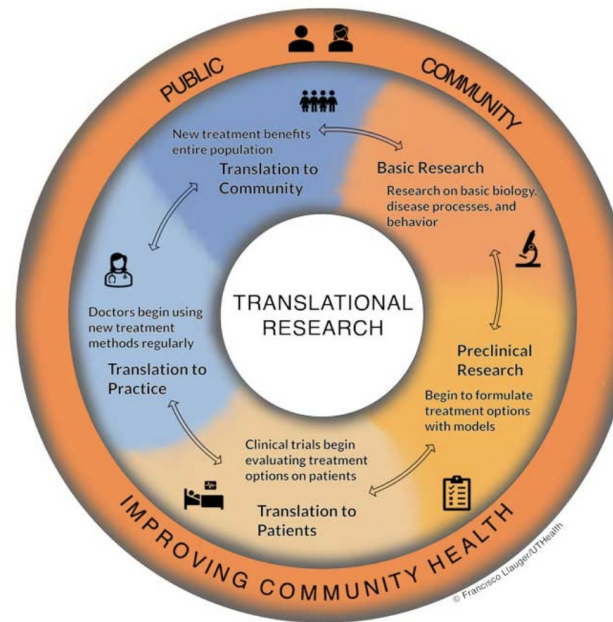
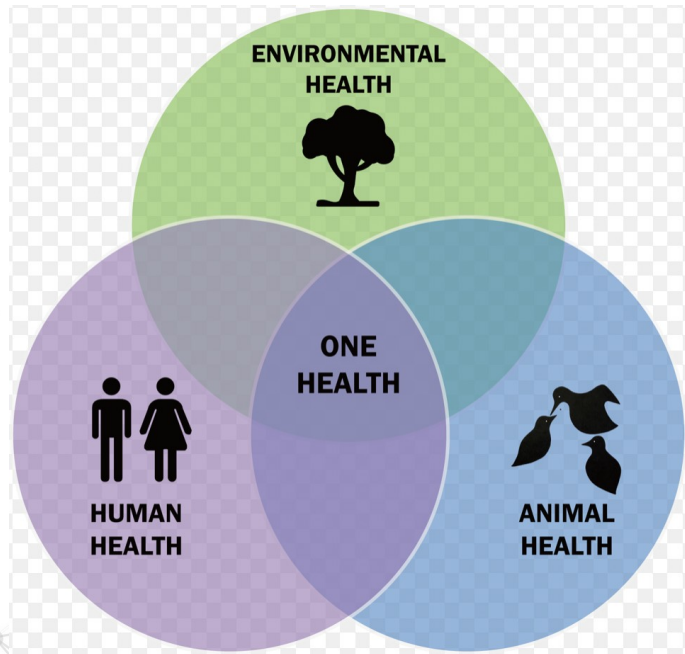
¡Son muy costosos!

¿Qué tienen en común?

- In its World Obesity Atlas 2023, the federation said the economic impact of overweight and obesity on the world is set to reach \$4.32tn—nearly 3% of global gross domestic product—annually by 2035. This is comparable with the impact of covid-19 in 2020.
- WASHINGTON, Jan 20 (Reuters) - The International Monetary Fund expects to raise its forecast that the COVID-19 pandemic will cost the global economy \$12.5 trillion through 2024, the head of the global lender said on Thursday.
- Destroying Biodiversity May Cost the World \$2.7 Trillion a Year: World Bank says existing economic models are too optimistic; Bank warns on risks to food from fisheries, forestry timber
- News 23 May 2022 DAVOS, SWITZERLAND Deloitte research reveals inaction on climate change could cost the world's economy US\$178 trillion by 2070

¿Qué estamos haciendo al respecto?

Intentando desarrollar marcos más adecuados



¿Qué más tienen en común?

- Los gobiernos no han podido solucionarlos
- Las sociedades no han podido solucionarlos
- La ciencia no ha podido solucionarlos

¿Por qué?

¡Son “Sistemas Complejos Adaptativos”!

- Multifactorial
- Múltiples causas directas y indirectas
- Multiescala
- Multidisciplinaria



Big car versus small car?

¡Nosotros somos la "causa"!



No Mask

Masks



Stairs versus escalator?

Hamburger versus salad?

Netflix or gym?

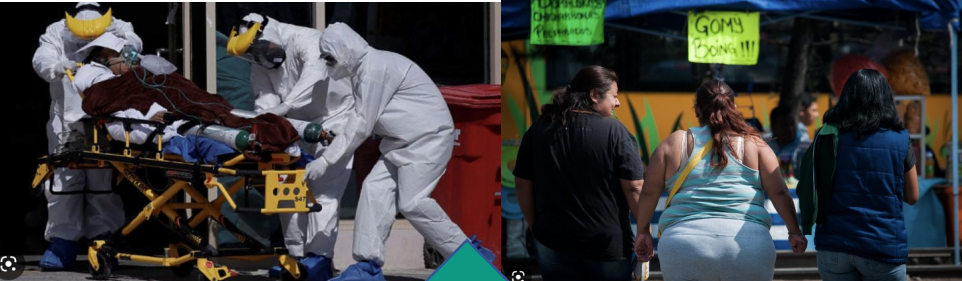
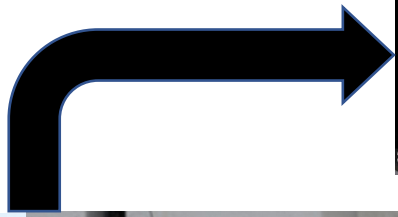
Son consecuencias de nuestras decisiones



Resist/don't resist?

Invade/don't invade?

¿Cómo tomamos nuestras decisiones?



¿Dónde habrá más pérdida de biodiversidad debido al cambio climático?
 ¿Quién está en más riesgo de infectarse por T.cruzi?
 ¿Qué es más importante en determinar la distribución de un especie particular?
 ¿Cuándo hay más riesgo de un aumento en casos de SARS-Cov-2?



López-Gatell, AMLO say no to enforcing use of coronavirus face masks

MND Staff October 28, 2020



The effectiveness of face masks is 'overstated', says Mexico's coronavirus czar.

Consecuencias
Objetivos y metas estratégicos



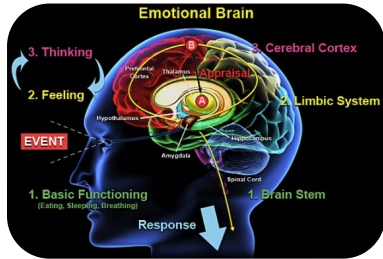
**NO HAY
 DECISION SIN
 PREDICCIÓN**

$$P(C|X)$$

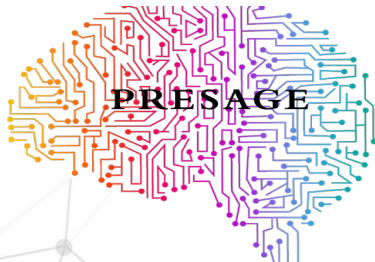
El Santo Grial para solucionar los problemas de los Sistemas Complejos Adaptativos

Potencia de procesamiento simbólico limitada ~ 30 bytes/seg

- No hablamos "base de datos"



+



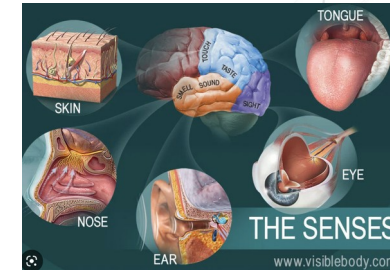
Las computadoras no "entienden" nada

- Produce máquinas - especialistas

¿Qué quiero predecir?

¿Con qué lo quiero predecir?

$$P(C|\mathbf{X}(t))$$



+



¿Qué modelo predictivo se usará?

LA PROBABILIDAD DE C DADO $\mathbf{X}(t) = (X_1, X_2, \dots, X_N)$

Nuestra predicción es una *inferencia estadística*

Tenemos un modelo de predicción
para cada una de estas situaciones.
¡Y muchas más!

$$P(C|X(t))$$

C = aguilas

X = mi moneda

X = todas nuestras monedas

$$P(C) = 0.5$$

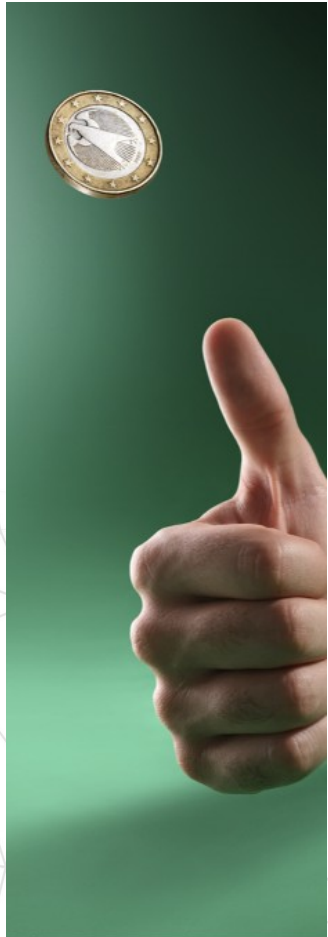
C = gato aterriza patas abajo

X = mi gato

X = gatos en general

$$P(C) \sim 1.0$$

Bayesian priors
Nuestros prejuicios
Nuestras creencias



¿Cómo se evalúan las predicciones?

- Se compone un “ensamble estadístico”
 - P.ej. se puede lanzar mi moneda N veces o N monedas una vez – representan diferentes comparaciones
 - $P(C)$ se convierte a una hipótesis nula. Se compara $P(\text{aguila} \mid \text{mi moneda})$, como hipótesis alternativa, con $P(\text{aguila}) = 0.5$ para ver como tan probable es las lanzas de la moneda representan monedas justas
 - Se calcula $P(\text{aguila} \mid \text{mi moneda})$ por lanzar la moneda N veces y ver con que frecuencia salio aguila. Se puede usar prueba de hipótesis para ver como tan probable era para observar una discrepancia entre $P(\text{aguila} \mid \text{mi moneda})$ y 0.5 por casualidad
 - Para el gato se puede formar un ensamble por dejar caer un gato N veces o N gatos una vez y comparar con la hipótesis nula que $P(\text{gato aterriza patas abajo}) = 1$, o podemos creerlo hasta vemos evidencia al contrario. Hay ensambles que son difíciles armar.
 - ¿Qué evidencia se necesita y cómo se incorpora en nuestro modelo?

¿Cómo ajustamos nuestras expectativas?

Aprendemos... esperamos

- $P(C) \square P(C|\mathbf{X}) = P(\mathbf{X}|C)P(C)/P(\mathbf{X}) \square P(C|\mathbf{X},\mathbf{X}') = P(\mathbf{X}'|C,\mathbf{X})P(C|\mathbf{X})/P(\mathbf{X}'|\mathbf{X}) \square \dots$
- “Corregimos” nuestro estimado inicial (prior) debido a la información \mathbf{X} a través de la Teorema de Bayes
 - $P(\mathbf{X}|C)$ es la verosimilitud (likelihood) que se verá los datos \mathbf{X} en los casos C y $P(\mathbf{X})$ es la evidencia
 - Llegamos a un nuevo estimado (el posterior) que toma en cuenta la nueva información que sirve como un nuevo prior
- “Corregimos” nuestro nuevo prior debido a la información \mathbf{X}' a través de la Teorema de Bayes de nuevo
 - $P(\mathbf{X}'|C,\mathbf{X})$ es el nuevo verosimilitud (likelihood) que se verá los nuevos datos \mathbf{X}' en los casos que tienen tanto C como \mathbf{X} y $P(\mathbf{X}'|\mathbf{X})$ es la evidencia acondicionado en la información inicial \mathbf{X}
 - Llegamos a un nuevo estimado (el posterior) que toma en cuenta la nueva información \mathbf{X}' que sirve como un nuevo prior etc etc

¿Qué X podemos/debemos incluir?

- ¿Un sitio web que dice que la teoría que los gatos aterrizan patas abajo es una conspiración neo-liberal/comunista/extraterrestre/...?
- ¿Un sitio web www.funanimalfacts.org que dice que los gatos no aterrizan patas abajo más que cualquier otro animal?
- Un estudio que apareció en Scientific American <https://www.scientificamerican.com/article/why-do-cats-land-on-their-feet-physics-explains/>
- Aunque esto no suena serio, es muy serio. La ciencia misma esta llena de prejuicios y sesgos tanto explícitos como implícitos. La disciplinariedad misma es un ejemplo de esto. Ni hablar de las autoridades y los individuos.

Ejemplos de cosas que se quiere predecir

1. ¿Es probable que se encontrará un lince en este lugar?
 C_1
2. ¿Es probable que *Artibeus jamaicensis* (un murciélago) es un hospedero importante del patógeno *Leishmania mexicana*? C_2
3. ¿Habrá un aumento en casos de COVID19 en este municipio en el próximo mes? C_3
4. ¿Es probable que esta persona fallecerá de COVID19?
 C_4
5. ¿Es probable que esta persona padece de hipoglucemia? C_5

¿Es probable que se encontrará un lince en “este lugar”? C_1

- $P(C_1 = \text{encontrar el lince} | \mathbf{X} = \text{este lugar})$
 - El problema con la Inteligencia Humana es que estimamos $P(C|\mathbf{X})$ sin saber que factores \mathbf{X} estamos metiendo y sin saber que es $P(\cdot | \cdot)$ - nuestro cerebro es “caja negra”
 - “Experiencia” (p.ej. Mastozoólogo vs. sociólogo) nos permite hacer una selección (sesgada hasta muy sesgada) de cuáles X_i meter
- ¿Cómo validamos la predicción si “este lugar” es único? Ensemble con $N = 1$.
- Se puede intentar formar un ensamble de “este lugar”es. ¿Cómo?
 - ¿Cuadrantes de 1m x 1m? ¿1km x 1km? ¿1000 km x 1000 km?...
 - Se puede contar en cuántos cuadrantes se encuentra (o se ha encontrado) un lince $\square P(C_1) = N(C_1)/N$
 - Pero así, todos los lugares son iguales. Yo quiero saber sobre “este lugar”.

¿Cómo se caracteriza “este lugar”?

- Imagina que hacemos la hipótesis que lo que distingue “este lugar”, donde se predice que es alta probabilidad de encontrar el lince, es X_i = el promedio anual de su temperatura
- Se puede ir a los N cuadrantes y contar $N(C_1 X_i) = N(\text{lince encontrado y promedio anual de temperature} = T)$ versus $N(\text{promedio anual de temperature} = T) = N(X_i)$
- $P(C_1 | X_i) = N(C_1 X_i) / N(X_i)$ ¡Así de simple! ¿o no?

¿Cómo se calcula $P(C|X)$?

- Dos problemas:

1. ¿Qué es la probabilidad de encontrar el lince si $T = 23.4523$ °C vs 23.4524 °C? $N(C_1 | X_i) = 0, 1$

2. ¿Se imagina que la presencia del lince únicamente depende de la temperatura annual? ¿Qué otras variables se debería meter?

- ¿Precipitación?

- ¿Promedio? ¿Varianza? ¿Mínimo? ¿Máximo?...

- ¿Diario? ¿Semanal? ¿Anual?...

- ¿Otras funciones de temperatura?

- ¿Promedio? ¿Varianza? ¿Mínimo? ¿Máximo?...

- ¿Diario? ¿Semanal? ¿Anual?...

- ¿Otros?

- ¿Sus presas? ¿Sus competidores? ¿Los alimentos de sus presas? ¿Los parásitos de sus presas?

- ¿Cuáles son sus presas, competidores etc.?

- ¿Tipo de suelo?

- ¿Factores humanos?

¿Cómo se calcula $P(C|\mathbf{X})$?

- Para una inferencia estadística de $P(C|X_i)$, y para distinguirla de la hipótesis nula $P(C)$, se requiere un ensamble ($N(X_i)$) de tamaño suficiente que permite distinguirlos a la confianza deseada. Esto requiere hacer un “coarse graining” (dividir en cubetas) de X_i
- $P(C_1|\mathbf{X}) = P(\text{encontrar el lince} | \text{temperatura anual} = T, \text{ precipitación máxima últimos tres meses}, \dots, \text{ canis latrans} = \text{ presente}, \text{ romeralagus diazi} = \text{ presente}, \dots, \text{ población humana} < 1000, \dots) = \frac{N(\text{encontrar el lince}, \text{ temperatura anual} = T, \text{ precipitación máxima últimos tres meses}, \dots, \text{ canis latrans} = \text{ presente}, \text{ romeralagus diazi} = \text{ presente}, \dots, \text{ población humana} < 1000, \dots)}{N(\text{temperatura anual} = T, \text{ precipitación máxima últimos tres meses}, \dots, \text{ canis latrans} = \text{ presente}, \text{ romeralagus diazi} = \text{ presente}, \dots, \text{ población humana} < 1000, \dots)} = 0, 1$
- NO se puede calcular $P(C_1|\mathbf{X})$ por puro conteo. Se requiere modelos de Aprendizaje de máquina/Inteligencia Artificial

¿Qué se quiere del modelo de $P(C|X)$?

- ¿Que es predictivo?
 - Se puede juzgar objetivamente como tan bueno es el modelo a través de muchas métricas estandares de desempeño
 - Se puede incluir y excluir variables X_i y juzgar si se aumentó/disminuyó el desempeño
- ¿Qué explica?
 - ¿Cuándo importa? Hay problemas donde los “por qué” no importan (p.ej. Traducción automática) versus otros donde los “por qué” si importan mucho (p.ej. diagnóstico médico). ChatGPT versus las plataformas de Chilam. Caja negra versus caja blanca.
- ¿Qué representa múltiples realidades?
 - La realidad actual es representado por cada elemento del ensamble c siendo representado por un estado $X(c)$. Con el modelo $P(C|X)$ se puede crear cualquier escenario de “que-si” para calcular $P(C|X'(c))$
- Predictabilidad es objetivo y explicabilidad es subjetivo (p.ej. muy dependiente de tu disciplina)
- Para optimizar la predictabilidad y crear nuevas realidades se requiere la Inteligencia Artificial y para optimizar la explicabilidad se requiere la Inteligencia Humana

¿Qué se quiere de la toma de decisión usando el modelo de $P(C|\mathbf{X})$?

- Si C representa algo bueno/malo se quiere aumentar/disminuir $P(C|\mathbf{X}) = P(C|X_1, X_2, \dots, X_N)$
- Se cambia $P(C|X_1, X_2, \dots, X_N)$ por cambiar uno o más X_i
- Estos cambios pueden ser debido a la dinámica actual del sistema – p.ej. Cambio climático – o por una intervención humana – p.ej. Un programa de reforestación.
- Hay demasiado posibles intervenciones para probarlas uno por uno y ver si ayudaron o no
- Se requiere saber el grado de causalidad entre C y X_i – X_i es directamente causal, indirectamente causal, pura correlación.
- Se requiere saber si X_i es “actionable” – es algo que tu intervención puede cambiar. P.ej. Temperatura anual no es actionable (en el corto plazo) pero uso de suelo si es
- $P(C|X_1, X_2, \dots, X_N)$ puede representar un enorme número de posibles mundos/realidades además de lo nuestro

Resumen

1. Cada problema de mayor grado que se enfrenta la humanidad se debe a una mala toma de decisión que tiene consecuencias adversas
2. Cada decisión requiere una predicción - predicción es el fundamento de la vida, la ciencia, la medicina,..., todo.
3. Únicamente los Sistemas Complejos Adaptativos toman decisiones
4. Para predecir SCA se requiere modelos de inferencia estadística $P(C|\mathbf{X})$ que: i) Son “complejos” - multifactoriales - y “adaptativos” - que las relaciones entre los C (“efectos”) y X_i (“causas”) pueden cambiar en tiempo.
5. El marco Bayesiano es lo más adecuado para modelar los SCA y calcular $P(C|\mathbf{X})$ - modela como los seres humanos (seres vivos) construimos modelos de predicción.
 - Incorpora el aprendizaje naturalmente.
 - Puede incorporar tanto “creencias” como datos

Resumen

6. La multi-factorialidad es clave. La disciplinarietà y el reduccionismo son los obstáculos principales porque la ciencia no ha avanzado más en solucionar o mitigar estos grandes problemas
7. La multifactorialidad implica la necesidad de BIG, DEEP data para representar en datos
8. Hoy en día estos datos son digitales y nosotros no hablamos base de datos
9. Así para calcular $P(C|\mathbf{X})$ se requiere algoritmos de ML/AI
10. Hay diferentes tipos de algoritmos, cada uno con sus pros y cons. Ninguno puede substituir la necesidad de la Inteligencia Humana para la interpretación de $P(C|\mathbf{X})$ y su uso en la toma de decisión.

Demanda de los grandes problemas nacionales e internacionales

1. La integración de grandes bases de datos que representan la riqueza (multifactorialidad/multidisciplinariedad) de estos problemas
2. Algoritmos de ML/AI que pueden convertir estos datos a modelos predictivos $P(C|\mathbf{X})$ que pueden ser tanto motores en los Ciclos de Decisión como objetos de investigación y de docencia
3. Plataformas abiertas disponibles a diferentes perfiles de usuario – tomadores de decisión, investigadores, estudiantes, público en lo general - sin que sean expertos en ML/AI



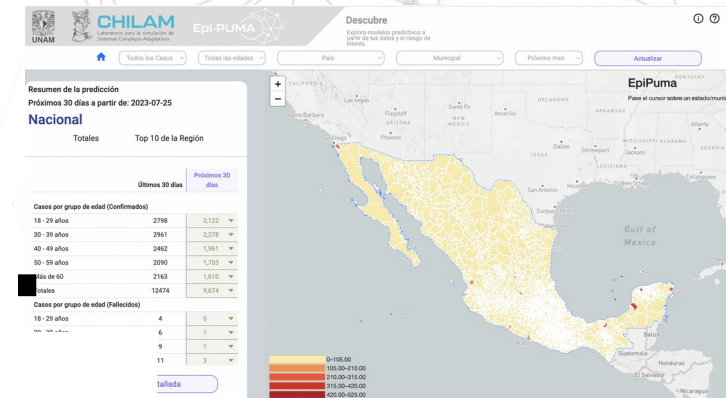
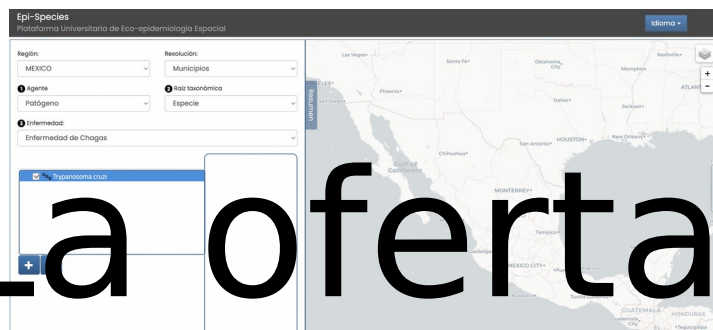
Project42



La oferta...

CHILAM

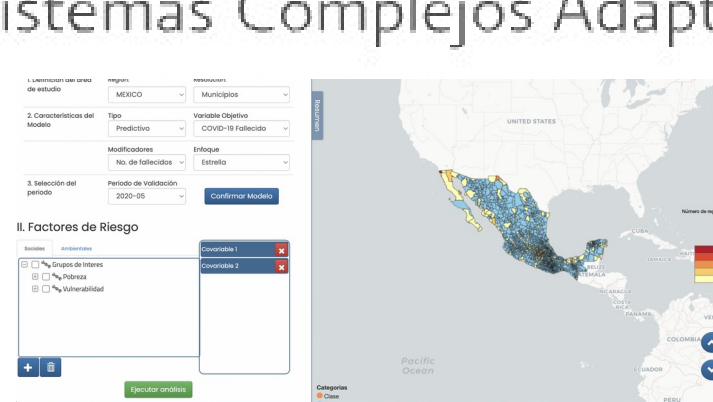
Laboratorio para la simulación de
Sistemas Complejos Adaptativos



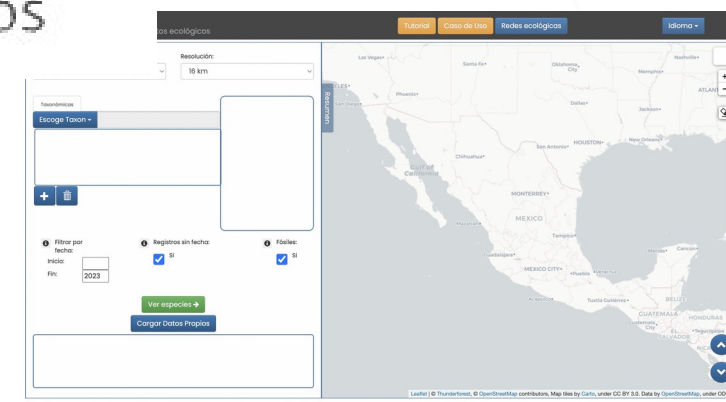
Epi-PUMA 2.0



Epi-PUMA 2.0



Epi-PUMA 1.0



SPECIES