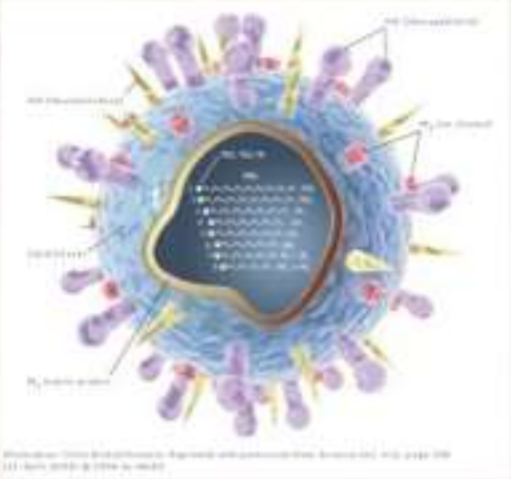
Predicting the “Why”, “Where” and “When” of Zoonoses from Co-occurrence Data

Chris Stephens
C3-Centro de Ciencias de la Complejidad y Instituto de Ciencias Nucleares, UNAM
Cary Institute 19/11/2020

Why do things happen?

Because of interactions;

and study of interactions among

interactions

Type of interaction	Sign	Effects
mutualism	+/+	both species benefit from interaction
commensalism	+/0	one species benefits, one unaffected
competition	-/-	each species affected negatively
predation, parasitism, herbivory	+/-	one species benefits, one is disadvantaged



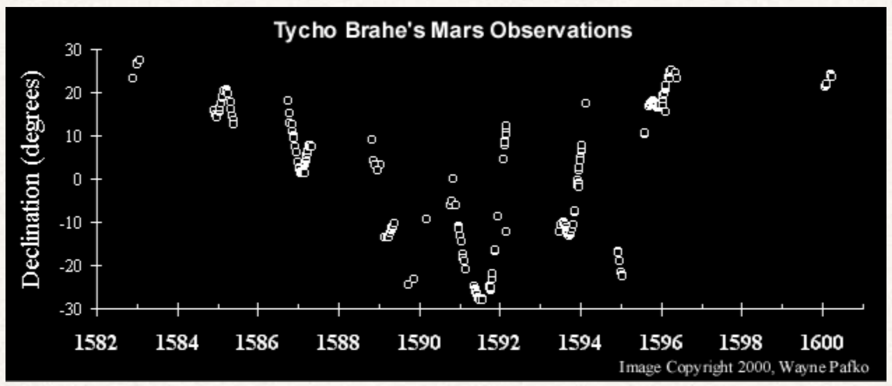


many types of interaction are
any interactions can we
SCIENCEphotoLIBRARY
gettyimages
Importancia médica
T. Infestans
T. barberi
T. pallidipennis
T. longipenis
T. recurva
T. neotomae
Culex quinquefasciatus

Inferring Interactions from Spatial Data…

A famous historical antecedent

Data —> Phenomenology —> Taxonomy —> Theory —> Isn't all science data science?



Data → **Phenomenology**

Kepler's Laws

1. The orbit of a planet is an ellipse with the Sun at one of the two foci.
2. A line segment joining a planet and the Sun sweeps out equal areas during equal intervals of time.
3. The square of the orbital period of a planet is proportional to the cube of the semi-major axis of its orbit.

Interaction labels

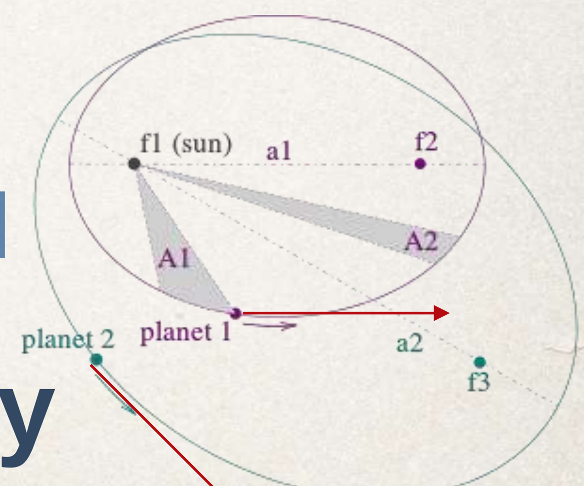
This is a "macro" interaction that emerges from a "micro" interaction that is the same

$F = ma$

$F = GMm/r^2$

Quantifying the interaction

Theory



Null hypothesis

Brahe's observations and Kepler's phenomenology can be explained by the existence of an interaction - gravity

"MICRO" interaction

Interaction labels:

Aepyceros melampus
Impala
Herbivore

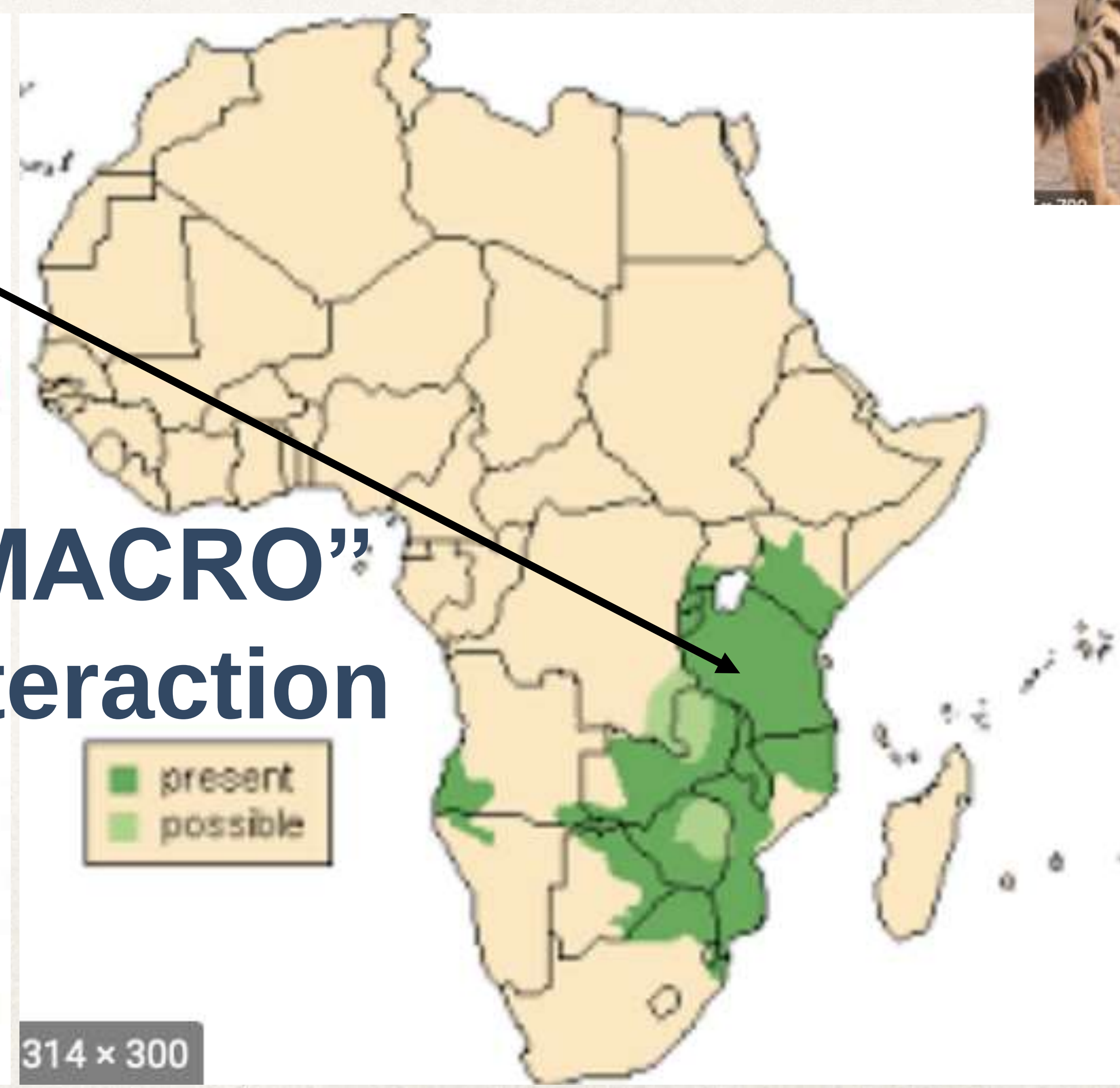
Acinonyx jubatus
Cheetah
Carnivore

Predation

Null hypothesis

This is a result of this… and many, many other interactions

"MACRO" interaction



What is an interaction?

Interactions make the state variables of agents different, compared to when the interaction is absent (Null hypothesis)

State variables - mass, charge, dead, alive, pregnant, satiated, mother, vector, host, pathogen,…, **x, y, z, t**

Agents - an electron, a planet, a carnivore, a cheetah, a tick, an Aedes aegyptii, a virion of Zika, an Artibeus literatus, a dog walker in the forest,…

How can we identify, categorise and quantify interactions?

Standard empirical approach:

- Observe an interaction - agents (Lutzomyia cruciata/Didelphis virginianus) and action (Lutzomyia takes blood meal from animal)
- Characterise it: Vector-host
- Concentrates on state variables that are labels: infected, blood gorged, dead, alive,…
- Micro-science/small data, time and human intensive

Inferential approach:

- Observe the spatio-temporal distributions (x, y, z, t) of agents and compare them to the Null hypothesis. If the distributions of the agents differ then we **DEFINE** that as an interaction.
- Use the other “labels” (electric charge, carnivore, male, mass, mother, Peromyscus yucatanicus, tick, vector, host, predator, prey,…) to categorise the interaction
- Use statistical analysis and machine learning-based modelling to quantify the interactions
- Micro-science/small data vs. **Macro-science/big data** - depends on the data

Ecological "Tycho Brahes"

Thanks to the DATA REVOLUTION we are now DROWNING in data

- Collection data
- Ecological niche data
- Ecological niche model data
- Socio-economic data
- Socio-demographic data
- Phenotypic data
- Vegetable and crop data
- Geographical data
- Medical and public health data…

Problems with spatial data:

Different sources

Different location, data base, access,…

Different data types

categorical, metric, continuous, discrete,…

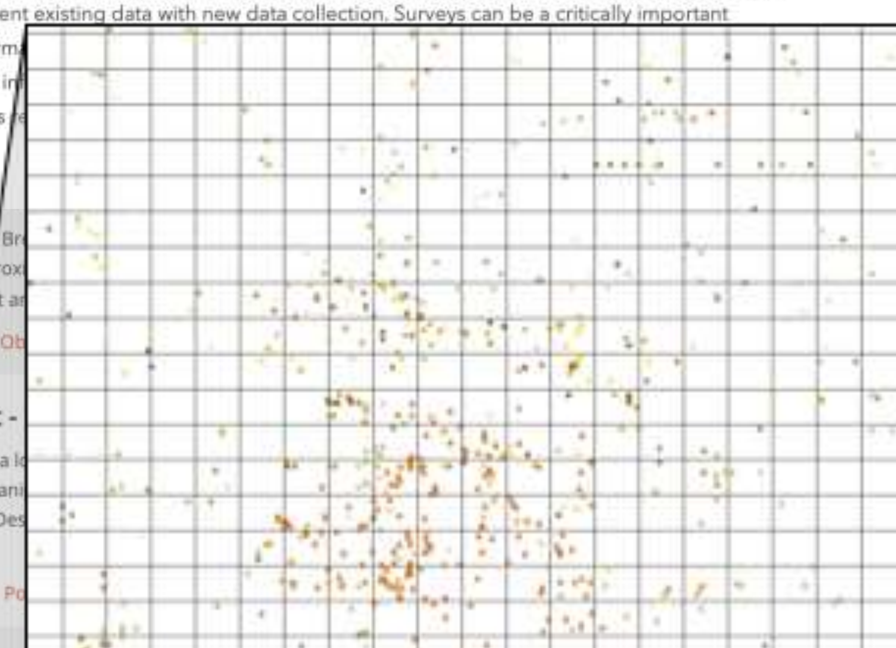
Different spatial resolution

Explicit – e.g., pixel by pixel in environmental layers

Implicit – 30,000,000 data points versus 30

"Quality" (e.g. Phenotypic characteristic) versus "quantity"

Abiotic versus biotic



The data are represented in space and time – spatial data mining

“Keplerian” Ecological models
Two agents

Example:

Take two agents: *Lutzomyia cruciata* C and *Artibeus literatus X* or *Lutzomyia cruciata* C and *Anual average temperature in the range 17-23 degrees C*

What’s the probability of finding the two agents together, or just one or neither? Both present. One present. Both absent/not present.

P(C, X) = P(C|X)P(C) = P(X|C)P(X)

P(X) and P(C) are the Null hypotheses for X and C

D(C,X) = (P(C|X) - P(C)) = (P(C, X) - P(C)P(X))/P(X)

is a measure of the interaction between C and X. This has a natural “unit” - the standard deviation of the binomial distribution

$$\varepsilon(C|X) = \frac{N_X(P(C|X) - P(C))}{\sqrt{(N_X P(C)(1 - P(C)))}}$$

is a measure of the statistical significance of the deviation of the co-distribution of C and X from the Null hypothesis. When the binomial can be approximated by the normal then $|\varepsilon(C|X)| > 1.96$ corresponds to the 95% confidence interval

“Keplerian” Ecological models

(N+M) agents

What do we want to predict?
C = (C1, C2, C3, …, CN)
the presence, or abundance, or,…
of one or more populations or taxa,
disease cases,…

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

What affects it?
The “**niche**”
X = (X1, X2, X3, …, XM)

A large part of the complexity is in the multi-factoriality of both C and X. Adaptation is inherent in the fact that P(C|X) can change in time.

This is a result of “all” interactions

$S(\mathbf{C}|\mathbf{X})$
Risk score

$$\mathbf{X} = X(sd)+X(se)+X(n)+X(ev)+X(g)+X(af)+X(hm)+X(i)+X(sp)+...$$

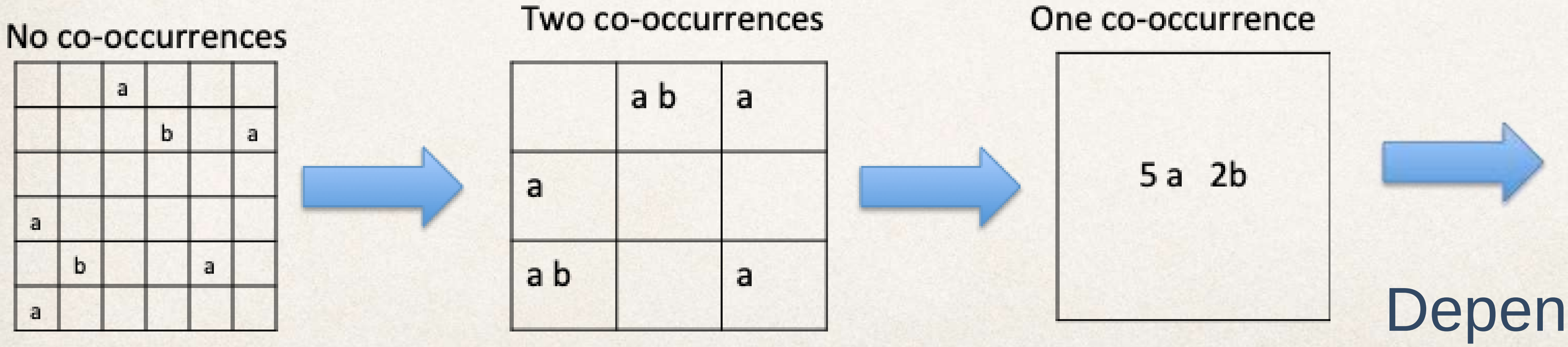
Macro-Climactic factors
Behavioural characteristics
Micro-Climatic factors
Phenotypic characteristics
Hydrography
Competitor species
Prey species
Predator species
Human activity

Problems of co-dependence and causality - of course!
Use Naive Bayes or Generalized Bayes approximation to calculate it

How do we calculate probabilities?
We count! But what do we count?

In standard data mining, for example: P(death|age) = N(death,age)/N(age); P(death|diabetes); P(death|age,diabetes); to **infer** that age is a risk factor for death, as is diabetes. Here, we count individuals who have different traits (labels). There is a preferred statistical unit - the individual, within which we can look for **coincidences/co-occurrences**. In spatial data mining this is not the case.

We must define **coincidences/co-occurrences** using an appropriate partition of our space (and time)



N(ab)/N(b)
= P(a|b)

Dependence of species a on niche variable b

Here we're in geographic space

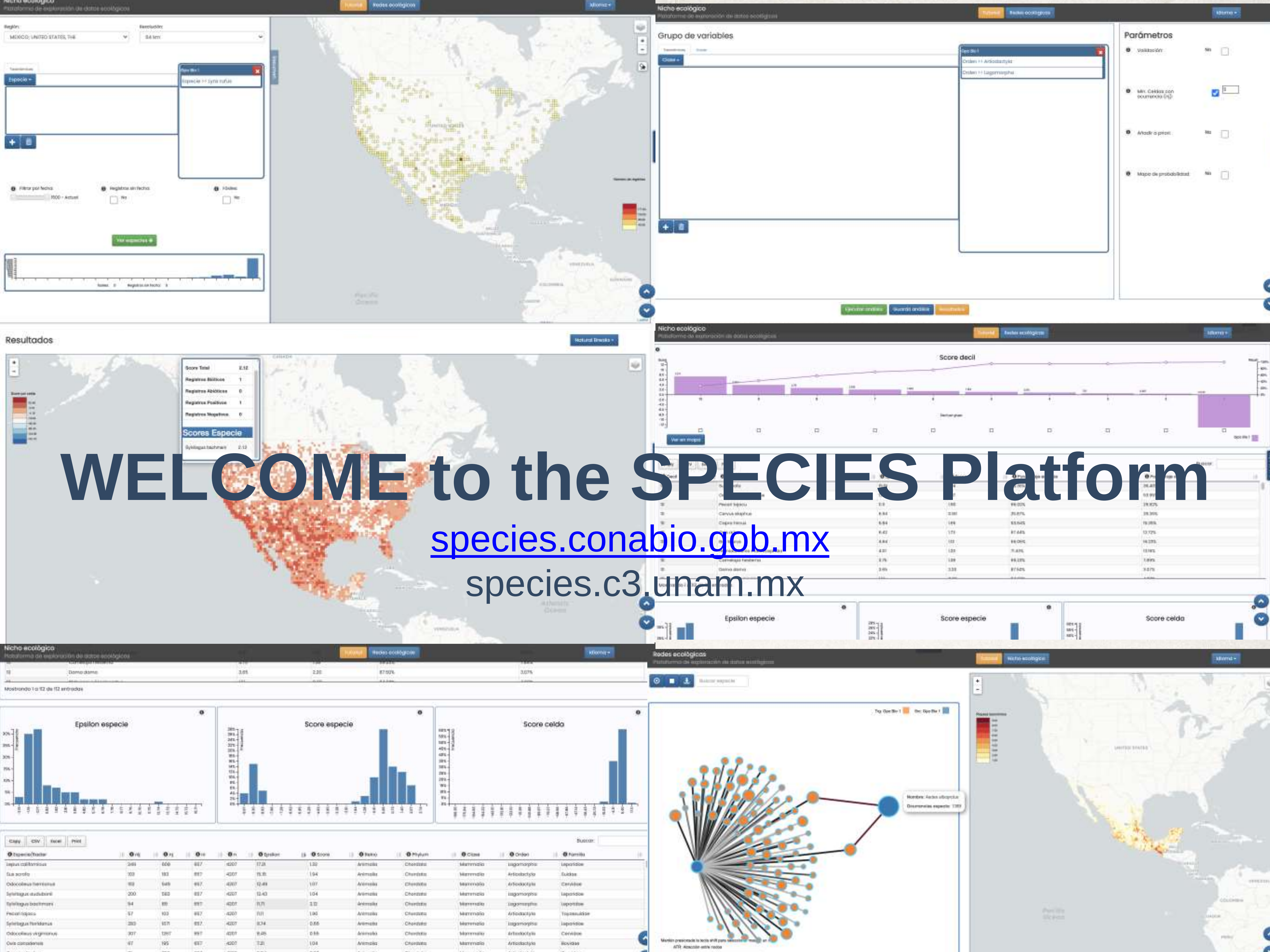
- Can we have a system that identifies, categorises and quantifies potential interactions and also constructs ecological niches and species distributions?

- Can we have a system that does that across hundreds of thousands of var

YES!

- Can we have a system that generates machine learning-based predictive

- Can we have it available open access Platform-as-a-Service?



WELCOME to the SPECIES Platform

species.conabio.gob.mx

species.c3.unam.mx

Some representative results

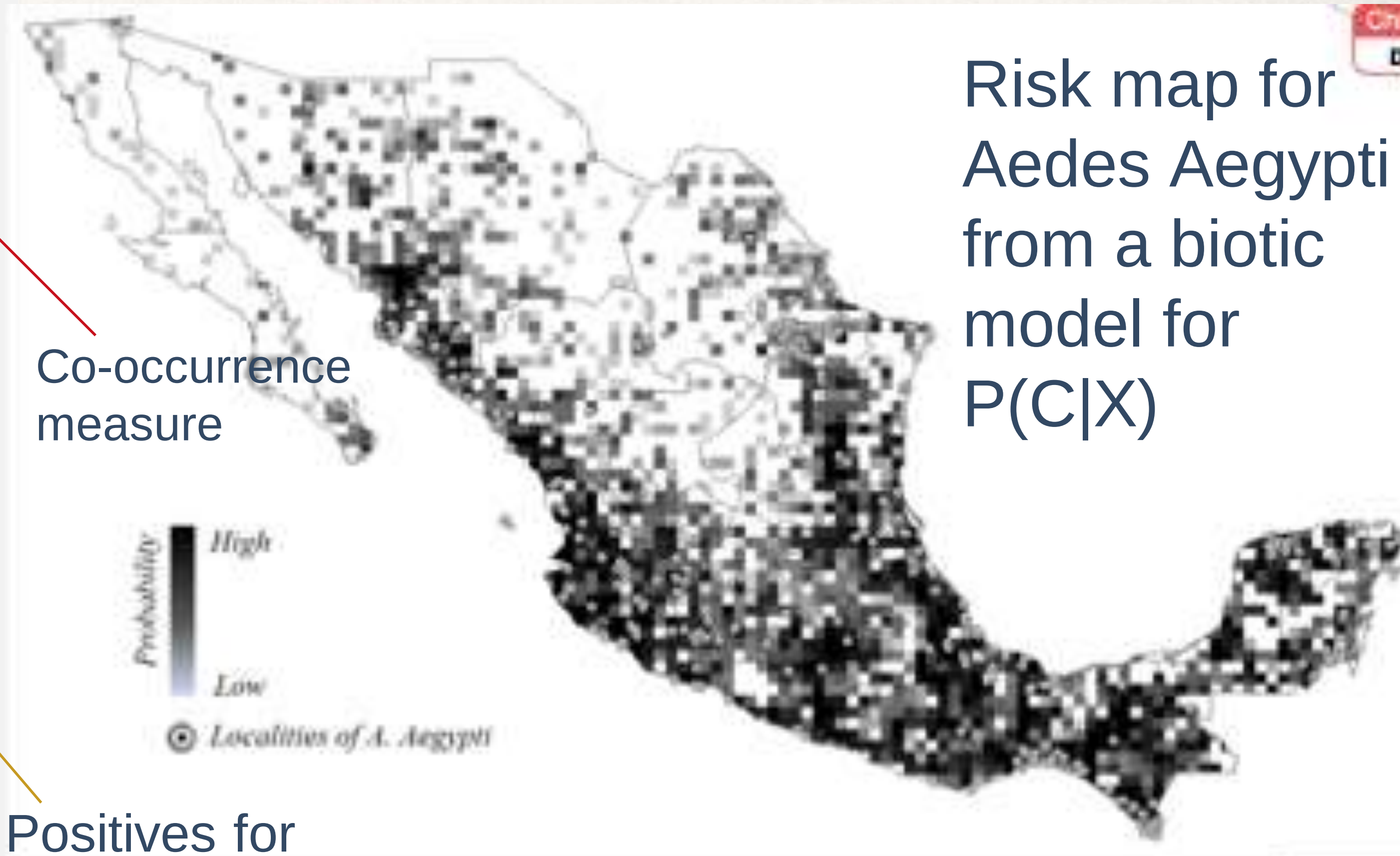
Predictive Model for potential hosts of ZIKV

Mammals with most statistically significant geographic overlap with Aedes Aegypti

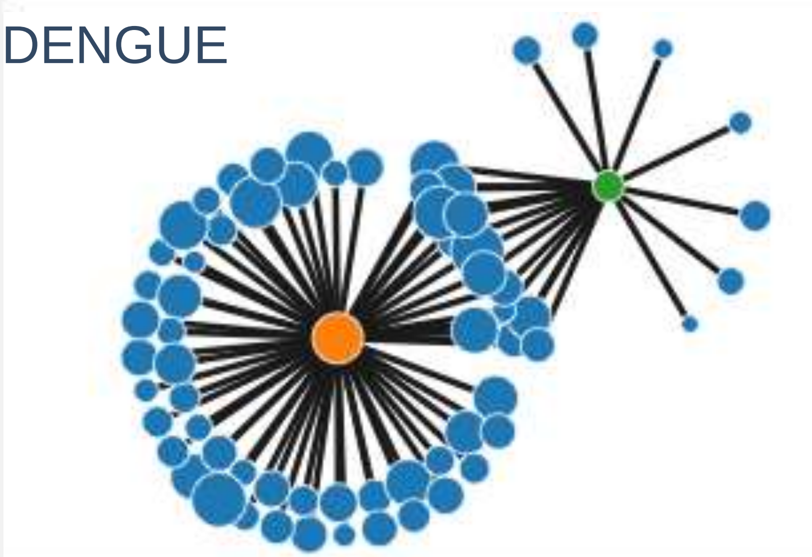
Rank	Mammal	epsilon	Rank	Mammal	epsilon
1	*Glossophaga soricina*	12.78	38	*Dasypus novemcinctus*	7.11
2	*Molossus rufus*	11.99	39	*Sigmodon hispidus*	7.02
3	*Artibeus jamaicensis**	11.68	40	*Uroderma bilobatum*	6.82
4	*Liomys pictus*	11.06	41	*Leptonycteris curasoae*	6.75
5	*Oryzomys couesi*	11.04	42	*Carollia perspicillata*	6.71
6	*Carollia subrufa*	10.49	43	*Centurio senex*	6.61
7	*Sturnira lilium*	10.28	44	*Sciurus colliaei*	6.59
8	*Artibeus lituratus**	9.91	45	*Lontra longicaudis*	6.49
9	*Choeroniscus godmani*	9.42	46	*Didelphis marsupialis*	6.49
10	*Liomys salvini*	9.33	47	*Cratogeomys bulleri*	6.35
11	*Oligoryzomys fulvescens*	9.15	48	*Carollia sowelli**	6.27
12	*Dermanura phaeotis*	9.12	49	*Myotis elegans*	6.12
13	*Rhogeessa tumida*	9.06	50	*Myotis nigricans**	6.06
14	*Pteronotus personatus*	9.05	51	*Sigmodon arizonae*	6.00
15	*Baiomys musculus*	8.97	52	*Rhynchonycteris naso*	5.95
16	*Glossophaga commissarisi*	8.80	53	*Tlacuatzin canescens*	5.87
17	*Didelphis virginiana*	8.58	54	*Leopardus pardalis*	5.84
18	*Pteronotus parnellii**	8.58	55	*Caluromys derbianus*	5.78
19	*Orthogeomys hispidus*	8.53	56	*Molossus molossus*	5.76
20	*Sciurus aureogaster*	8.52	57	*Oryzomys rostratus*	5.76
21	*Molossus sinaloae*	8.51	58	*Osgoodomys banderanus*	5.76
22	*Desmodus rotundus*	8.23	59	*Myotis carteri*	5.66
23	*Saccopteryx bilineata*	8.22	60	*Micronycteris microtis*	5.52
24	*Lasiurus intermedius*	8.15	61	*Sylvilagus brasiliensis*	5.47
25	*Phyllostomus discolor*	8.12	62	*Sylvilagus floridanus*	5.37
26	*Philander opossum*	8.10	63	*Spermophilus annulatus*	5.36
27	*Peromyscus gymnotis*	7.90	64	*Peromyscus leucopus*	5.30
28	*Balantiopteryx plicata*	7.81	65	*Conepatus leuconotus*	5.30
29	*Eptesicus furinalis*	7.69	66	*Chaetodipus pernix*	5.27
30	*Pteronotus davyi*	7.55	67	*Sciurus yucatanensis*	5.23
31	*Dermanura tolteca*	7.48	68	*Sigmodon mascotensis*	5.13
32	*Sciurus variegatoides*	7.48	69	*Eira barbara*	5.12
33	*Mormoops megalophylla*	7.45	70	*Ateles geoffroyi*	5.11
34	*Oryzomys melanotis*	7.42	71	*Neotoma phenax*	5.07
35	*Artibeus intermedius*	7.40	72	*Noctilio leporinus*	5.06
36	*Chaetodipus artus*	7.20	73	*Reithrodontomys fulvescens*	4.95
37	*Nasua narica*	7.18			

Co-occurrence measure

Positives for DENGUE



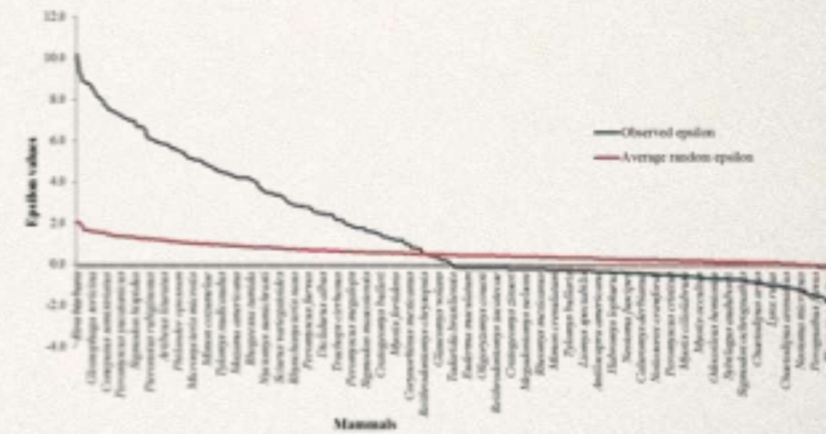
Risk map for Aedes Aegypti from a biotic model for P(C|X)



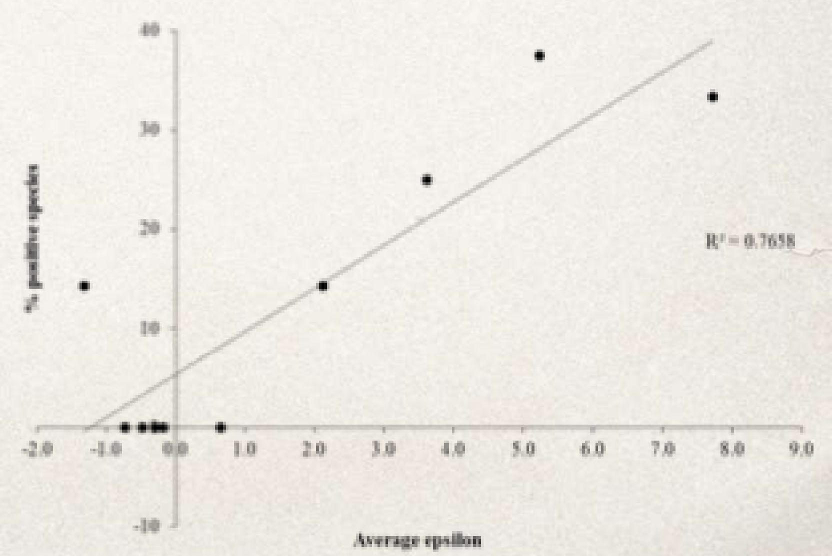
Complex Inference Network for Aedes aegypti and Aedes albopictus

Data-Predictions-Experiment
Test zoonosis - Leishmaniasis

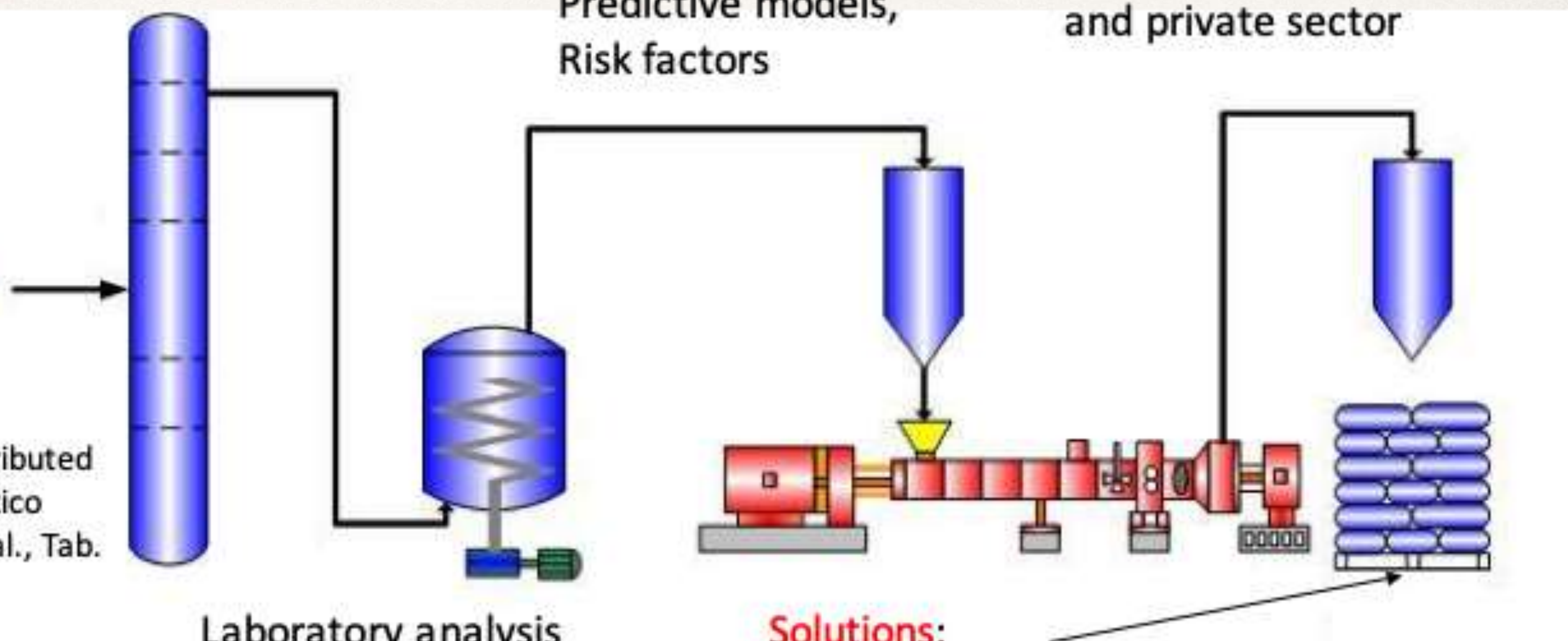
	Mammals	Epsilon	Conf.
1	Eira barbara	10.1683	
2	Rhogeessa aeneus	9.3649	
3	Artibeus intermedius	9.1628	
4	Reithrodontomys gracilis	8.8921	Yes
5	Carollia sowelli	8.8303	
6	Heteromys gaumeri	8.8000	Yes
7	Peromyscus mexicanus	8.7859	
8	Heteromys desmarestianu	8.7164	Yes
9	Molossus rufus	8.6277	
10	Glossophaga soricina	8.5713	
11	Carollia perspicillata	8.5030	
12	Orthogeomys hispidus	8.3468	
13	Pteronotus parnellii	8.1632	
14	Desmodus rotundus	8.1519	
15	Dasyprocta mexicana	8.1128	
16	Sturnira lilium	8.0290	
17	Dermanura phaeotis	8.0055	
18	Dasyprocta punctata	7.9678	
19	Oryzomys couesi	7.7253	
20	Potos flavus	7.7246	
21	Conepatus semistriatus	7.6879	
22	Ototylomys phyllotis	7.5587	Yes
23	Ateles geoffroyi	7.4787	
24	Cryptotis magna	7.4207	
25	Cuniculus paca	7.3220	
26	Lampronycteris brachyotis	7.2852	
27	Sigmodon hispidus	7.2805	Yes
28	Peromyscus yucatanicus	7.2486	Yes
29	Oryzomys chapmani	7.1242	
30	Didelphis virginiana	7.1150	
31	Peromyscus melanocarpus	7.0260	
32	Microtus umbrosus	6.9630	
33	Thyroptera tricolor	6.9630	
34	Nasua narica	6.8953	
35	Megadontomys cryophilus	6.6830	
36	Oryzomys alfaroi	6.6816	
37	Sorex veraepacis	6.6797	
38	Carollia subrufa	6.6316	
39	Peromyscus aztecus	6.6173	
40	Didelphis marsupialis	6.4390	Yes
41	Sciurus yucatanensis	6.3865	
42	Philander opossum	6.2546	
43	Habromys ixtlani	6.1120	
44	Microtus waterhousii	6.1120	
45	Pteronotus rubiginosus	6.1120	
46	Reithrodontomys microdon	6.0967	
47	Coendou mexicanus	6.0268	
48	Centurio senex	6.0076	
49	Artibeus jamaicensis	5.9786	
50	Glossophaga morenoi	5.8847	
51	Molossus sinaloae	5.8518	
52	Artibeus lituratus	5.8422	
53	Mormoops megalophylla	5.8374	
54	Habromys lepturus	5.7848	
55	Myotis keaysi	5.6148	
56	Chiroderma villosum	5.5562	
57	Tamandua mexicana	5.4845	
58	Tylomys nudicaudus	5.4510	
59	Saccopteryx bilineata	5.2984	
60	Macrotus mexicanus	5.2472	
61	Sciurus aureogaster	5.2267	
62	Baiomys musculus	5.2092	
63	Rhogeessa tumida	5.1950	
64	Sciurus deppei	5.1414	
65	Dermanura watsoni	5.1338	
66	Otonyctomys hatti	5.1338	
67	Orthogeomys grandis	5.0556	
68	Alouatta palliata	5.0457	
69	Choeroniscus godmani	5.0457	
70	Peropteryx macrotis	5.0457	
71	Pteronotus personatus	5.0266	
72	Lontra longicaudis	4.9330	
73	Reithrodontomys mexicanu	4.9120	
74	Oryzomys rostratus	4.8681	
75	Mimon cozumelae	4.8327	
76	Pteronotus davyi	4.7943	
77	Herpailurus yagouaroundi	4.7100	
78	Glossophaga leachii	4.6849	
79	Rhogeessa gracilis	4.6317	
80	Sylvilagus brasiliensis	4.6317	
81	Hodomys alleni	4.5155	
82	Leopardus wiedii	4.4420	
83	Peromyscus simulatus	4.4195	
84	Sigmodon alleni	4.3707	
85	Bassariscus sumichrasti	4.3110	
86	Oryzomys fulvescens	4.3110	
87	Diphylla ecaudata	4.3013	
88	Oryzomys melanotis	4.2907	Yes
89	Micronycteris microtis	4.2338	
90	Mazama americana	4.2274	
91	Microtus oaxacensis	4.2061	
92	Rheomys thomasi	4.2061	
93	Oryzomys saturatior	4.2061	
94	Myotis elegans	4.2024	
95	Oligoryzomys fulvescens	4.1984	
96	Natalus stramineus	4.0626	
97	Balantiopteryx io	4.0522	
98	Nyctinomops laticaudatus	4.0522	
99	Tlacuatzin canescens	4.0119	
100	Odocoileus virginianus	3.9265	
101	Balantiopteryx plicata	3.8590	
102	Peromyscus leucopus	3.7994	
103	Sturnina ludovici	3.7888	
104	Enchisthenes hartii	3.6929	
105	Vampyrodes caraccioli	3.6929	
106	Eptesicus furinalis	3.6453	
107	Liomys pictus	3.6107	
108	Glossophaga commissaris	3.4861	
109	Lonchorhina aurita	3.4781	
110	Phyllostomus discolor	3.4781	
111	Peromyscus gymnotis	3.4516	
112	Anoura geoffroyi	3.4201	
113	Platyrrhinus helleri	3.3586	
114	Eumops bonariensis	3.3398	
115	Sciurus variegatoides	3.3398	
116	Uroderma bilobatum	3.3373	
117	Lasiurus intermedius	3.2197	
118	Lasiurus ega	3.1739	
119	Peromyscus megalops	3.1410	
120	Eumops glaucinus	3.0564	
121	Urocyon cinereoargenteus	2.9697	
122	Procyon lotor	2.9502	
123	Hylonycteris underwoodi	2.9343	
124	Rhynchonycteris naso	2.8580	
125	Eptesicus brasiliensis	2.8106	
126	Myotis albescens	2.8106	
127	Lophostoma evotis	2.8106	
128	Tapirus bairdii	2.8106	
129	Vampyrum spectrum	2.8106	
130	Marmosa mexicana	2.7731	Yes
131	Peromyscus furvus	2.7731	
132	Myotis velifera	2.5757	
133	Spilogale putorius	2.5411	
134	Microtus mexicanus	2.5268	
135	Dasypus novemcinctus	2.4725	
136	Myotis nigricans	2.4704	
137	Lophostoma brasiliense	2.4407	
138	Diclidurus albus	2.4407	
139	Sciurus niger	2.4407	
140	Leptonycteris curasoae	2.4268	
141	Nyctomys sumichrasti	2.4026	
142	Sigmodon mascotensis	2.3815	
143	Alouatta pigra	2.3374	
144	Peromyscus melanophrys	2.2204	
145	Dermanura tolteca	2.1920	
146	Trachops cirrhosus	2.1663	
147	Bauerus dubiaquercus	2.1612	
148	Spilogale pygmaea	2.1612	
149	Leptonycteris nivalis	2.1402	
150	Sylvilagus floridanus	2.1002	



Biotic facilitation seems to be the norm. Species are not distributed randomly



To Link Data-Predictions-Experiment
The Emerging Disease “production line”



Predictions-Experiment
Test zoonosis - Leishmaniasis

Species	s	Negative	Positive	Total	% positive	(95%)
Carollia sowelli	8.83	43	2	45	4.4	-1 - 14
*Heteromys gaumeri**	8.8	5	0	5	0	-15 - 29
Peromyscus mexicanus	8.79	115	6	121	5	2 - 11
*Heteromys desmarestianus**	8.72	30	0	30	0	-2 - 16
Molossus rufus	8.63	1	0	1	0	-42 - 56
Glossophaga soricina	8.57	19	7	26	26.9	-3 - 16
Carollia perspicillata	8.5	8	0	8	0	-11 - 24
Pteronotus parnellii	8.16	4	0	4	0	-18 - 31
Desmodus rotundus	8.15	13	1	14	7.1	-6 - 20
Sturnira lilium	8.03	56	7	63	11.1	1 - 13
Artibeus phaeotis	8.01	35	1	36	2.8	-1 - 15
Oryzomys couesi	7.73	2	0	2	0	-28 - 41
*Ototylomys phyllotis**	7.56	9	1	10	10	-9 - 22
*Sigmodon hispidus**	7.28	36	4	40	10	-1 - 14
*Peromyscus yucatanicus**	7.25	3	0	3	0	-22 - 35
Didelphis virginiana	7.12	3	0	3	0	-22 - 30
Didelphis marsupialis	6.44	11	0	11	0	-8 - 21
Philander opossum	6.25	6	1	7	14.3	-12 - 25
Centurio senex	6.01	1	0	1	0	-42 - 56
Artibeus jamaicensis	5.98	81	5	86	5.8	1 - 12
Artibeus lituratus	5.84	38	3	41	7.3	-1 - 14
Myotis keaysi	5.61	2	0	2	0	-28 - 41
Chiroderma villosum	5.56	5	0	5	0	-15 - 29
Saccopteryx bilineata	5.3	1	0	1	0	-42 - 56
Sciurus aureogaster	5.23	71	8	79	7.3	1 - 12
Baiomys musculus	5.21	2	0	2	0	-28 - 41
Artibeus watsoni	5.13	2	0	2	0	-28 - 41
Choeroniscus godmani	5.05	10	3	13	23.1	-7 - 20
Pteronotus personatus	5.03	3	1	4	25	-18 - 31
Reithrodontomys mexicanus	4.91	1	0	1	0	-42 - 56
Oryzomys rostratus	4.87	22	1	23	4.3	-4 - 17
Micronycteris microtis	4.23	1	0	1	0	-42 - 56
Oligoryzomys fulvescens	4.2	6	0	6	0	-13 - 27
Peromyscus leucopus	3.8	22	4	26	15.4	-3 - 16
Sturnira ludovici	3.79	24	1	25	4	-3 - 17
Vampyrodes caraccioli	3.69	1	0	1	0	-42 - 56
Liomys pictus	3.61	47	1	48	2.1	0 - 14
Glossophaga commissarisi	3.49	2	6	8	75	-11 - 24
Lonchorhina aurita	3.48	1	0	1	0	-42 - 56
Phyllostomus discolor	3.48	0	1	1	100	-42 - 56
Platyrrhinus helleri	3.36	5	0	5	0	-22 - 35
Uroderma bilobatum	3.34	4	0	4	0	-18 - 31
Urocyon cinereoargenteus	2.97	1	0	1	0	-42 - 56
Procyon lotor	2.95	1	0	1	0	-42 - 56
Myotis velifer	2.58	3	0	3	0	-18 - 31
Microtus mexicanus	2.53	16	0	16	0	-6 - 19
Myotis nigricans	2.47	2	0	2	0	-28 - 41
Leptonycteris yerbabuenae	2.43	1	1	2	50	-28 - 41
Reithrodontomys fulvescens	2.08	20	0	20	0	-4 - 18
Neotoma mexicana	1.99	5	0	5	0	-15 - 29
Eptesicus fuscus	1.82	1	0	1	0	-42 - 56
Peromyscus levipes	1.34	1	0	1	0	-42 - 56
Sorex saussurei	1.29	3	0	3	0	-22 - 35
Orgoodomys banderanus	1.21	9	0	9	0	-10 - 23
Liomys irroratus	1.16	8	0	8	0	-11 - 24
Myotis auriculus	0.22	2	0	2	0	-28 - 41
Tadarida brasiliensis	-0.09	1	0	1	0	-42 - 56
Peromyscus hylocetes	-0.28	2	0	2	0	-28 - 41
Antrozous pallidus	-0.34	1	0	1	0	-42 - 56
Peromyscus zarhynchus	-0.46	2	0	2	0	-28 - 41
Chaetodipus hispidus	-0.71	4	0	4	0	-18 - 31
Peromyscus pectoralis	-0.73	2	0	2	0	-28 - 41
Neotomodon alstoni	-0.9	17	0	17	0	-5 - 19
Baiomys taylori	-1.16	10	3	13	23.1	-7 - 20
Chaetodipus nelsoni	-1.24	3	0	3	0	-22 - 35
Neotoma micropus	-1.27	16	0	16	0	-6 - 19
Peromyscus maniculatus	-1.37	58	2	60	3.3	0 - 13
Peromyscus eremicus	-1.41	0	1	1	100	-42 - 56
Perognathus flavus	-1.52	1	0	1	0	-42 - 56
Dipodomys merriami	-2.01	1	0	1	0	-42 - 56

- Only about 50 (2.5%) of mammals on the American continent have been identified as hosts of Leishmania
- In Mexico only 8 out of 419 (2.1%) had been identified as hosts
- We collected 922 individuals from 70 species
- Predicted and confirmed 21 new species of mammal as carriers of Leishmania in Mexico
- 13 of them are bats, identified for the first time in Mexico
- Squirrels identified as carriers
- 33% of collected species were confirmed as hosts
- Overall infection rate was 6.7%
- No species could be rejected as a host at this infection rate at the 95% confidence level
- Changes the picture for control of Leishmania totally;
- Leishmania and Lutzomyias are eclectic in their host source.
- Linnean classification is NOT ecologically relevant

What’s more important - biotic or abiotic interactions?

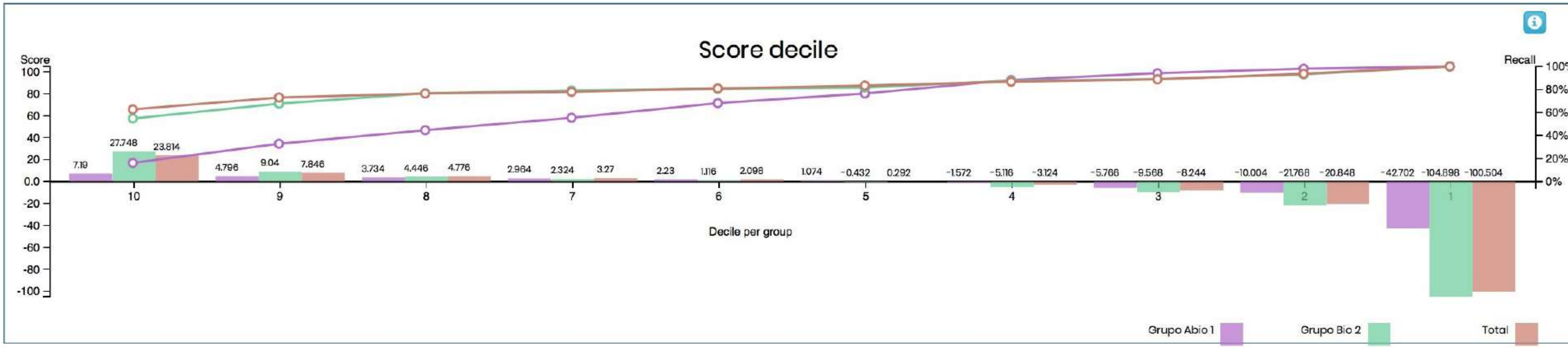


Figure 2. Performance of predicted distribution models for the bobcat based on abiotic variables only (WorldClim), biotic variables only (mammals) and a combination.

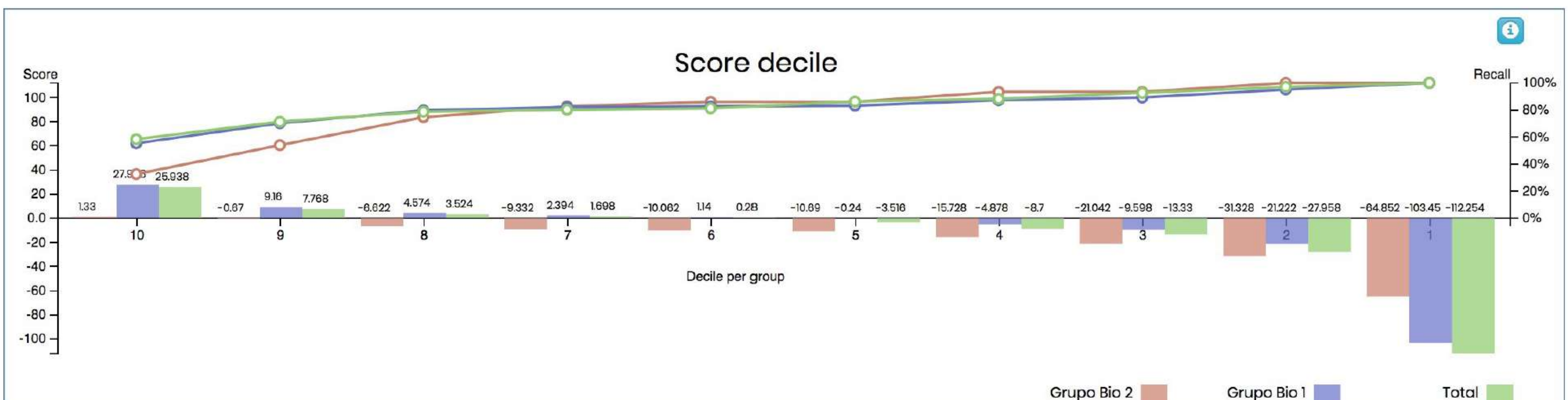
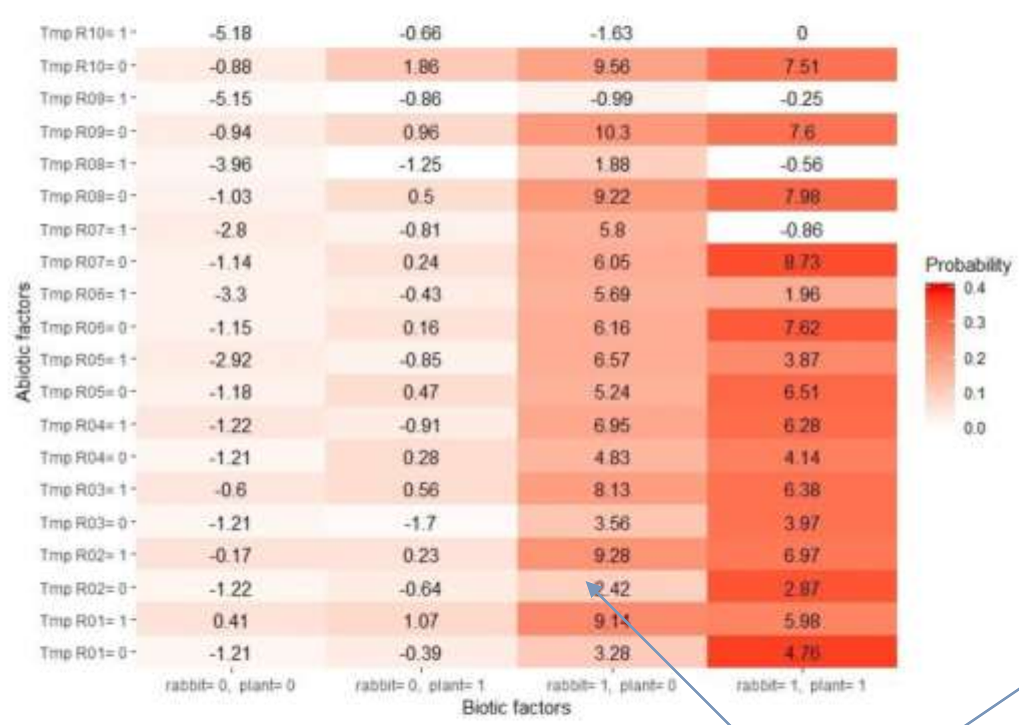


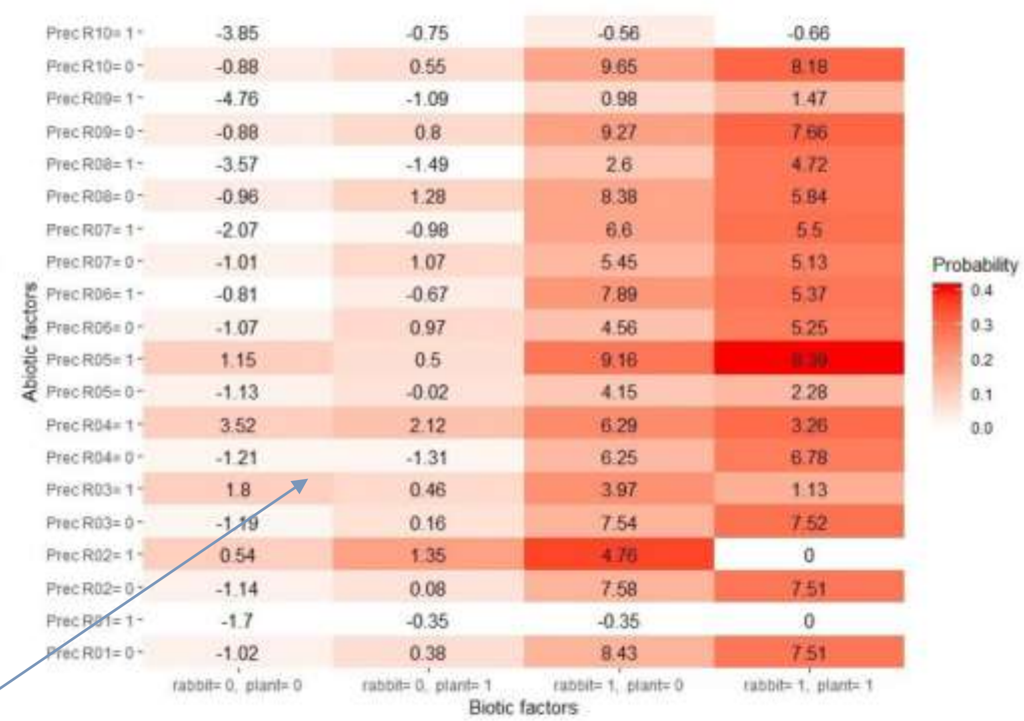
Figure 3. Performance of predicted distribution models for the bobcat based on two classes of biotic variables: Group Bio 1 = Mammalia and Group Bio 2 = Magnoliales, and their combination.

Disentangling causality

Does climate confound biotic factors or vice versa?

Table 1. Probability $P(C|X_\alpha X_\beta X_\gamma)$ and $\varepsilon(C|X_\alpha X_\beta X_\gamma)$ for the bobcat with respect to a prey species, a food source of that prey species and climate.

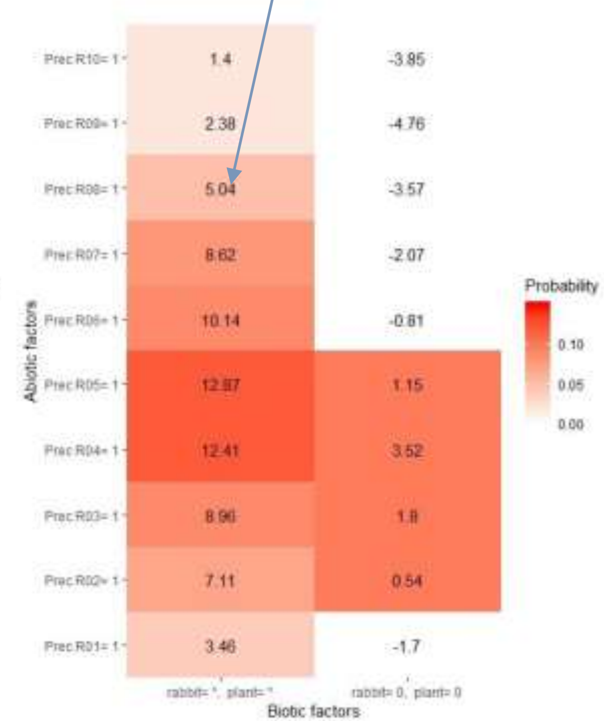




This shows that climate is confounded by biotic factors not vice versa!

This shows that the presence of the bobcat's rabbit prey is more important as a **direct** niche variable than the presence of the food of the rabbit as an **indirect** niche variable which in turn is more important than climate (average annual temperature or average annual precipitation) as an **indirect** niche variable

Table 2: Probability $P(C|X_\alpha X_\beta X_\gamma)$ and $\varepsilon(C|X_\alpha X_\beta X_\gamma)$ for the bobcat in the absence of biotic factors, and $P(C|X_\gamma)$ and $\varepsilon(C|X_\gamma)$, where X_γ represents just climate.



Predicting predator-prey interactions
The bobcat

Table 4: The top 57 highest ranked species by ε corresponding to those species with the most important interaction with the bobcat. The true positive rate in this group is 22.4% compared to the null (random) benchmark of 0.1%.

Species	nij	nj	ni	n	Epsilon	Score	Class	Order	Prey
Canis latrans	106	400	238	26944	54.75	3.7	Mammalia	Carnivora	0
Urocyon cinereoargenteus	85	535	238	26944	37.09	3.05	Mammalia	Carnivora	0
Taxidea taxus	32	87	238	26944	35.79	4.18	Mammalia	Carnivora	0
Lepus californicus	64	383	238	26944	33.1	3.11	Mammalia	Lagomorpha	1
Peromyscus maniculatus	99	871	238	26944	33.06	2.67	Mammalia	Rodentia	1
Otospermophilus variegatus	54	339	238	26944	29.61	3.06	Mammalia	Rodentia	1
Procyon lotor	56	371	238	26944	29.25	2.99	Mammalia	Carnivora	1
Tadarida brasiliensis	66	520	238	26944	28.78	2.79	Mammalia	Chiroptera	0
Sylvilagus audubonii	58	417	238	26944	28.43	2.9	Mammalia	Lagomorpha	1
Puma concolor	33	143	238	26944	28.36	3.52	Mammalia	Carnivora	0
Mephitis macroura	46	279	238	26944	27.86	3.1	Mammalia	Carnivora	1
Odocoileus virginianus	71	633	238	26944	27.78	2.65	Mammalia	Artiodactyla	0
Bassariscus astutus	45	270	238	26944	27.72	3.11	Mammalia	Carnivora	0
Sayornis saya	92	1045	238	26944	27.36	2.38	Aves	Passeri-formes	0
Thomomys bottae	51	351	238	26944	27.32	2.95	Mammalia	Rodentia	0
Haemorhous mexicanus	118	1648	238	26944	27.23	2.16	Aves	Passeri-formes	0
Conepatus leuconotus	41	236	238	26944	27.07	3.16	Mammalia	Carnivora	1
Bubo virginianus	62	519	238	26944	26.93	2.72	Aves	Strigiformes	0
Dipodomys merriami	79	814	238	26944	26.9	2.49	Mammalia	Rodentia	1
Corvus corax	110	1504	238	26944	26.65	2.18	Aves	Passeri-formes	0
Spizella passerina	96	1197	238	26944	26.39	2.28	Aves	Passeri-formes	0
Regulus calendula	89	1044	238	26944	26.39	2.35	Aves	Passeri-formes	0
Icterus parisorum	65	590	238	26944	26.31	2.63	Aves	Passeri-formes	0
Reithrodontomys megalotis	58	488	238	26944	25.97	2.72	Mammalia	Rodentia	1
Sylvilagus floridanus	62	564	238	26944	25.66	2.63	Mammalia	Lagomorpha	1
Ursus americanus	16	43	238	26944	25.46	4.2	Mammalia	Carnivora	0
Lanius ludovicianus	108	1573	238	26944	25.36	2.11	Aves	Passeri-formes	0
Accipiter cooperii	81	939	238	26944	25.36	2.36	Aves	Accipitri-formes	0

This model is 224 times better than chance!

But why is the coyote number 1?

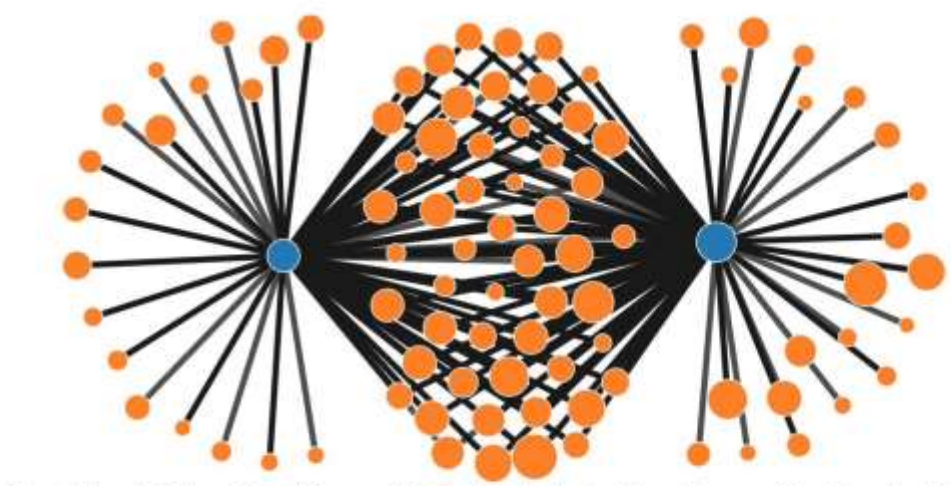


Figure 1. Complex Inference Network between the bobcat (blue circle to the left) and the coyote (blue circle to the right) and the set of potential prey species (orange circles) from the orders Lagomorpha, Artiodactyla and Rodentia. Only the most important interactions corresponding to ε> 8 are shown.

Because they have very similar diets

Can we model using more than just co-occurrence labels?

Build a model using just taxonomic labels as they are a proxy for a host of relevant characteristics of the prey species

Table 3: Model performance for 4 model types: i) ε (unsupervised); ii) taxonomic labels (supervised); iii) ε, s and taxonomic labels (meta-model); and iv) ε and taxonomic labels (meta-model).

Lynx/prey	Total candidate prey species	Total true positives	AUC (ε)	AUC (Taxonomic labels score)	AUC (ε, s(X) and taxonomic labels score)	AUC (ε and taxonomic labels score)
Total	53722	67	0.98	0.99 Std error 0.00	0.99 Std error 0.00	0.99 Std error 0.00
Mammalia	496	50	0.91	0.70 Std error 0.00	0.71 Std error 0.00	0.71 Std error 0.00
Lagomorpha	14	6	0.95	0.78 Std error 0.02	0.94 Std error 0.01	0.94 Std error 0.01

Using a "mixed" model that uses co-occurrence labels and taxonomic labels reduces the false positives. Species co-occur for multiple reasons not just predation. The taxonomic labels help distinguish between the different reasons they co-occur.

Table 5: Impact of taxonomic labels on ranking by ε for most important macro interactions with the bobcat.

LYNX/PREY	Rank by (ε)	Rank by taxonomic labels	Rank by Epsilon, score and taxonomic labels	Rank by Epsilon and taxonomic labels
Canis latrans	1	174	170	172
Urocyon cinereoargenteus	2	145	80	81
Taxidea taxus	3	76	102	102
Lepus californicus (prey)	4	2	2	2
Peromyscus maniculatus (prey)	5	34	39	3

Conclusions

- Ecosystems and in particular EIDs are Complex Adaptive Systems
  - Multi-question, multi-factorial, multi-scale, multi-discipline —> **multi-interaction**
  - There are too many interactions to observe directly
  - Standard mathematical techniques model only a few factors
- Interactions change states and in particular spatio-temporal distributions
  - Interactions can be inferred from comparing spatio-temporal distributions to a non-interacting Null hypothesis
- The Data Revolution has made available large amounts of data with which the spatio-temporal data about organisms, relative to each other (biotic) and relative to the environment (abiotic), can be used to deduce the nature of their interactions
  - This can be done at the niche level (one to many) and at the community level (many to many)
  - Obtaining and integrating data is a huge challenge - political and technical
- Because of their multi-question/multi-discipline/multi-factorial nature, we need to integrate large data sets, which can then be used for rapid hypothesis construction, validation and interpretation by multiple stakeholders and decision makers
- Our solution is open access modelling platforms, such as SPECIES, EPISPECIES and EPIPUMA
- Our work on various zoonosis show the utility of innovative approaches that use data of arbitrary spatial resolution and format, such as predicting host range.
  - Importance of a Data-Predictions-Experiment production line approach to emerging diseases
  - Importance of a multi-pathogen, multi-vector, multi-host approach

SPECIES

Grupo de Trabajo

C3 - Centro de Ciencias de la Complejidad, UNAM; Instituto de Biología, UNAM; CONABIO; Universidad Catolica de Chile; Facultad de Medicina, UNAM; FMVZ, UNAM

1.- Dr. Christopher R. Stephens
2.- Dr. Raúl Sierra Alcocer
4.- Dr. Constantino González Salazar
5.- M. en C. Enrique del Callejo
6.- M. en C. Everardo Robredo
7.- Lic. Juan Carlos Salazar Carrillo
8.- Ma. Juan Barrios
9.- Ing. Raúl Jiménez *
10.- Pedro Romero Martínez
11.- Dr. Pedro Marquet
12.- Dr. Gerardo Suzan
13.- Dr. Víctor Sánchez-Cordero
9.- Dra. Inge Becker
10.- Dr. Angel Rodríguez
11.- Dr. Gabriel Gutierrez
12.- Dra. Miriam Berzunza
13.- Dr. Eduardo Rebollar
14.-15.- Dr. Gabriel Garcia
16.- M. Jesus de Sotomayor
17.- M. Maria José Tolsa
18.- L. Fabiola Nieto
19.- M. Laura Rengifo
20.- Dr. Carlos Ibarra

Publications

Competitive interactions between felid species may limit the southern distribution of bobcats Lynx rufus
V Sánchez-Cordero, D Stockwell, S Sarkar, H Liu, CR Stephens, ...
Ecography 31 (6), 757-764, 2008

Using biotic interaction networks for prediction in biodiversity and emerging diseases
CR Stephens, JG Heau, C González, CN Ibarra-Cerdeña, ...
PLoS One 4 (5), e5725, 2009

Exploratory analysis of the interrelations between co-located boolean spatial features using network graphs
R Sierra, CR Stephens
International Journal of Geographical Information Science 26 (3), 441-468, 2012

Constructing ecological networks: a tool to infer risk of transmission and dispersal of Leishmaniasis
C González-Salazar, CR Stephens
Zoonoses and public health 59 (s2), 179-193, 2012

Comparing the relative contributions of biotic and abiotic factors as mediators of species' distributions
C González-Salazar, CR Stephens, PA Marquet
Ecological Modelling 248, 57-70, 2013

Leishmania (L.) mexicana Infected Bats in Mexico: Novel Potential Reservoirs
M Berzunza-Cruz, Á Rodríguez-Moreno, G Gutiérrez-Granados, ...
PLoS neglected tropical diseases 9 (1), e0003438-e0003438, 2015

Predicting the potential role of non-human hosts in Zika virus maintenance
C González-Salazar, CR Stephens and V. Sanchez-Cordero
Eco-health14 (1), 171-177

Leishmania (L.) mexicana infected bats in Mexico: novel potential reservoirs
M Berzunza-Cruz, Á Rodríguez-Moreno, G Gutiérrez-Granados, ...
PLoS Negl Trop Dis 9 (1), e0003438

Trypanosoma cruzi reservoir—triatomine vector co-occurrence networks reveal meta-community effects by synanthropic mammals on geographic dispersal
CN Ibarra-Cerdeña, L Valiente-Banuet, V Sánchez-Cordero, ...
PeerJ 5, e3152

UNDERSTANDING TRANSMISSIBILITY PATTERNS OF CHAGAS DISEASE THROUGH COMPLEX VECTOR-HOST NETWORKS
Laura Rengifo-Correa, Constantino González-Salazar, Juan J. Morrone, Juan Luis Téllez-Rendón, Christopher Stephens, Parasitology 144 (6), 760

Can you judge a disease host by the company it keeps? Predicting disease hosts and their relative importance using complex networks
CR Stephens et al, PLoS neglected tropical diseases 10 (10), e0005004

Can Ecological Interactions be Inferred from Spatial Data?
CR Stephens, C González-Salazar, M Villalobos, P Marquet
Biodiversity Informatics 15 (1), 11-54

SPECIES: Supporting big-data-driven research
R Sierra-Alcocer, C Stephens, J Barrios, C González-Salazar, ...
Biodiversity Information Science and Standards 3, e36095

SPECIES: A platform for the exploration of ecological data
CR Stephens, R Sierra-Alcocer, C González-Salazar, JM Barrios, ...
Ecology and evolution 9 (4), 1638-1653

Bayesian inference of ecological interactions from spatial data
CR Stephens, V Sánchez-Cordero, C González Salazar
Entropy 19 (12), 547

Epi-PUMA/Epi-SPECIES

Chris Stephens C3 y ICN, UNAM
Rodolfo Rivas IMSS
David Barros IMSS
Victor Villalobos CENAPRECE
Raúl Sierra Alcocer CONABIO
Juan Carlos Salazar CONABIO
Pedro Romero CONABIO
Gabriel Garcia FMZN, UNAM
Constantino Gonzalez CCA, UNAM
Julián Velasco CCA, UNAM
Antonio Barajas FM UNAM
Luís Flores FM UNAM
Carlos A. Hernández IIMAS UNAM
Marco Rosas, C3, UNAM
Adriana Robles, C3 UNAM
Estefa Espitia, C3, UNAM
Miriam Barajas, C3, UNAM
Jonathan Easton C3, UNAM
Hugo Flores C3, UNAM
Dagmara Wrzecionkowska F Psic, UNAM
Ivette Rizzo C3, UNAM
Rolando Díaz-Loving, Fac. Psic., UNAM
Manuel Suarez Igg, UNAM
Juan Pablo Gutierrez FM, UNAM
Gustavo Olaiz CIPSS, UNAM
Armando Sanchez IIE, UNAM

Jésica Tapia C3, UNAM
Lilian Castillo IIMAS, UNAM
Alejandro Ruíz IIMAS, UNAM
Alejandro Salinas IIMAS, UNAM
Tanya Arenas C3, UNAM
Adrian Ghilardo CIGA, UNAM
Gerardo Suzán FMVZ, UNAM
Romel Calero C3, UNAM
José Luis Gordillo C3, UNAM
Hector Resendiz, C3 UNAM

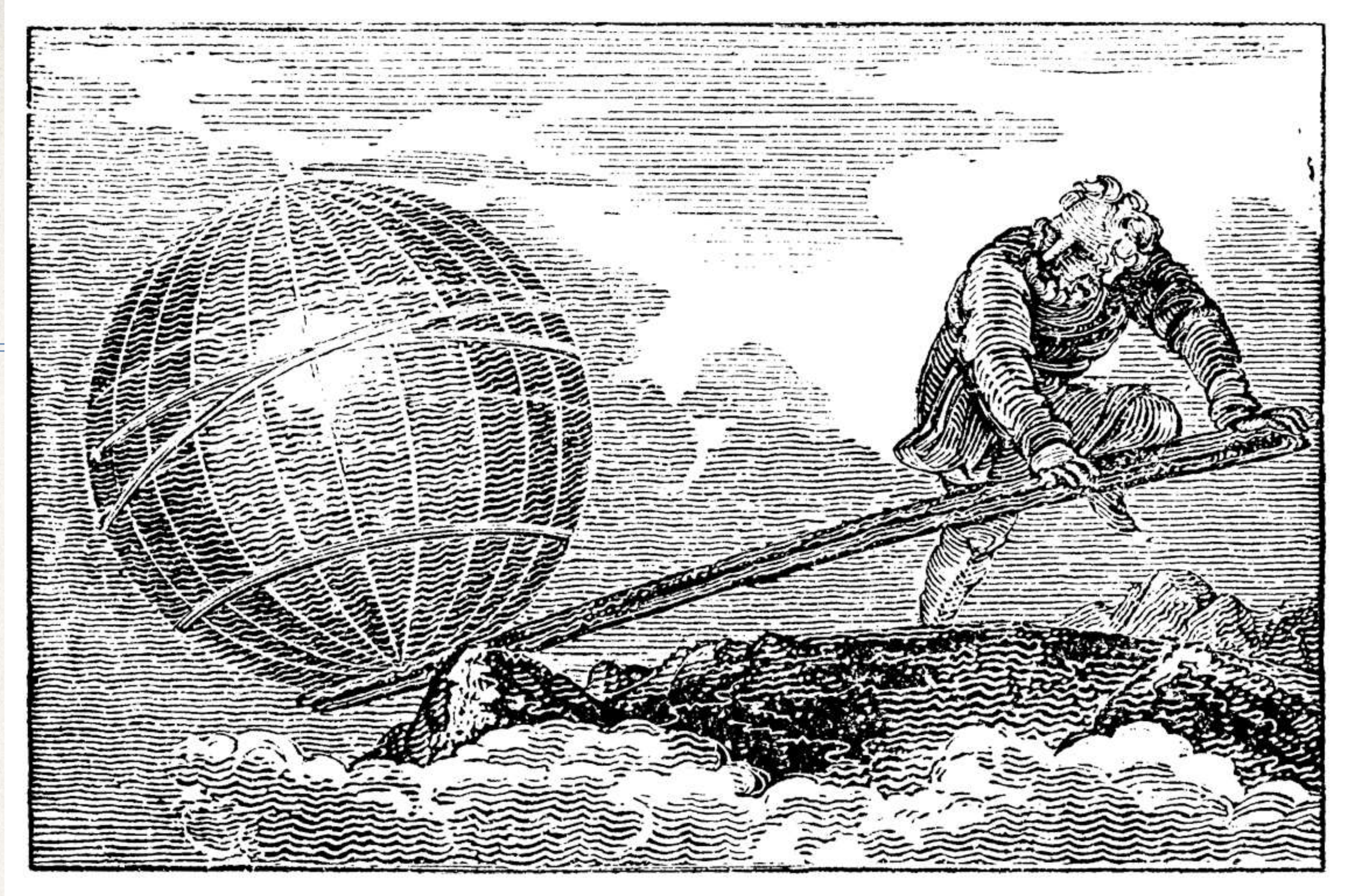
CONABIO

CONACYT

dgapa
Dirección General de Asuntos del Personal Académico

SECTEI

Microsoft



δῶς μοι πᾶ στῶ καὶ τὰν γᾶν κινάσω

Give me a place to stand on and I´ll move the earth

Give me enough data and I´ll predict anything

The Data Revolution will revolutionise our ability to model and understand ecology