How Firms Can Apply Data Science To Save Species

Timothy C. Haas, Associate Professor (Statistics)

Sheldon B. Lubar School of Business, University of Wisconsin-Milwaukee, haas@uwm.edu, https://sites.uwm.edu/haas/

Biodiversity is going away

- The sixth mass extinction in the history of the planet is underway.
- Most large, wild mammals, many fish species, and many rare plants will be gone by 2060.

Cycads, sharks, and elephants

- For instance, the cycad plant, poached as a status symbol and investment, has been on this planet for about 280 million years. Dinosaurs didn't show up until 245 million years ago.
- The great white shark, a particular species of fish is endangered.
- And the African savanna elephant was added to the IUCN Red List in 2021.

What to do

- To curb this non-reversible destruction, the wholesale killing of animals and plants needs to stop, and habitat destruction needs to be curtailed.
- But achieving these two goals will require initiatives that move people away from these behaviors.
- These initiatives need to be derived from credible models of those **political-ecological** systems that host endangered species.

Private enterprise could save biodiversity

- Private, for-profit firms could single-handedly save biodiversity by developing and running these initiatives.
- **REALLY? HOW?**
- In short, by making biodiversity conservation profitable.

Selling to biodiversity-concerned customers

Step 1: A firm identifies a species they want to save.

Step 2: They launch a new product or service called a **biodiversity offering** that is attached to a **biodiversity project**.

Step 3: They maintain a public-facing **biodiversity dashboard** that contains real-time, detailed, and audited information on the project and the species being saved.

Why would this work?

- Firms exist to make a profit. Biodiversity offerings would be priced and managed to be profitable.
- Biodiversity offerings would tap into a giant market niche of people wanting to help but feeling powerless.
- Customers buying the offering would know that they are paying a premium that will cause a measurable improvement in the species' sustainability.

Firms are resourceful and distributed

- Firms hold most of the world's wealth and expertise. Collectively, they have the resources to solve a planet-level problem.
- Because firms are not all under a single, hierarchical control structure, a few newly-minted tyrants would not be able to shutdown all firm-level biodiversity projects.

Building a biodiversity project

- First, the firm builds a model of the political-ecological system that contains their species.
- Based on this model, the firm identifies a biodiversity project that will encourage species-preserving behaviors.
- Finally, the firm implements this project and an attendant monitoring program that feeds real-time data to the firm's biodiversity dashboard.

Suggested biodiversity projects (ranked by effectiveness)

1. Improving habitat by reducing the firm's raw material demands from species-hosting countries.
2. Improving habitat by relocating/shutting operations within species-hosting countries.
3. Reducing poaching by owning and operating a production/service facility in a city that draws people away from ecological **hotspots**.

Biodiversity projects (continued)

4. Providing data, software, and hardware to international law enforcement teams fighting wildlife trafficking.
5. Owning and operating a private wildlife reserve.
6. Owning and operating captive breeding facilities for nearly-extinct species.

Needed technology

- To be seen as credible, political-ecological models need to be statistically fitted to data streams of parsed news articles, and ecological metrics such as species abundance.
- These fitted models need to be simulated under different biodiversity projects to find ones that work.
- To support these tasks, firms need to develop data science techniques in the following eight areas.

Data acquisition and social networks

1. Methods from computational linguistics to detect actions from online news outlets and social media sites.
2. Satellite/drone-to-cloud solutions to remotely sense animal and plant populations.
3. Methods from machine learning for fitting and evaluating social network models of wildlife trafficking syndicates.

High-performance computing and graphics

4. Optimization of high-dimensional, non-convex, and expensive functions to support the tasks of:
   4.1 statistically estimating models containing many parameters
   4.2 assessing model out-of-sample predictive performance
   4.3 assessing parameter sensitivity
   4.4 searching a space of model-computed biodiversity projects for the optimal one.
5. Graphical methods to display a categorical time series of political-ecological actions.

Federated databases that are secure

6. Distributed databases that enable law enforcement agencies to securely share between themselves, criminal intelligence on wildlife traffickers.
7. Tools to ensure the physical and electronic security of political-ecological action databases so as to keep them from being used by poachers to locate targets.
8. Algorithms for running queries against these massive and distributed relational databases.

Examples

Four examples are described:

I. Remote sensing of wildlife populations
II. Social media mining and social network analysis of wildlife trafficking syndicates
III. High performance computing
IV. Political-ecological databases

Example I: Remote sensing

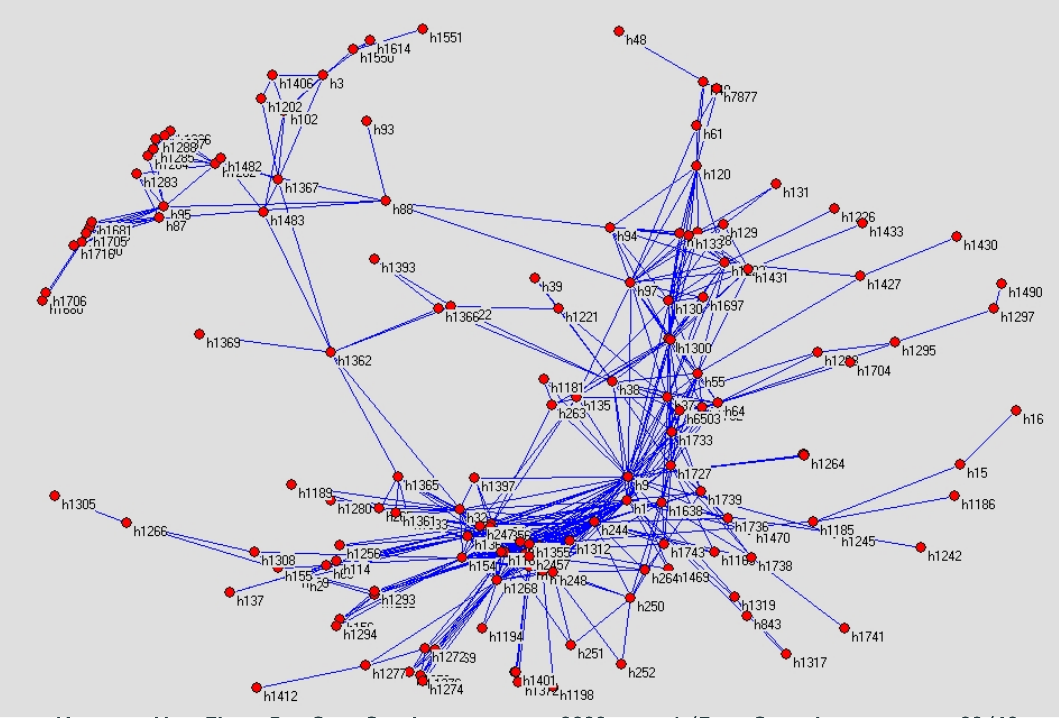
- The next slide is a commercial satellite image of “Elephant Valley” at San Diego Zoo Safari Park.
- I’ve written an algorithm that correctly counts the thirteen elephants therein using only this image.

Example II: Wildlife trafficking networks

- Criminal syndicates kill and transport wildlife from source country to consumer country.
- For instance, in 2012, there were 10,000 rhinos in Kruger National Park (KNP), South Africa. Today there are about 3,000.
- Rhinos are shot for their horns that are then sold in Vietnam and China as traditional medicine.

Even a local trafficking syndicate is large

The next slide exhibits the estimated KNP rhino poaching network as of December 2014. Cell phone intercepts were used to fit this model. Players are poachers, couriers, or middlemen.


h1551
h1614
h48
h7877
h1406
h3
h1202
h102
h93
h61
h1288
h1482
h120
h131
h1283
h1367
h88
h1226
h1433
h95
h87
h1483
h94
h129
h1430
h1681
h1705
h1393
h1431
h1427
h39
h97
h1490
h1706
h1221
h130
h1697
h1297
h1369
h1300
h1295
h1362
h1704
h55
h1181
h38
h135
h263
h64
h6503
h16
h1733
h1264
h1727
h15
h1365
h1397
h9
h1189
h1739
h1305
h1280
h1638
h1186
h1266
h244
h1736
h1185
h1470
h1245
h1256
h1355
h1312
h1743
h1242
h1308
h2457
h114
h1738
h248
h1469
h137
h1268
h1293
h250
h1319
h843
h1294
h1194
h1741
h251
h1317
h1412
h1277
h1274
h1401
h252
h1198

What is needed

- Algorithms to detect hidden links between traffickers using incomplete criminal intelligence
- Data science systems that can automatically summarize a syndicate's social network and from that, deliver **actionable intelligence** to law enforcement
- This intelligence includes particular traffickers to arrest, particular wire transfers to block, and particular wildlife contraband shipments to interdict.

Example III: High performance computing

Data science can provide tools to build models of political-ecological systems that can be used to discover politically feasible, species-saving biodiversity projects.

Modeling a political-ecological system

- Agent-based submodels make decisions about actions that affect an at-risk species.
- Agents include poachers, kingpins, consumers, farmers, wildlife protection agencies, and governments.
- An individual-based submodel of how the targeted species' abundance affects and is affected by agent decisions.

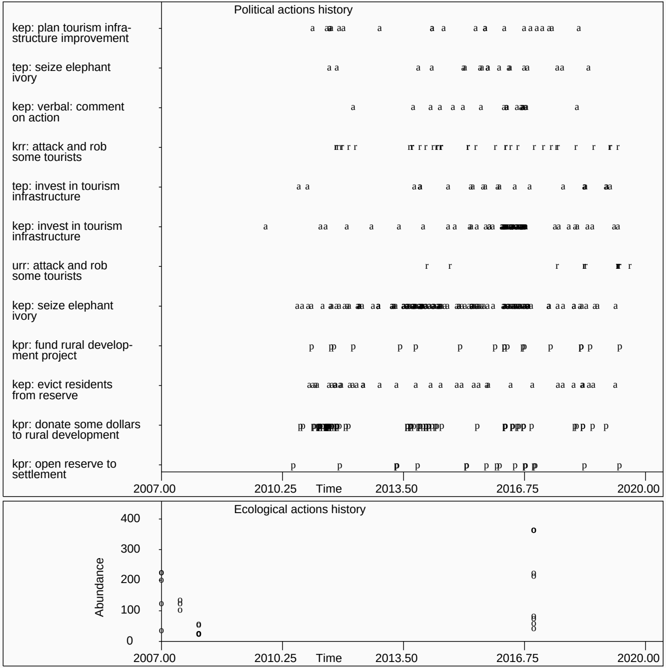
Need for massive computing

Runs of these stochastic models are expensive. Hence, massive computing resources are needed to compute:

- Statistical estimates of model parameters
- Out-of-sample prediction error rates
- Parameter sensitivities
- Politically-feasible and species-saving biodiversity projects

East African actions from online news

This plot is for the period from January 2007 through June 2019. The symbol "p" indicates an action taken by a presidential office, "a" an action taken by an EPA, "r" an action taken by rural residents, "s" an action taken by pastoralists, and "n" an action taken by an NGO. Selected out-combinations only are labeled. The bottom plot is observed cheetah abundance.

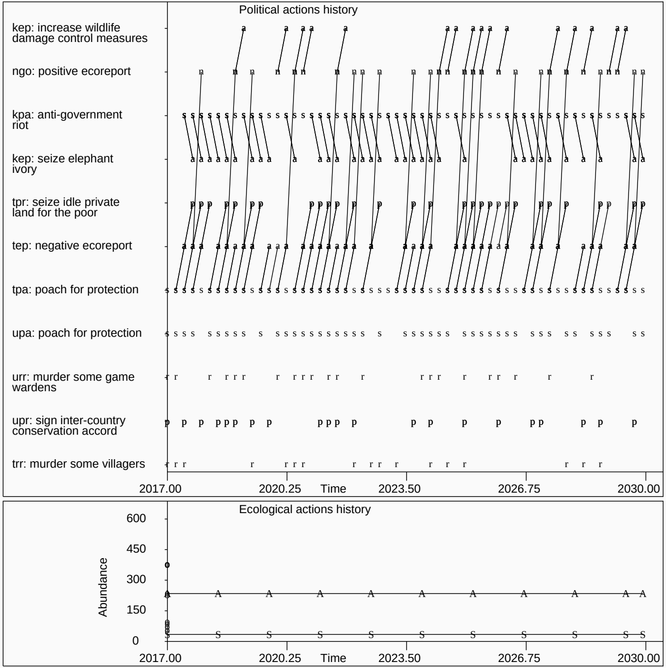

Political actions history
kep: plan tourism infrastructure improvement
tep: seize elephant ivory
kep: verbal: comment on action
krr: attack and rob some tourists
tep: invest in tourism infrastructure
kep: invest in tourism infrastructure
urr: attack and rob some tourists
kep: seize elephant ivory
kpr: fund rural development project
kep: evict residents from reserve
kpr: donate some dollars to rural development
kpr: open reserve to settlement
2007.00
2010.25
Time
2013.50
2016.75
2020.00
Ecological actions history
Abundance
400
300
200
100
0
2007.00
2010.25
Time
2013.50
2016.75
2020.00

Parameter estimation

Find submodel parameter values that maximize the agreement between submodel-generated actions/abundance, and observed actions/abundance while simultaneously minimizing the Hellinger distance to a set of hypothesized parameter values.

Forecasted actions under a proposed project

- The following is a plot of the cheetah system simulator's actions history under a proposed project.
- Lines connect action-reaction sequences.
- One frequent action sequence or **episode** is:
  poaching event $\rightarrow$
  negative ecosystem status report $\rightarrow$
  land gift to the poor.


Political actions history
kep: increase wildlife damage control measures
ngo: positive ecoreport
kpa: anti-government riot
kep: seize elephant ivory
tpr: seize idle private land for the poor
tep: negative ecoreport
tpa: poach for protection
upa: poach for protection
urr: murder some game wardens
upr: sign inter-country conservation accord
trr: murder some villagers
2017.00
2020.25
Time
2023.50
2026.75
2030.00
Ecological actions history
Abundance
600
450
300
150
0

What is needed

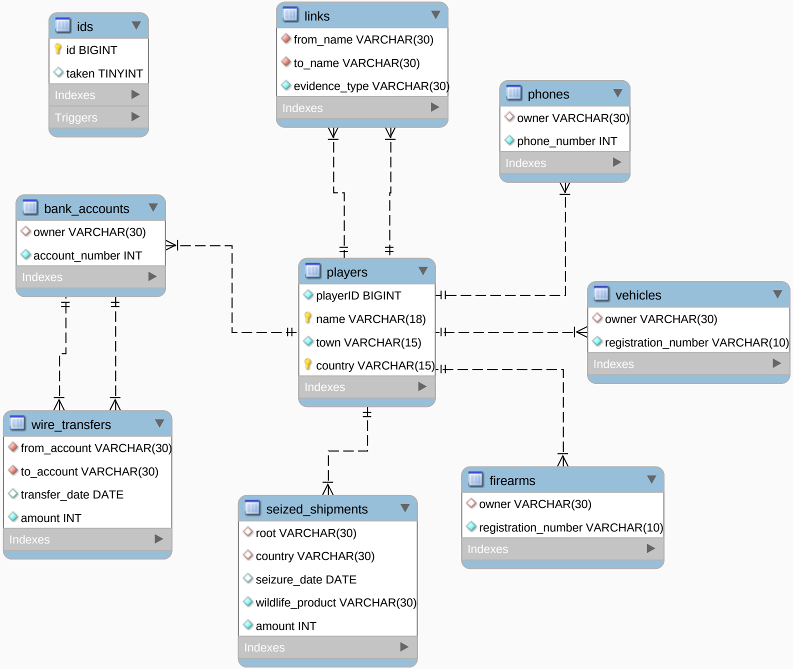
- This graphic is too busy, hard-to-read, and unattractive. New methods in computational graphics are needed that can display a categorical time series in a way that is easily interpretable.
- Algorithms are needed that can detect significant features of this **actions history**. Episode detection is a first but primitive step in this direction.

Example IV: Political-ecological databases

- Two database types: one for criminal intelligence on wildlife traffickers; and one on political-ecological actions that impact a species-hosting ecosystem.
- Both types need to be federated and secure against attacks by traffickers aiming to acquire information on criminal investigations and/or species locations.
- Both database types can become large. Hence, scalable query algorithms are needed.

1: A criminal intelligence database

- Poachers and/or couriers may be in contact with several different poaching rings.
- Players may have fled to another country to avoid prosecution.
- Shared criminal intelligence would help law enforcement prosecute these wildlife traffickers.
- The next slide is the entity-relationship diagram of a federated criminal intelligence database.


ids
id BIGINT
taken TINYINT
Indexes
Triggers
links
from_name VARCHAR(30)
to_name VARCHAR(30)
evidence_type VARCHAR(30)
Indexes
phones
owner VARCHAR(30)
phone_number INT
Indexes
bank_accounts
owner VARCHAR(30)
account_number INT
Indexes
players
playerID BIGINT
name VARCHAR(18)
town VARCHAR(15)
country VARCHAR(15)
Indexes
vehicles
owner VARCHAR(30)
registration_number VARCHAR(10)
Indexes
wire_transfers
from_account VARCHAR(30)
to_account VARCHAR(30)
transfer_date DATE
amount INT
Indexes
seized_shipments
root VARCHAR(30)
country VARCHAR(30)
seizure_date DATE
wildlife_product VARCHAR(30)
amount INT
Indexes
firearms
owner VARCHAR(30)
registration_number VARCHAR(10)
Indexes

2: A political-ecological actions database

Actions may be:

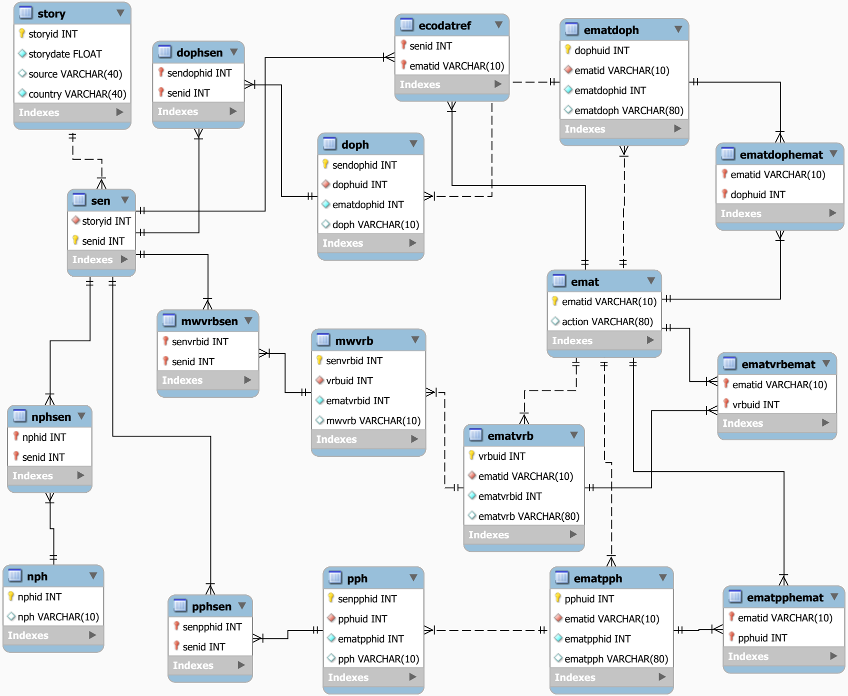
- local to a species-hosting country;
- concern interactions between consumption/donor countries and a species-hosting country;
- be taken by ecosystem managers;
- or be reactions by the ecosystem, itself.

Stream parsing of online news articles

- Parsing algorithms are needed that match actions reported in online articles and social media to actions in a taxonomy.
- Associated learning algorithms are needed that allow the taxonomy to be updated and extended so that it can represent new ecosystem-affecting actions.
- These tasks are more difficult than performing classic sentiment analysis.

Entity-relationship diagram

- The following is an entity-relationship diagram of one such database.
- Rectangles are entities. Rows within rectangles are attributes that take on values.
- A double bar into an entity indicates a source entity can map to only one entity whereas a trident indicates a source entity can map to many entities.


story
storyid INT
storydate FLOAT
source VARCHAR(40)
country VARCHAR(40)
Indexes
dophsen
sendophid INT
senid INT
Indexes
ecodatref
senid INT
ematid VARCHAR(10)
Indexes
ematdoph
dophuid INT
ematid VARCHAR(10)
ematdophid INT
ematdoph VARCHAR(80)
Indexes
doph
sendophid INT
dophuid INT
ematdophid INT
doph VARCHAR(10)
Indexes
ematdophemat
ematid VARCHAR(10)
dophuid INT
Indexes
sen
storyid INT
senid INT
Indexes
emat
ematid VARCHAR(10)
action VARCHAR(80)
Indexes
mwvrbsen
senvrbid INT
senid INT
Indexes
mwvrb
senvrbid INT
vrbuid INT
ematvrbid INT
mwvrb VARCHAR(10)
Indexes
ematvrbemat
ematid VARCHAR(10)
vrbuid INT
Indexes
nphsen
nphid INT
senid INT
Indexes
ematvrb
vrbuid INT
ematid VARCHAR(10)
ematvrbid INT
ematvrb VARCHAR(80)
Indexes
nph
nphid INT
nph VARCHAR(10)
Indexes
pphsen
senpphid INT
senid INT
Indexes
pph
senpphid INT
pphuid INT
ematpphid INT
pph VARCHAR(10)
Indexes
ematpph
pphuid INT
ematid VARCHAR(10)
ematpphid INT
ematpph VARCHAR(80)
Indexes
ematpphemat
ematid VARCHAR(10)
pphuid INT
Indexes

Thank you!