

Global Correlates of Emerging Zoonoses

**Anthropogenic, Environmental,
and Biodiversity Risk Factors**

Toph Allen, Kris A. Murray, Carlos Zambrana-Torrel, Stephen S. Morse,
Carlo Rondinini, Moreno Di Marco, Nathan Breit, Kevin J. Olival, Peter Daszak



Background

Jones et al. (2008) – “Global trends in emerging infectious diseases”

EID events

“...the first temporal emergence of a pathogen in a human population which was related to the increase in distribution, increase in incidence or increase in virulence or other factor which led to that pathogen being classed as an emerging disease.”

Table 1 | Socio-economic, environmental and ecological correlates of EID events

Pathogen type	Zoonotic: wildlife	
Number of EID event grid cells	147–156	
	<i>b</i>	<i>B</i>
log(JID articles)	0.34–0.37***	1.41–1.45
log[human pop. density (persons per km ²)]	0.56–0.64***	1.75–1.90
Human pop. growth (change in persons per km ² ,1990–2000)†	0.09–0.45	1.09–1.56
Latitude (decimal degrees)	0.002–0.017	1.00–1.02
Rainfall (mm)	(0.14–0.06) x 10 ^{−3}	1.00–1.00
Wildlife host richness	0.008–0.013**	1.01–1.01
Constant	−9.81 to −8.78***	

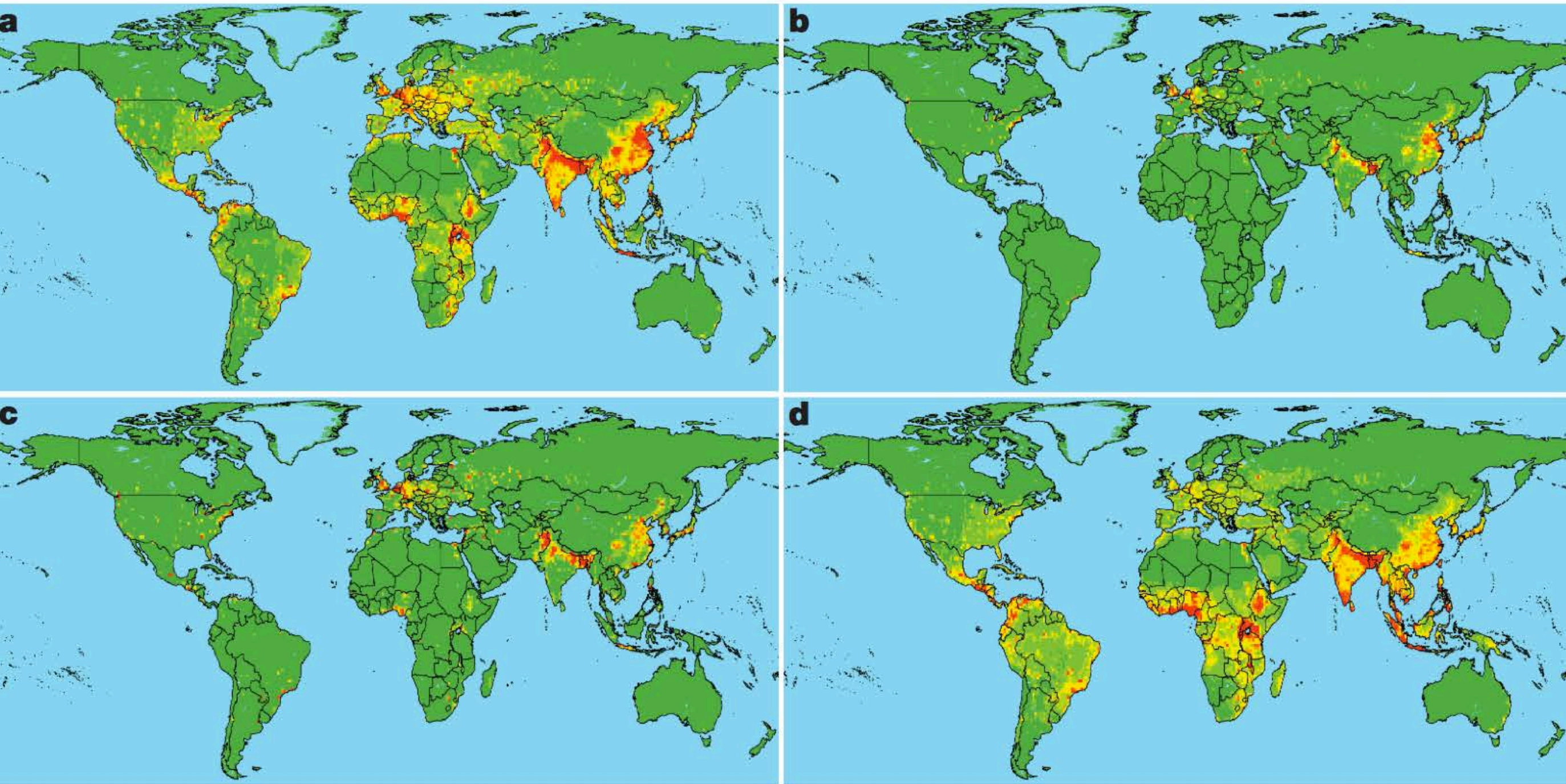


Figure 3 | Global distribution of relative risk of an EID event. Maps are derived for EID events caused by **a**, zoonotic pathogens from wildlife, **b**, zoonotic pathogens from non-wildlife, **c**, drug-resistant pathogens and **d**, vector-borne pathogens. The relative risk is calculated from regression coefficients and variable values in Table 1 (omitting the variable measuring reporting effort), categorized by standard deviations from the mean and mapped on a linear scale from green (lower values) to red (higher values).

Goals

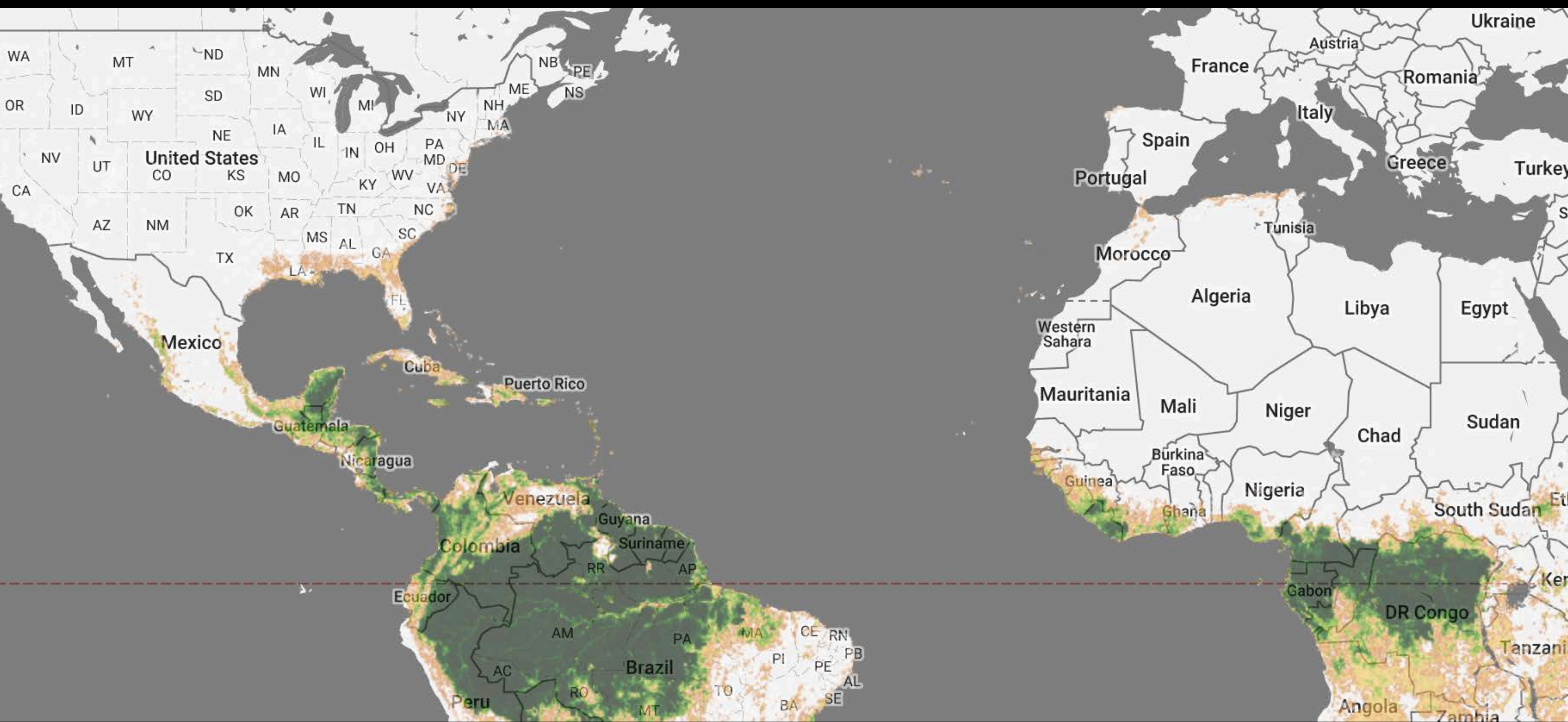
- Hypotheses based on biological reasoning
- Updated set of predictors
- Re-think reporting effort measure
- Updated modeling framework:
 - Quantify uncertainty (sampling and spatial)
 - More robust to non-linear effects, sparse data, etc.

Biological reasoning

- Analyze only wildlife-origin zoonoses
- Select predictors based on hypothesized mechanisms

Updated predictors

- New spatial layers for mammal species richness, land cover, climate
- Temporal matching of variables where possible



EarthEnv dataset, “evergreen broadleaf trees”

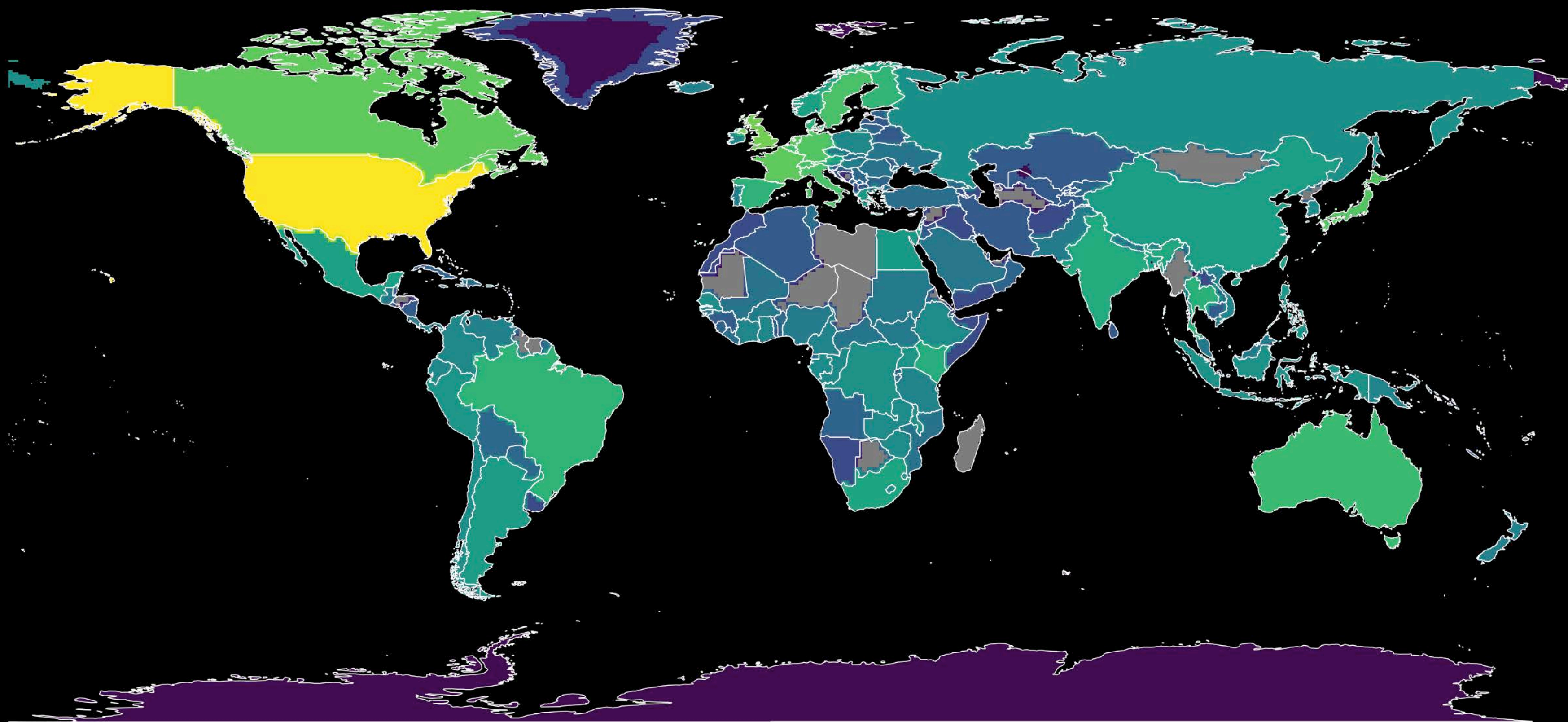
Reporting effort

Reporting effort

Goal: “factor bias out”

Prior reporting effort measures

- Jones et al. (2008): Summed author counts from JID articles 1973 – 2008
- Yang et al. (2012): Summed results for PubMed searches for “infectious disease + [country name]”
- All country-level



Jones et al.'s measure of reporting effort

PubCrawler

<https://github.com/ecohealthalliance/pubcrawler/>

PubCrawler

- New measure of reporting effort
- High-resolution
- Based on PubMed Central Open Access Subset (PMCOAS)

PubCrawler

Extracted and aggregated
toponyms from articles in the
PMCOAS

PubCrawler

???

PubCrawler

1. Filter

PubCrawler

1. Filter
- 2. Resolve toponyms (place names)**

PubCrawler

1. Filter
2. Resolve toponyms (place names)
3. **Normalize and aggregate**

PubCrawler

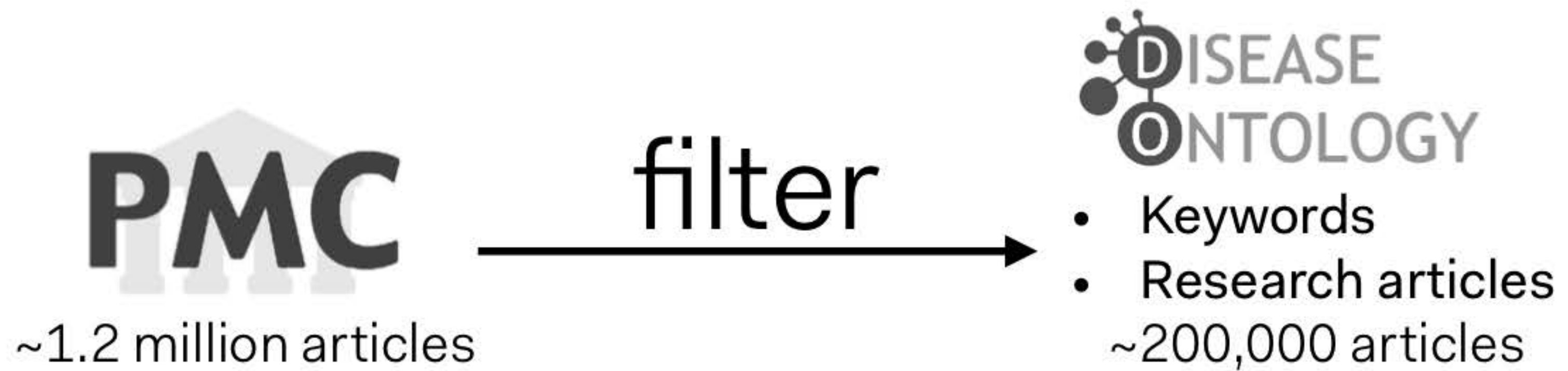
1. Filter
2. Resolve toponyms (place names)
3. Normalize and aggregate
4. **Fit boosted regression tree model**

PMC
~1.2 million articles

filter

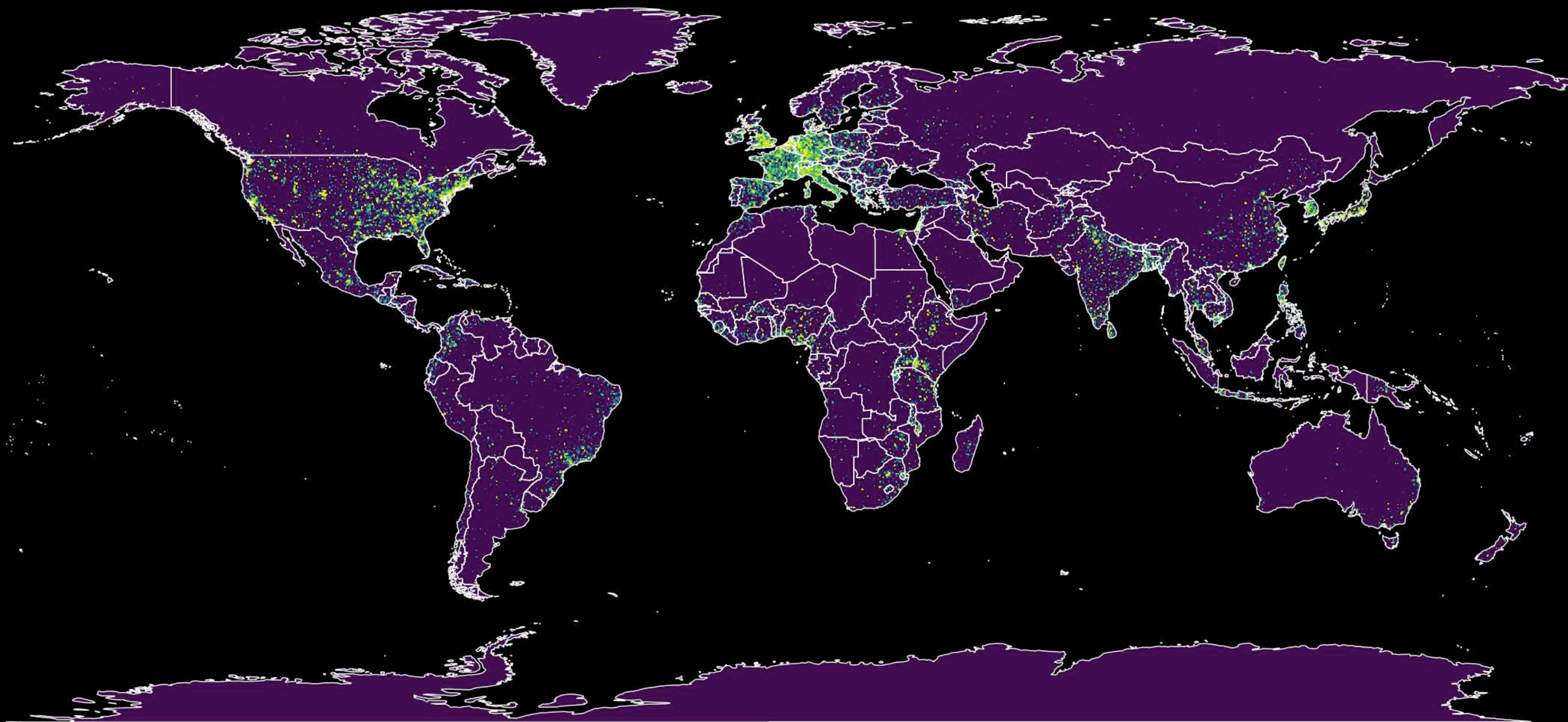
 **DISEASE
ONTOLOGY**

- Keywords
- Research articles
~150,000 articles

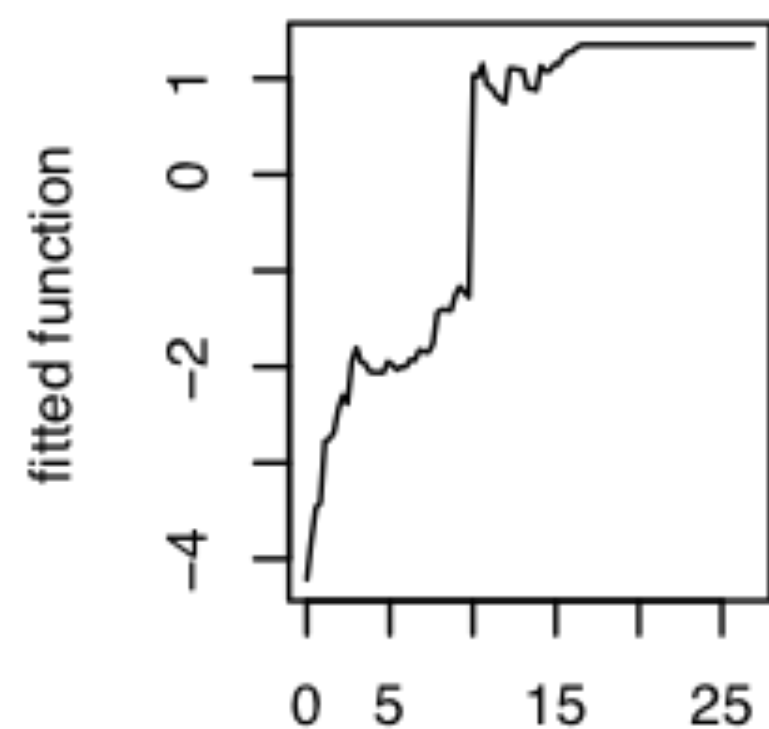


toponym resolution

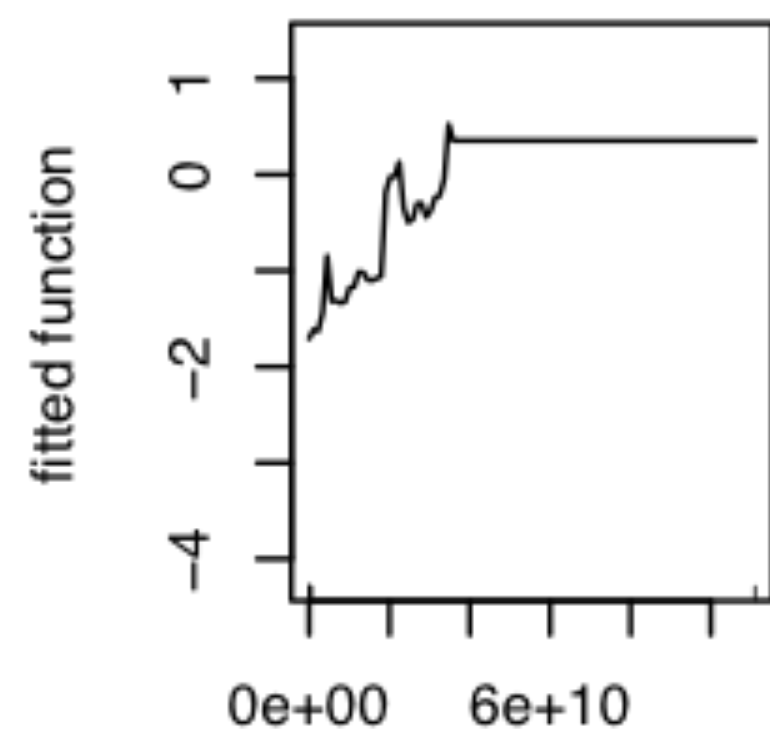
1. Match all potential locations
2. Score matches on various heuristics
3. Cull low-scoring matches
4. Select highest-scoring set of matches



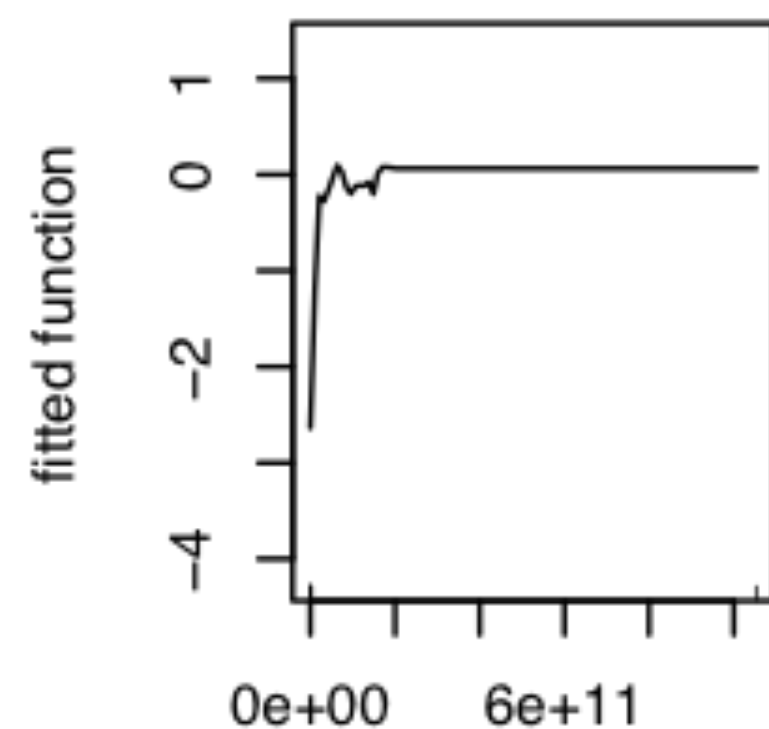
Matched toponyms, normalized by paper
and summed to a 1/4-degree grid



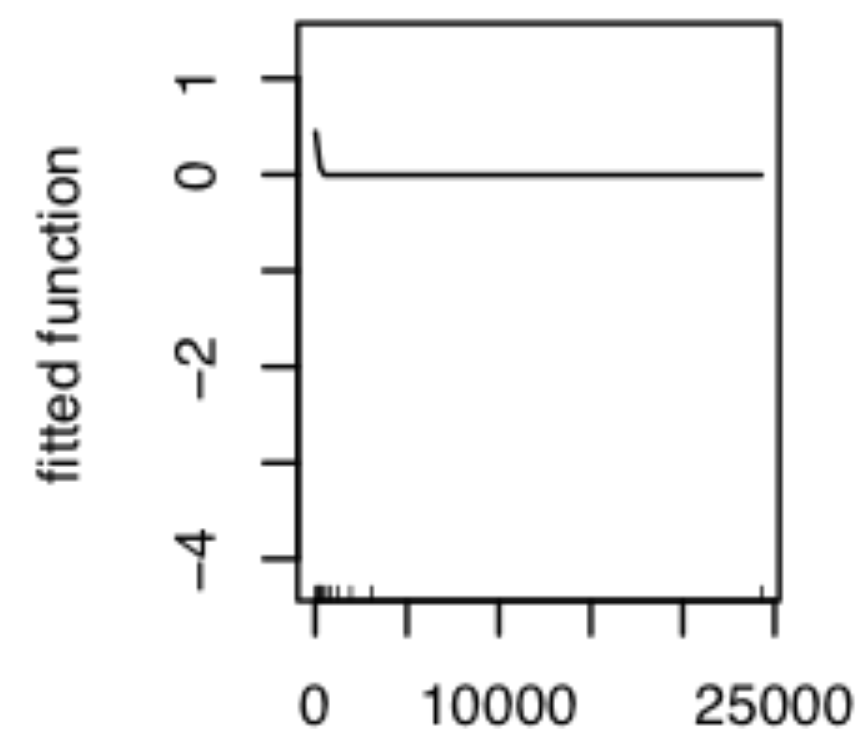
earth9_urban (45.2%)



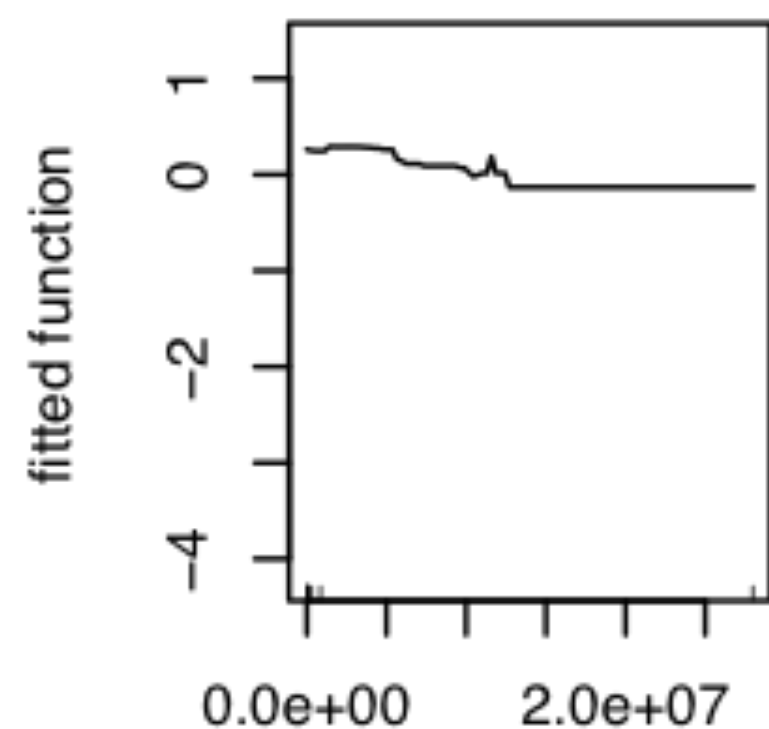
health_exp (30.5%)



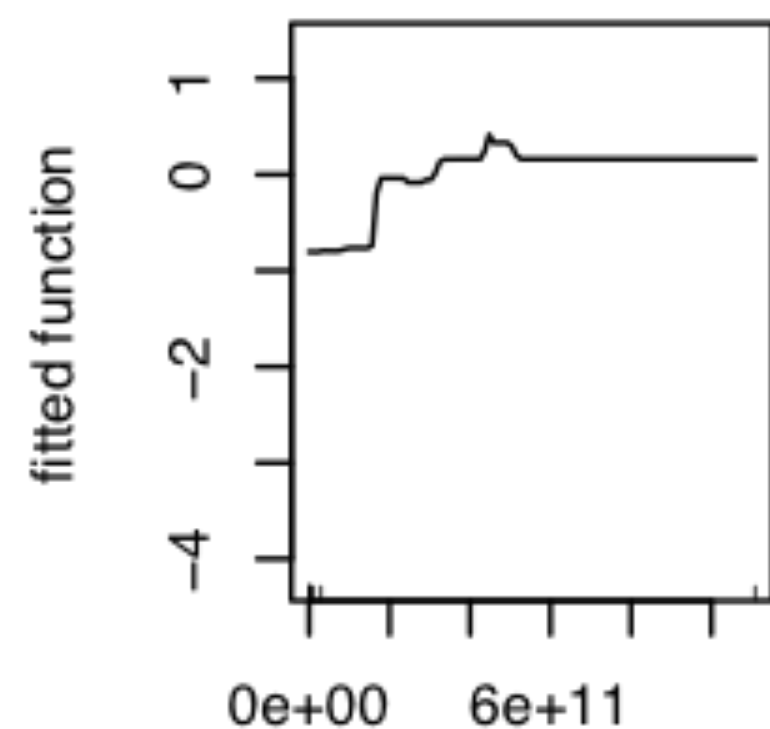
gdp (12.6%)



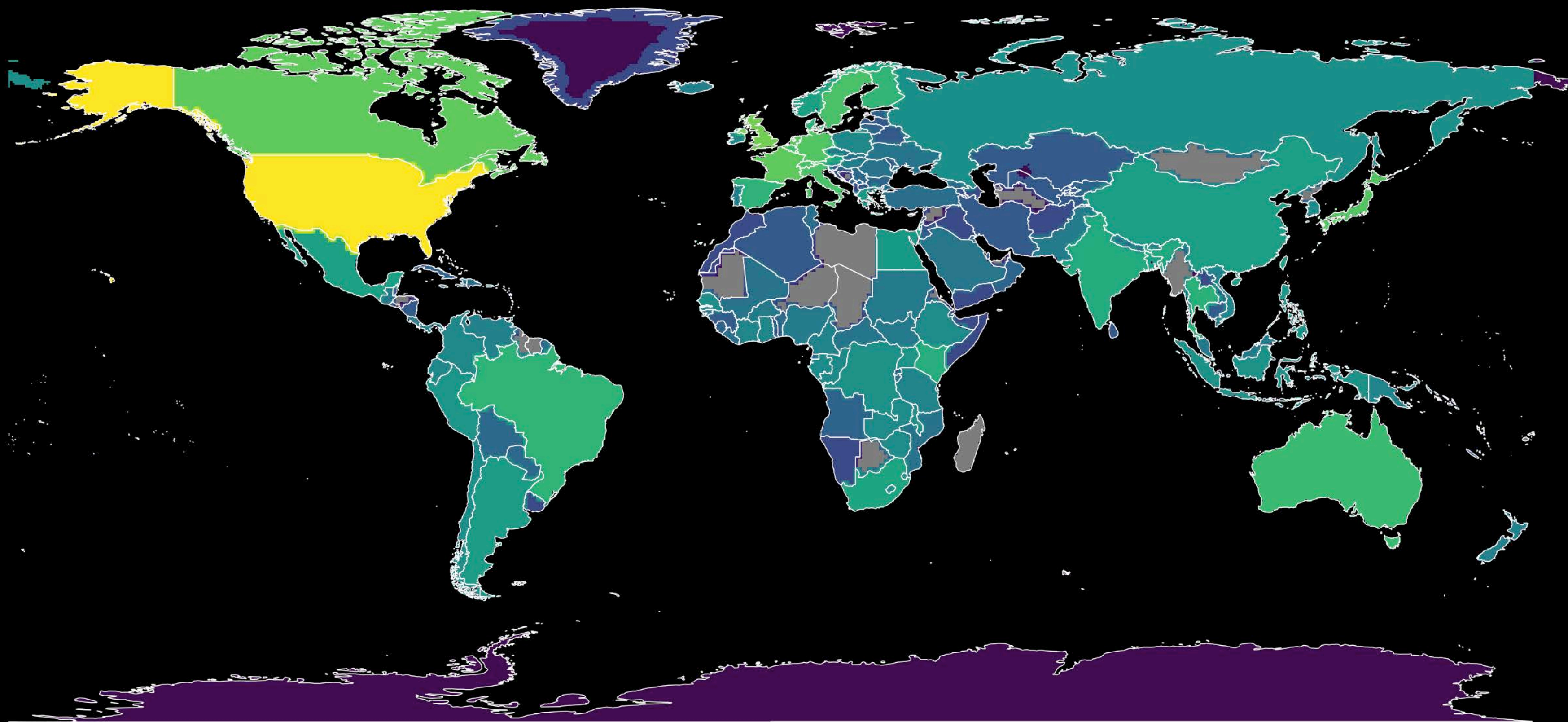
acc_50k (6.4%)



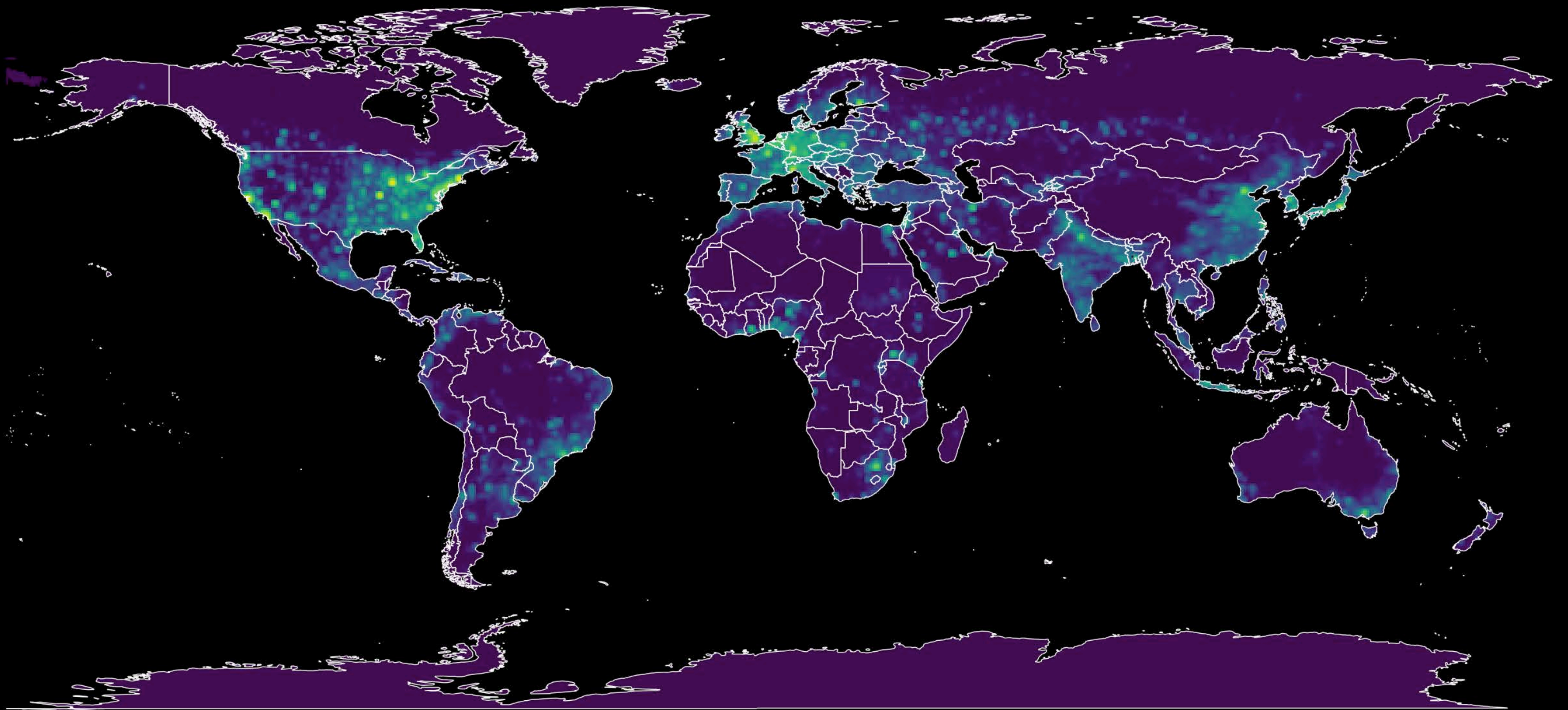
pop (3.6%)



dalys (1.7%)



Jones et al.'s measure of reporting effort



Output of PubCrawler BRT model

Open-source:

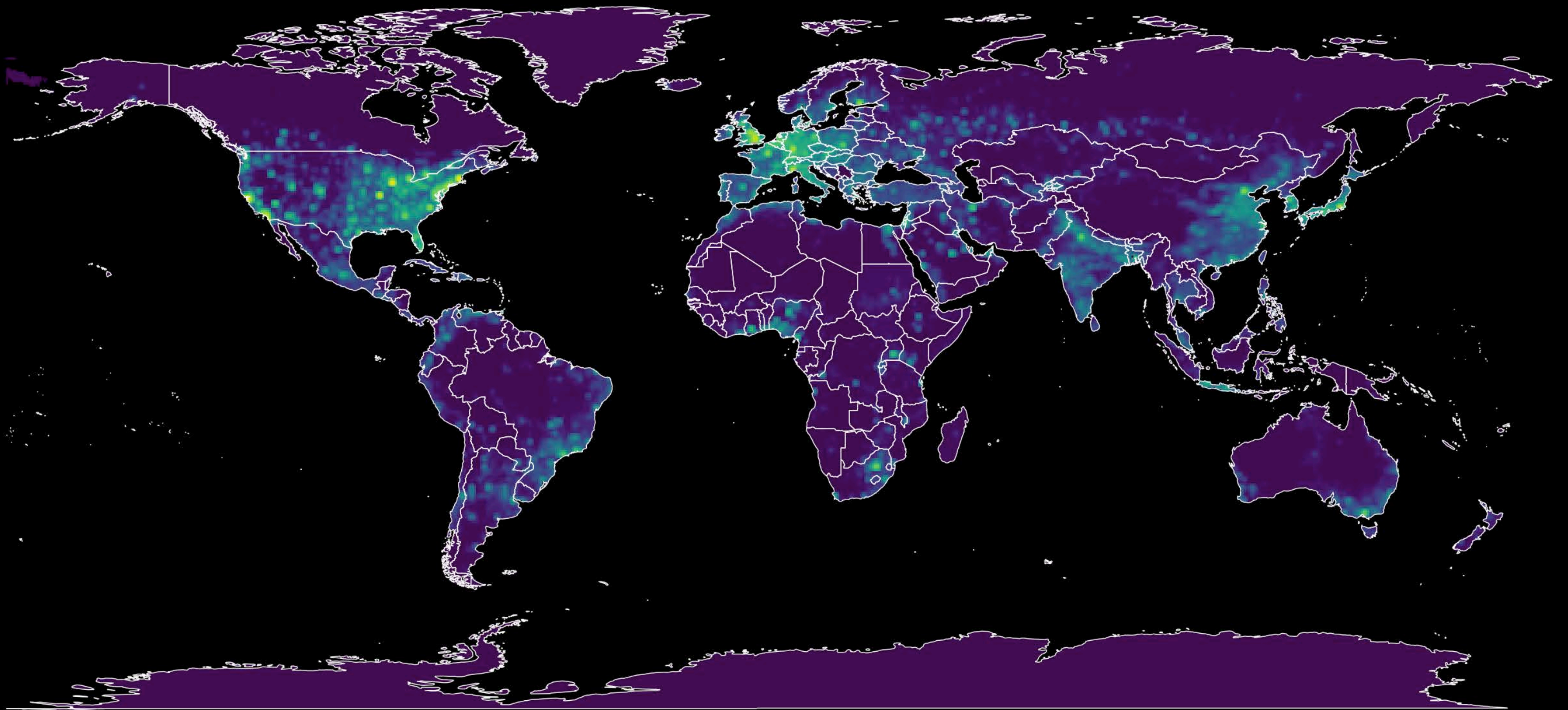
<https://github.com/ecohealthalliance/pubcrawler/>

(Please excuse the mess; under ongoing development)

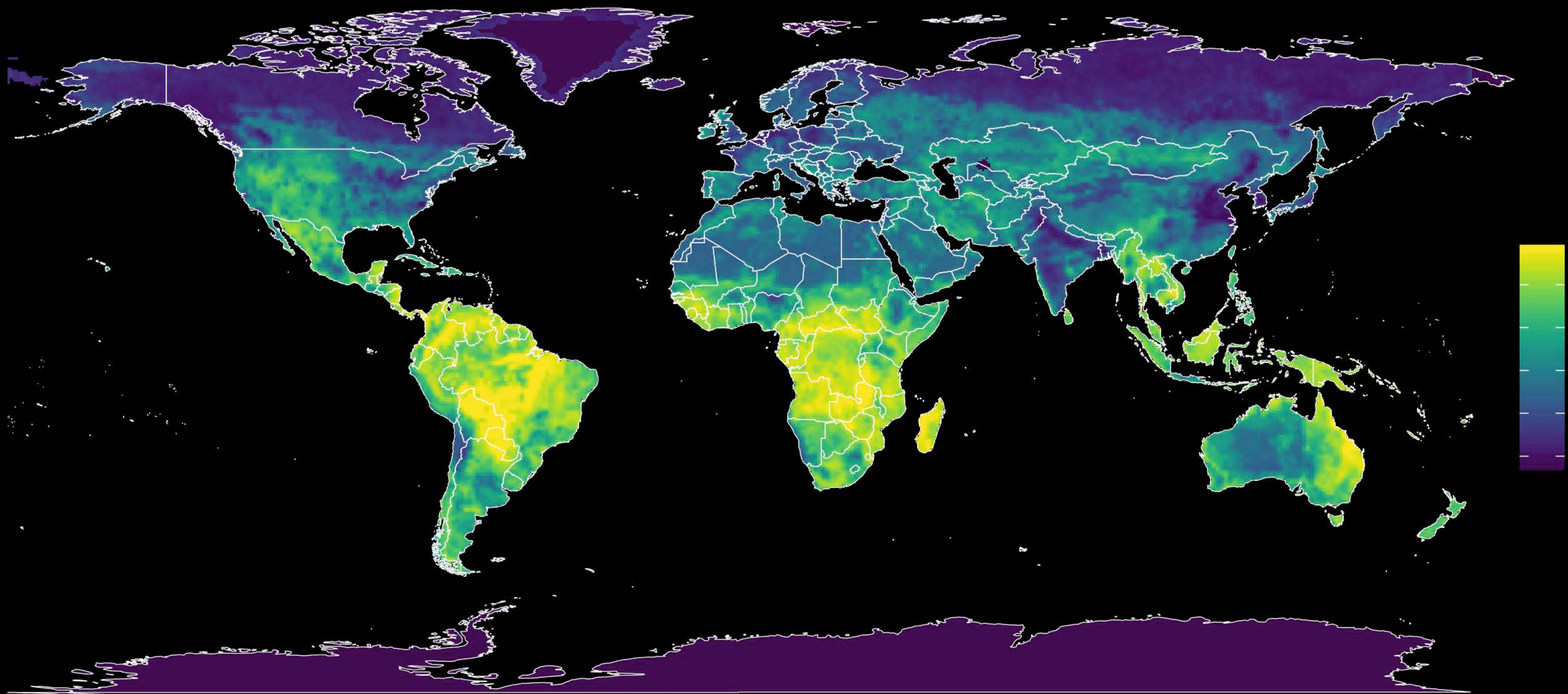
Modeling framework

- Quantify uncertainty: **resampling**
 - Bootstrap resampling (1000 iterations)
 - Exhaustive leave-one-out cross-validation (10 iterations per event)
- More robust model type: **boosted regression trees**

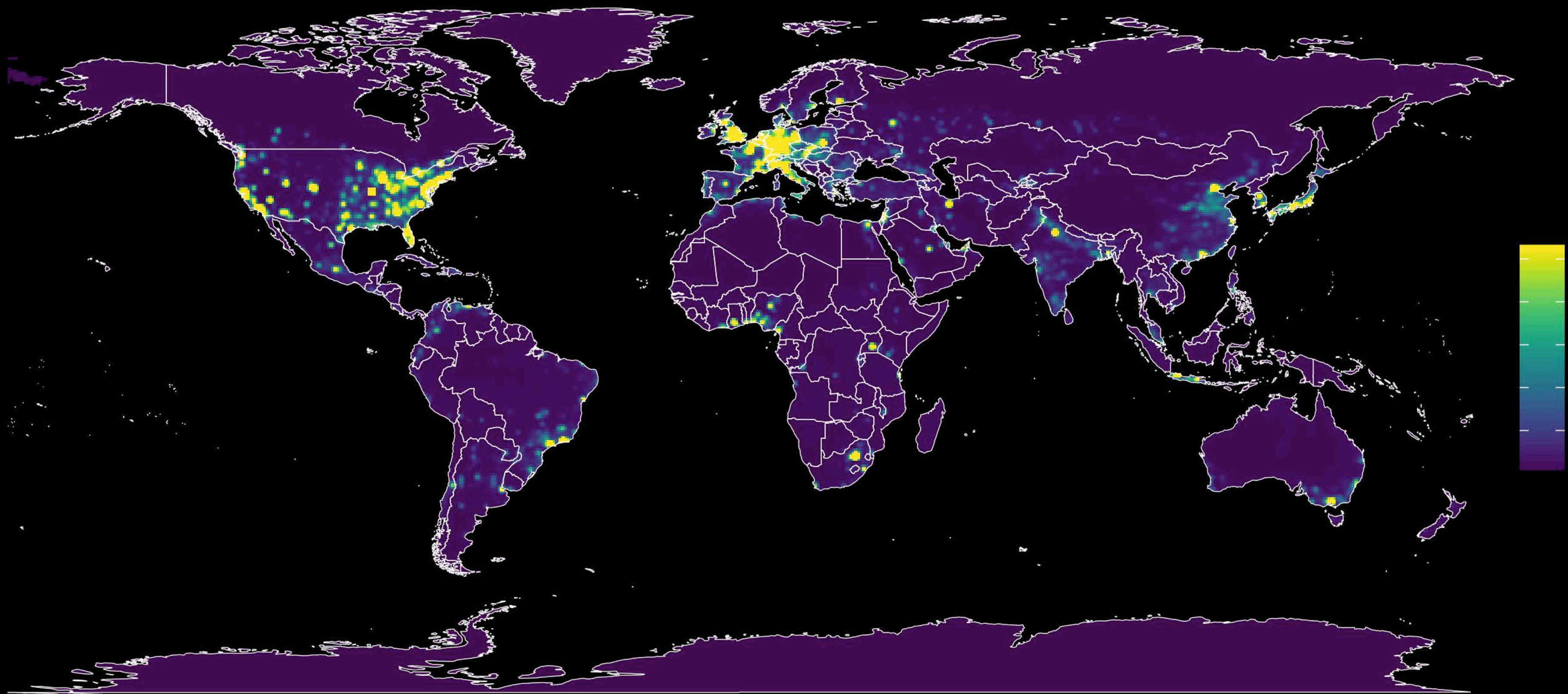
Results



Measure of reporting effort (PubCrawler)

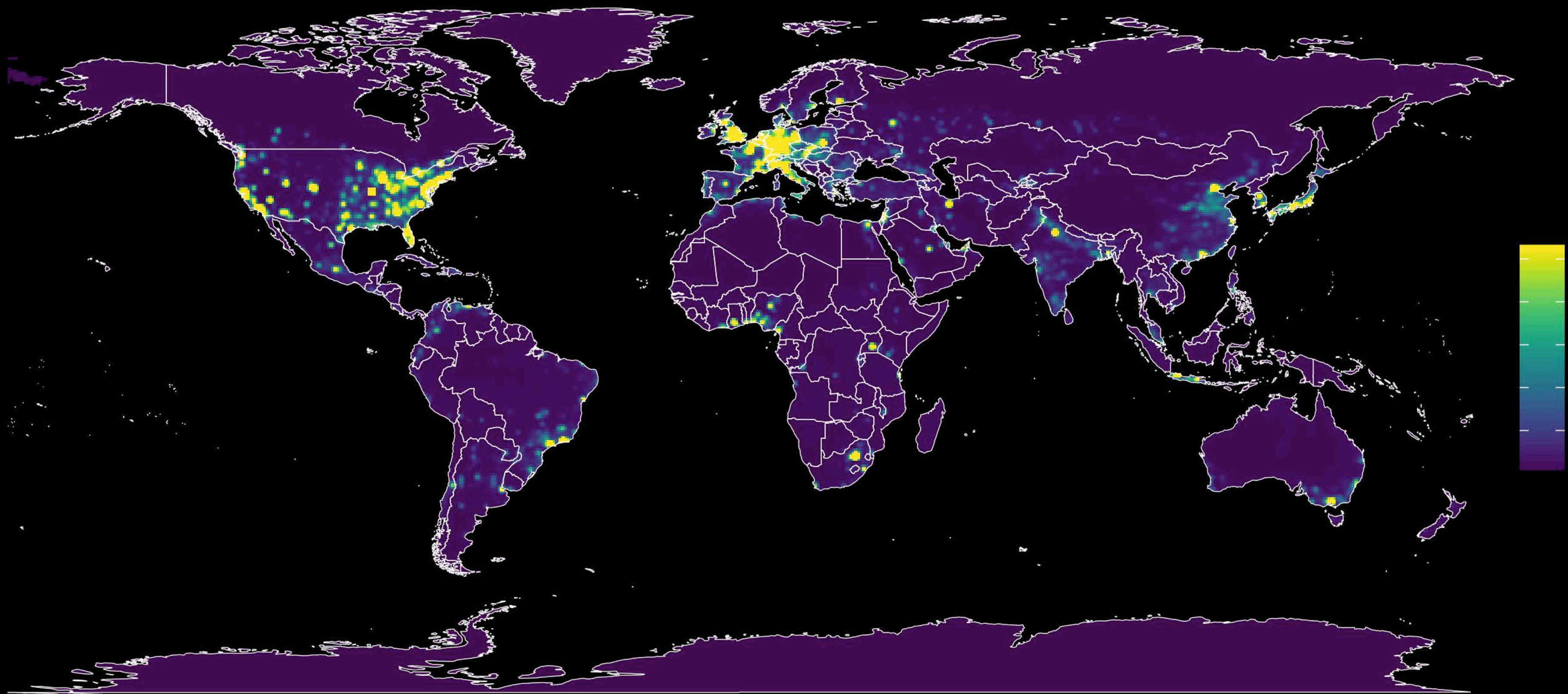


Output of BRT model, relative to reporting effort
response

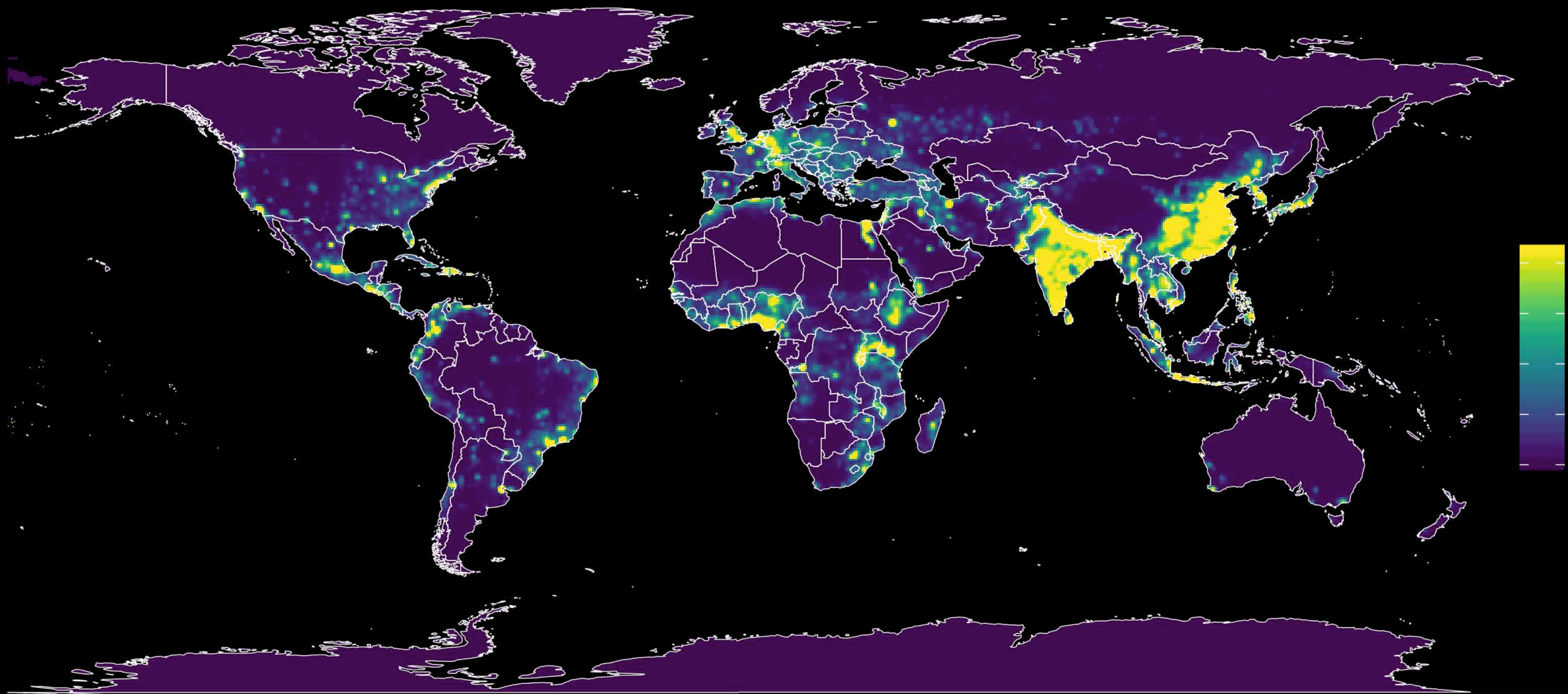


Model's predicted observations of EID events
response × reporting effort

bias = reporting effort /
population



EID risk index (observed)
response × reporting effort

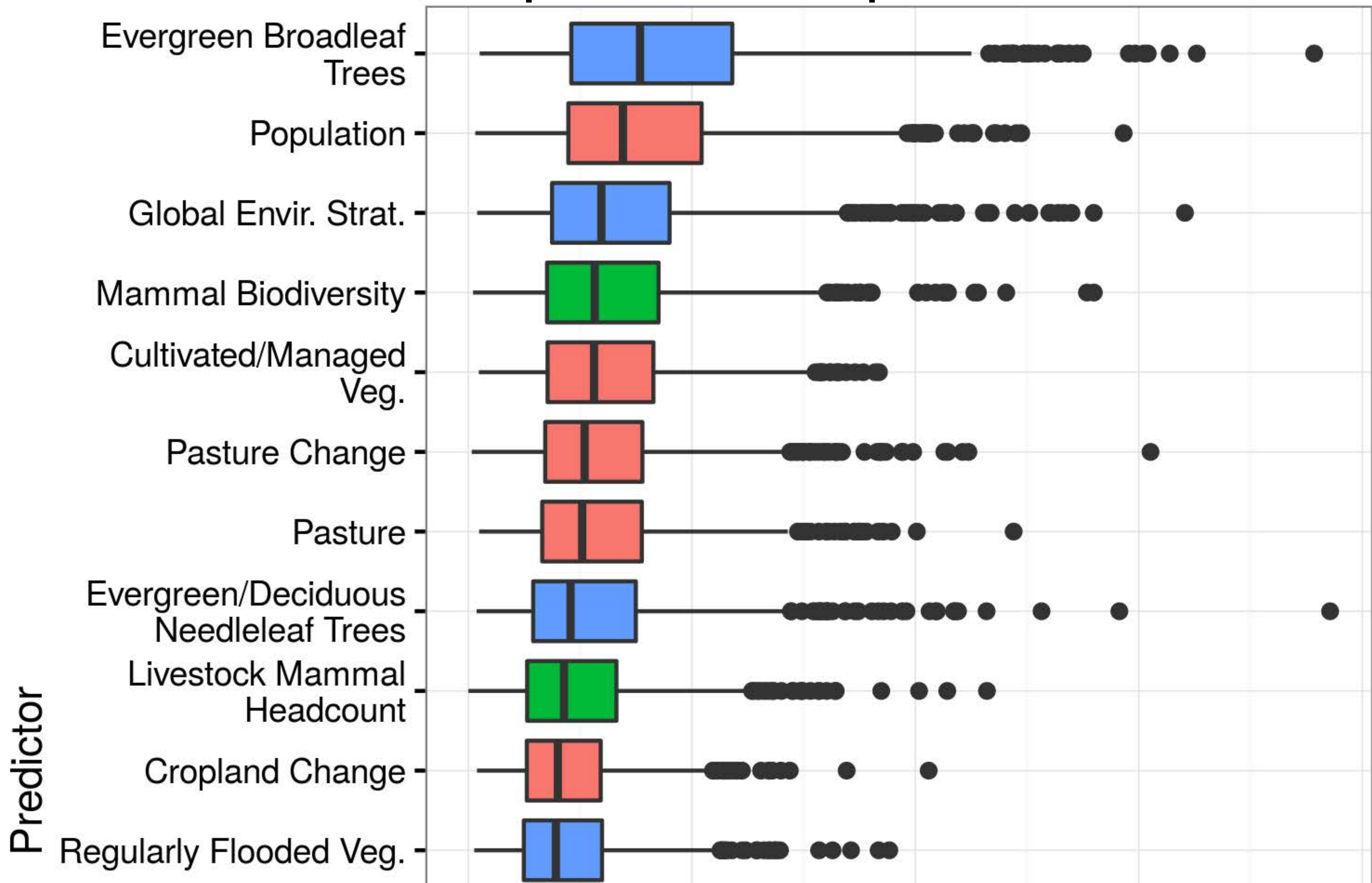


EID risk index (corrected for reporting bias)
response × population

Other model output
gets technical fast

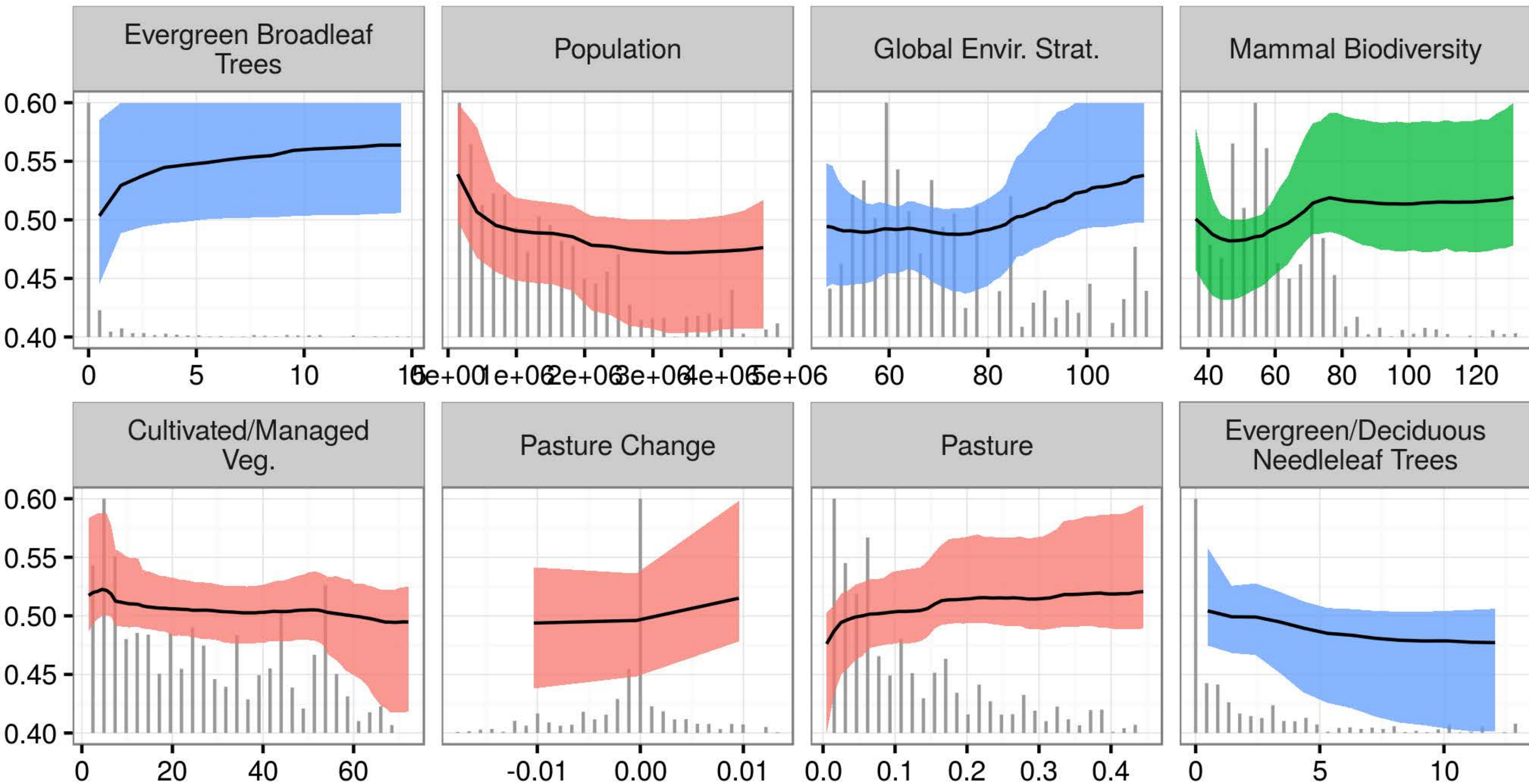
Relative influence of predictors

over 1000 replicate resampled models



Partial dependence plots

Effects of predictors, holding all others constant



Thanks!



EcoHealth Alliance