Environmental Determinants of Enteric Infectious Disease

PI: BENJAMIN ZAITCHIK, JOHNS HOPKINS UNIVERSITY
CO-I: MARGARET KOSEK, UNIVERSITY OF VIRGINIA
CO-I: HAMADA BADR, JOHNS HOPKINS UNIVERSITY
CO-I: JIM NELSON, BRIGHAM YOUNG UNIVERSITY
POSTDOC: JOSH COLSTON, UNIVERSITY OF VIRGINIA

Mortality

Globally, diarrhea kills 2,195 children every day

This is 1 out of 9 child deaths, worldwide

It is more than AIDS, malaria, and measles combined

It is the second leading cause of death in children less than five years old

Liu et al. (2012)

Morbidity

Impaired **cognitive development**

Stunting

Reduced **vaccine response**

https://borgenproject.org/what-causes-stunting/

EID are preventable and treatable

In some cases, **vaccines** are available

Improved Water, Sanitation and Hygiene (**WASH**) infrastructure and behavior is critical

Those suffering from diarrhea can be treated with **oral rehydration therapy**

Hector Retamal/AFP/Getty Images

Project goal

Establish the feasibility of Earth Observation-informed EID risk mapping, monitoring, and prediction systems

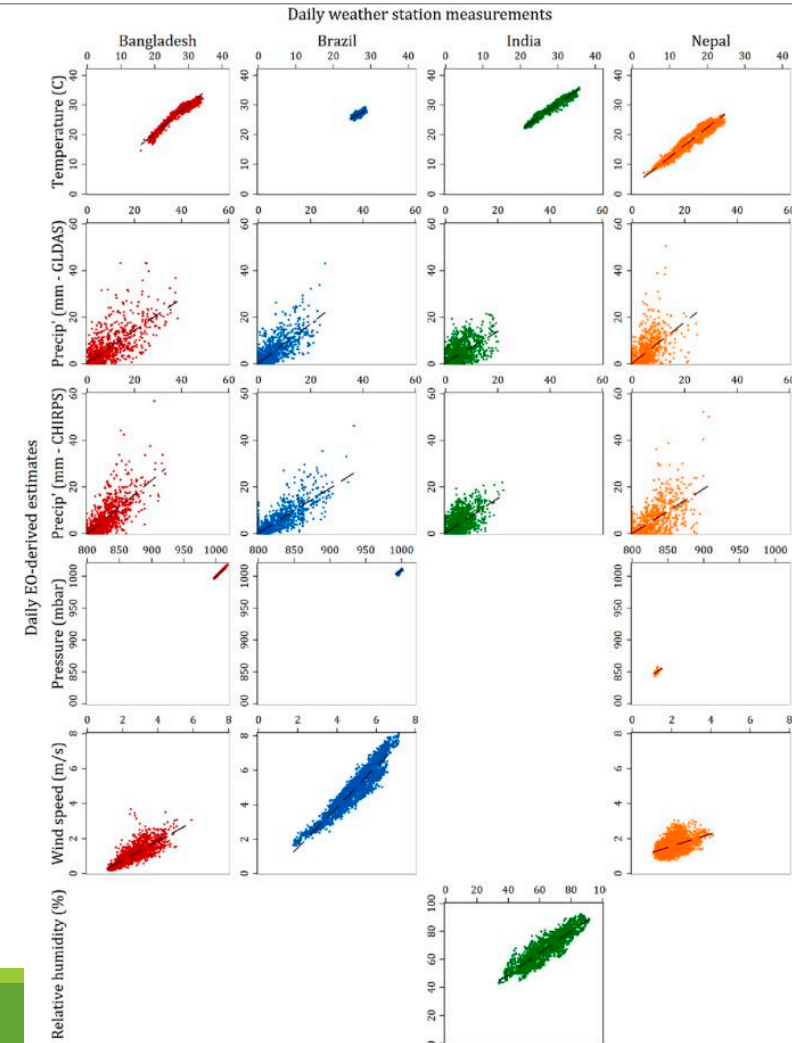We are doing this through collaboration with multiple EID studies performed at sites around the world

Current list of collaborating studies
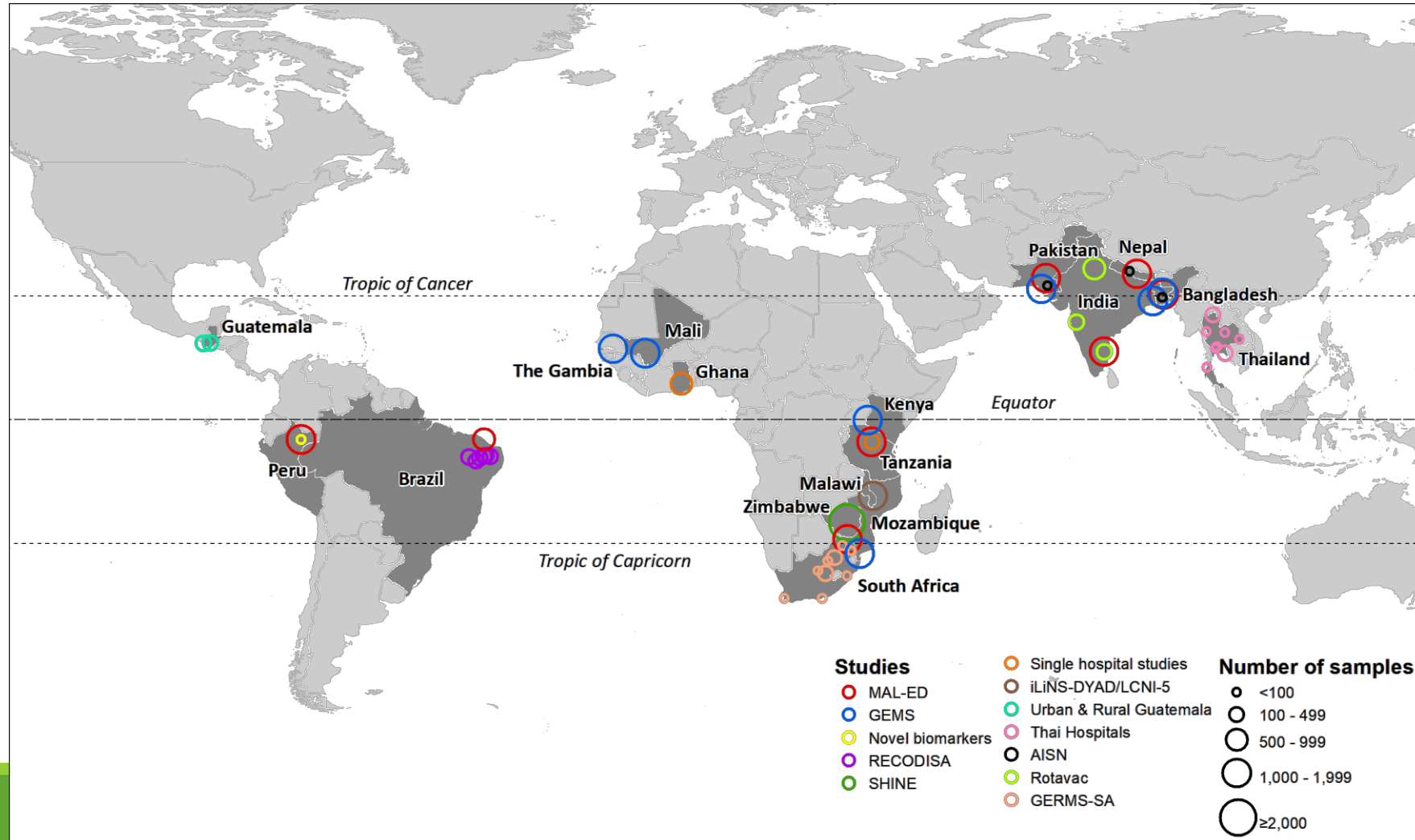
Earth Observation data

None of these infection studies included collection of data on climate or environment.

Earth Observations offer an opportunity to fill this gap.

Colston et al. (2018)
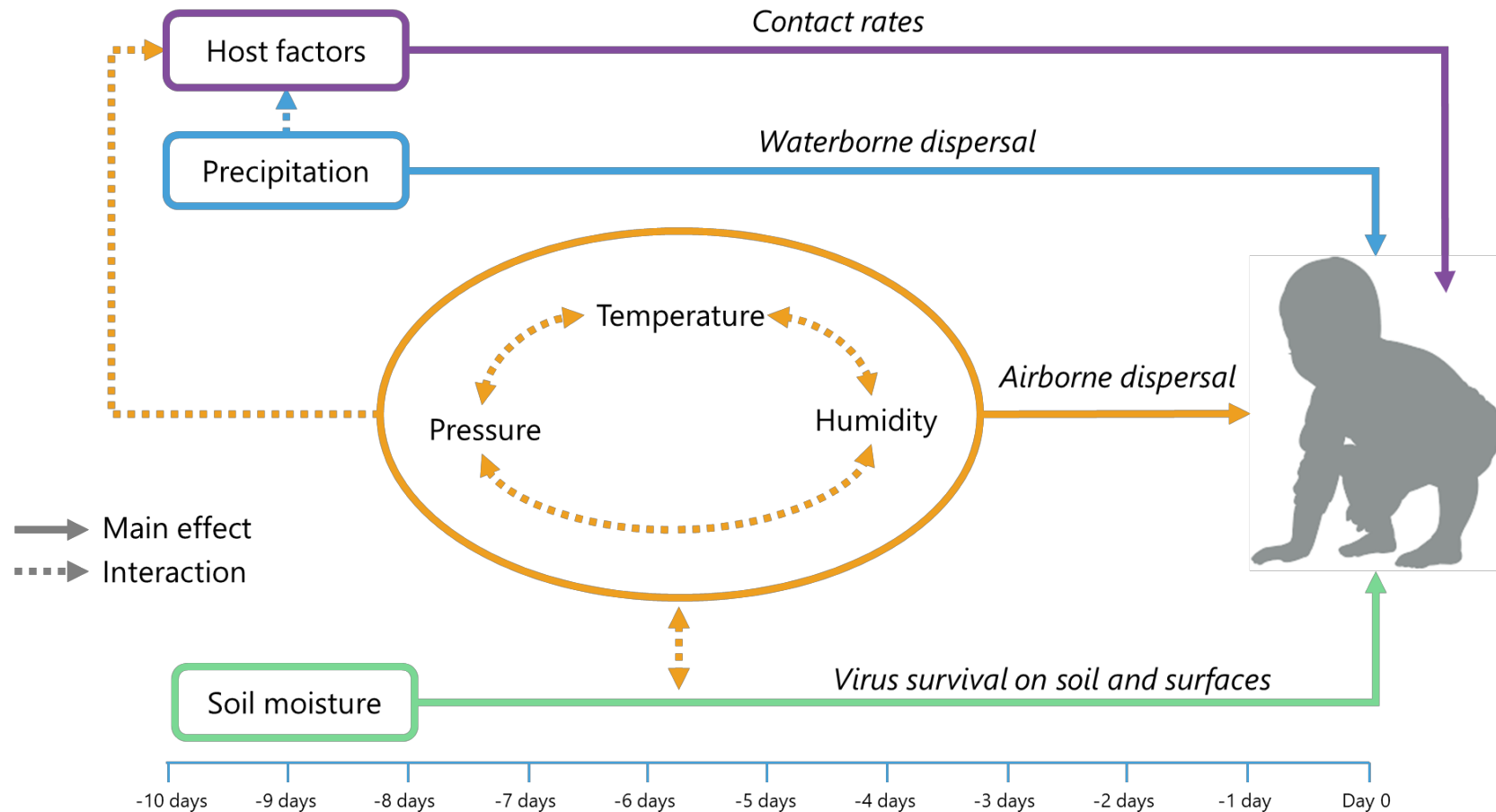
Accomplishments

PY1:

1. Evaluated EO performance at MAL-ED sites, and published results collaboratively with MAL-ED site PIs (Colston et al., 2018)

2. Generated a preliminary rotavirus prediction model based on MAL-ED site data and Earth Observations

PY2:

1. Published the results of the rotavirus model collaboratively with site PIs (Colston et al., 2019)

2. Performed preliminary regionalization based on rotavirus predictors

3. Built template visualization app in Tethys

4. Participated in NASA's pilot commercial data buy program

Rotavirus transmission pathways

Accomplishments

PY3:

1. Performed a targeted study of ENSO influence on EID at our Peru MAL-ED site (Colston et al., 2019)

2. Substantially enhanced our database of predictor variables

3. Nearly completed models of Shigella

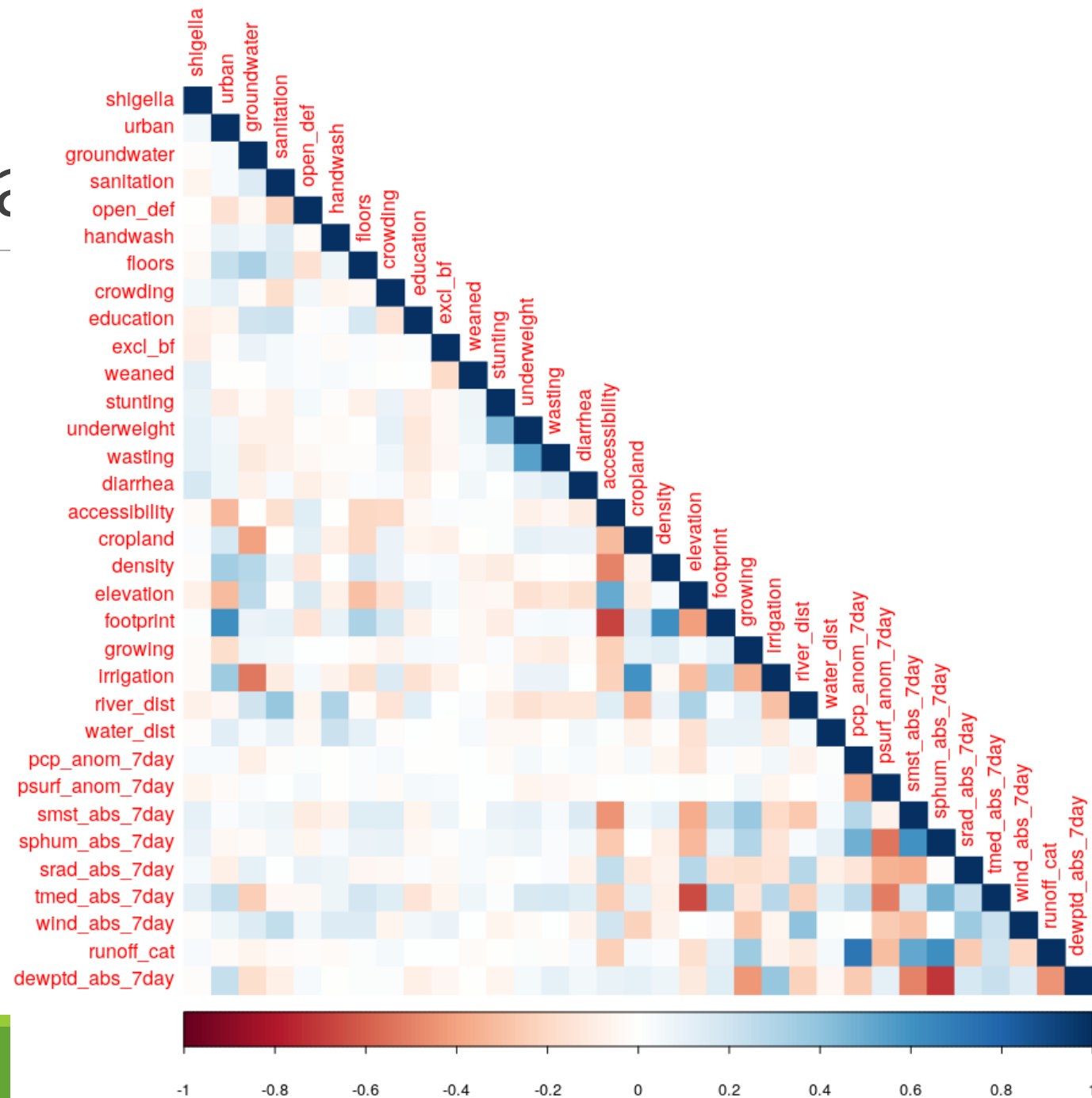4. Produced maps of Shigella risk for dissemination to partners

Impacts of the 2011-20[...]
on enteric infections i[...]

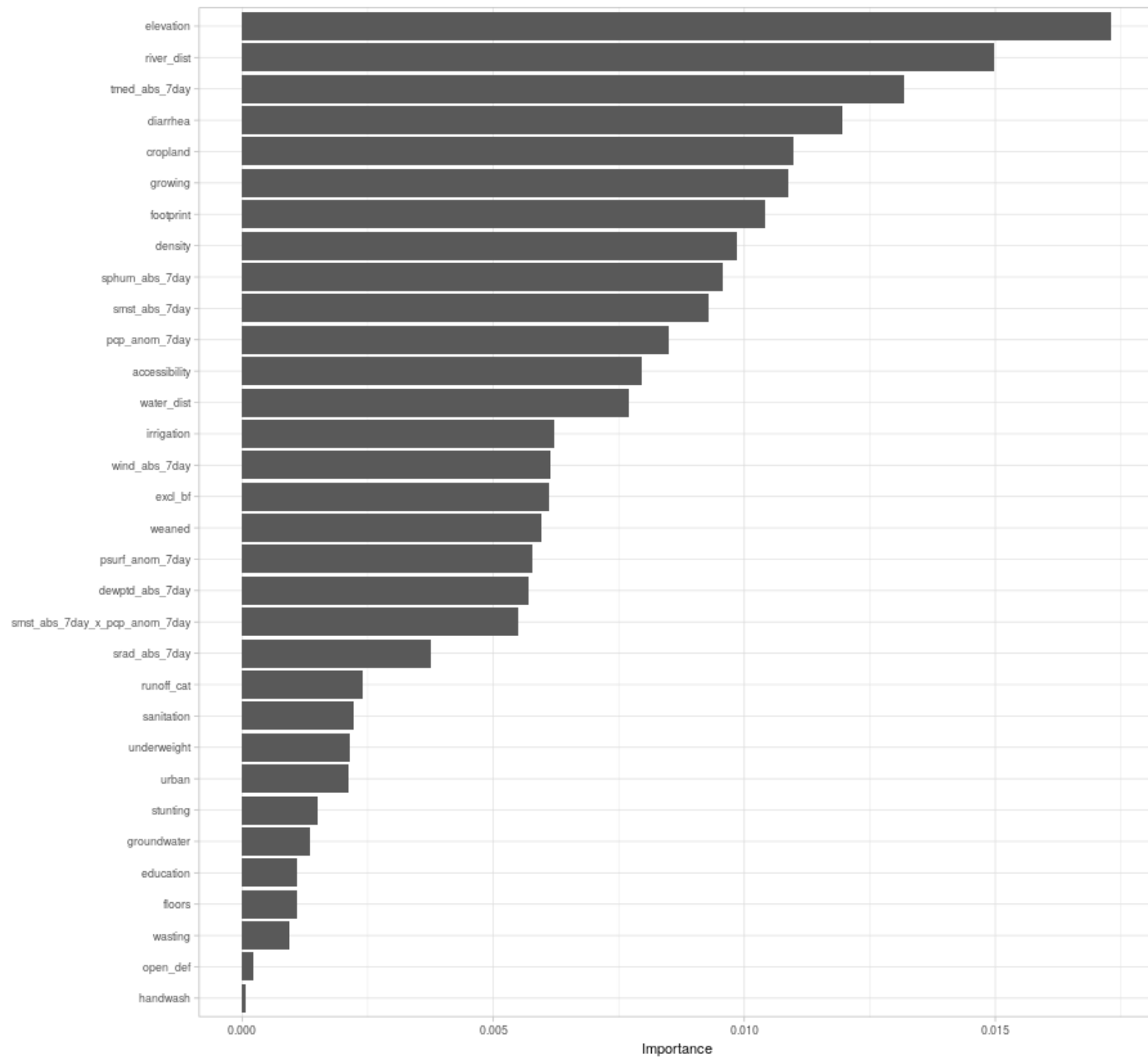Colston et al. (2020)
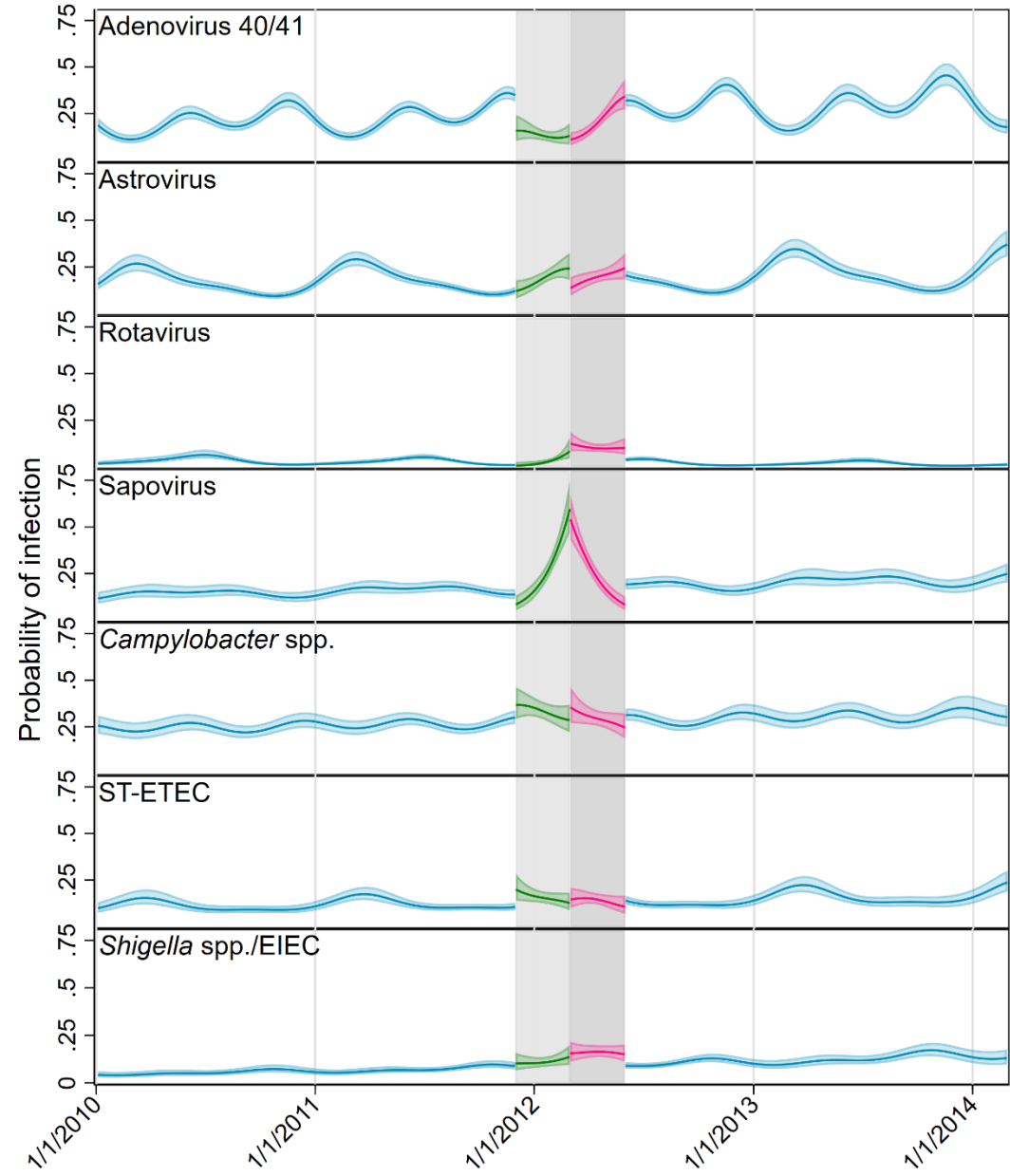
Shigella modeling results

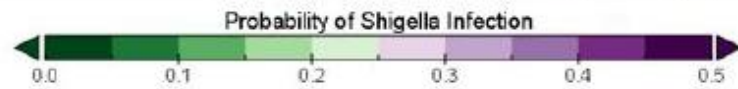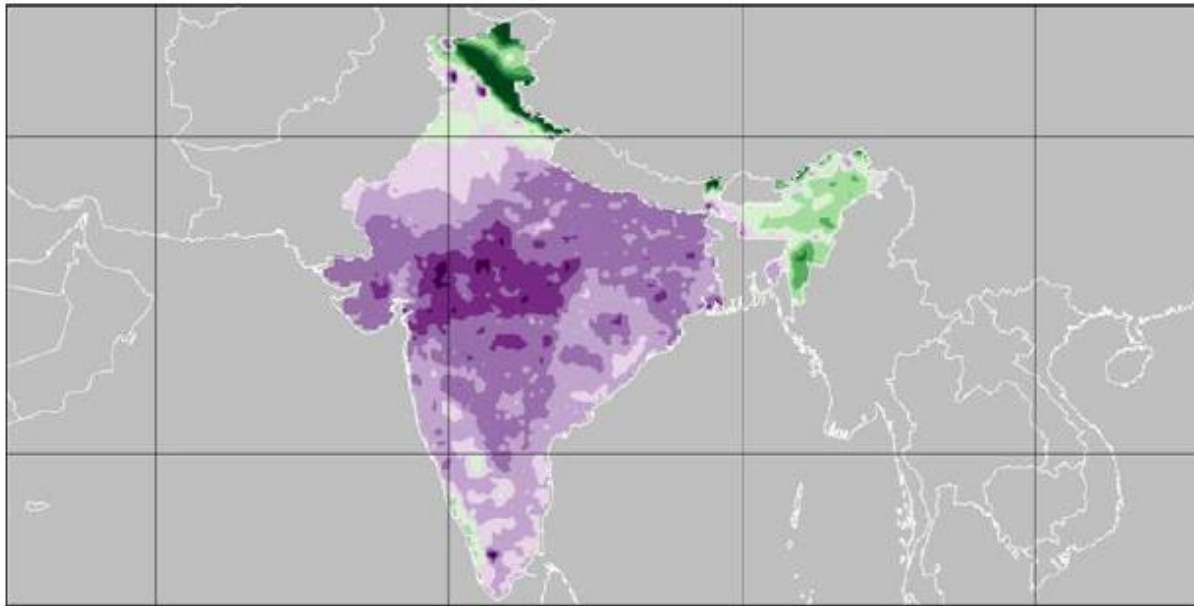Correlations between shigella and its predictors

Shigella

Variable Importance Plot

Impurity-corrected Random Forest;
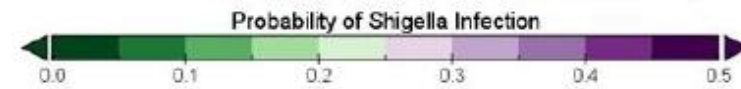Unconditional permutation scheme;
cross-validated results

Shigella risk maps

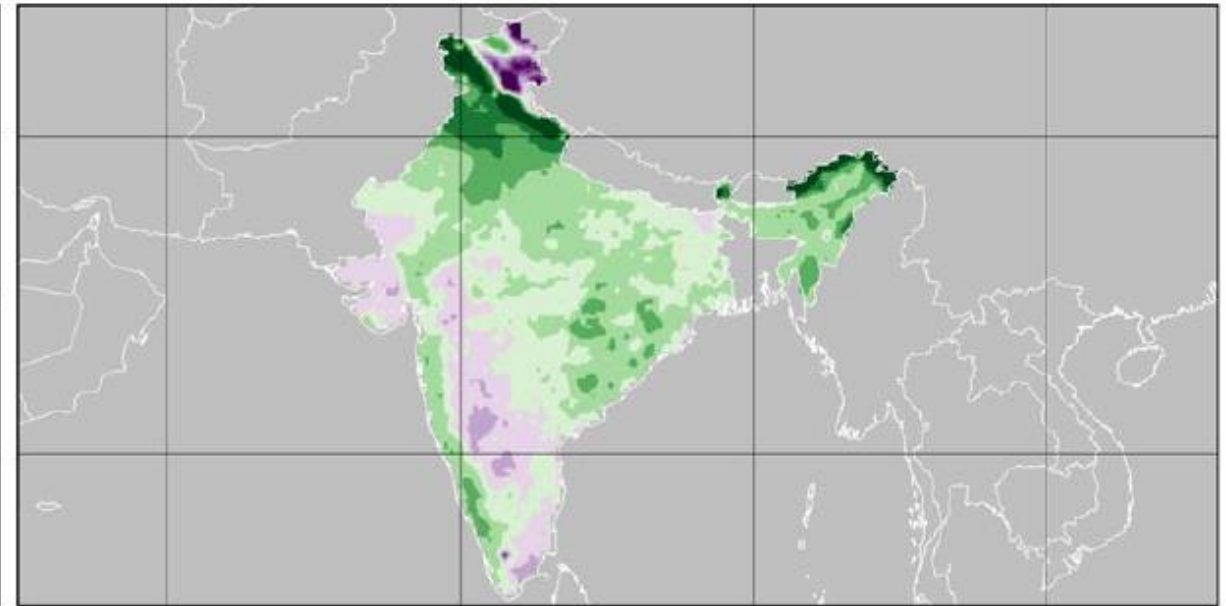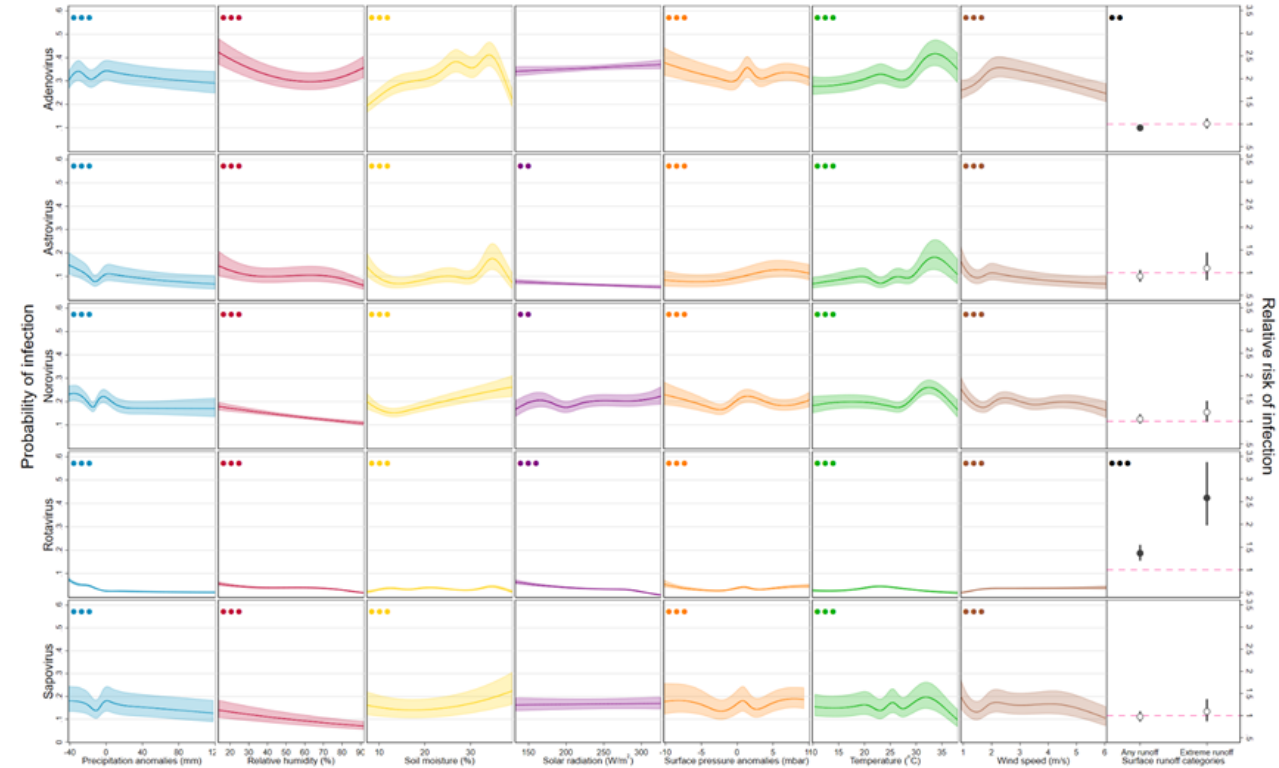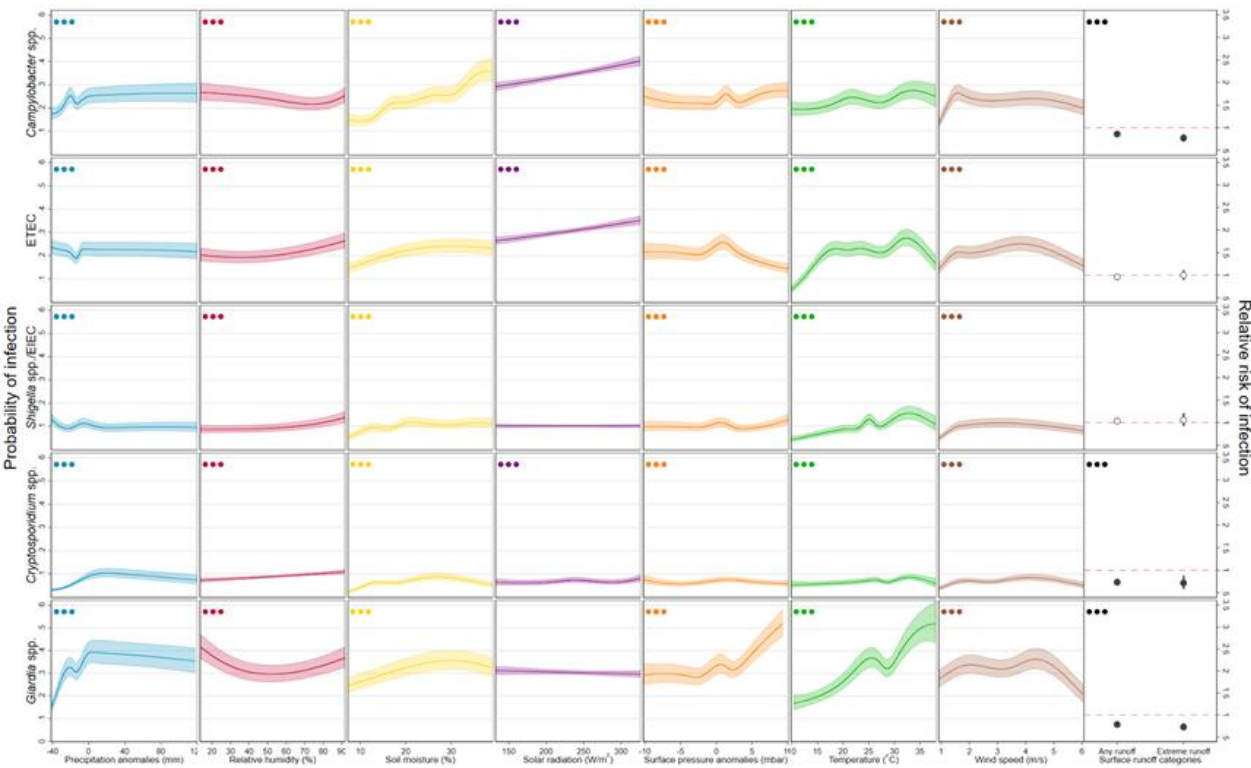# Understanding mechanism

Understanding mechanism

Precipitation RH Soil Moisture Solar Rad Surface Pres Temperature Wind Speed Runoff
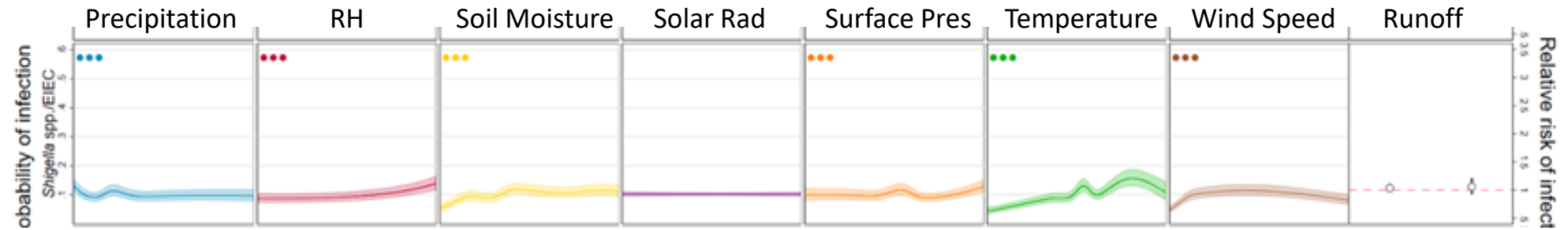
Timeline and Risks

	PY1				PY2				PY3			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Objective 1												
Perform and evaluate retrospective LDAS simulations	███	███										
Create unified database of EID predictors		███	███									
Develop and evaluate statistical EID models	███	███	███	███								
Objective 2												
Perform EID-specific regionalization			███	███	███							
Characterize regionalization uncertainty						███	███					
Objective 3												
Generate maps of EID potential by disease and season						███	███					
Implement monitoring/warning systems for selected EID								███	███	███		
Produce projections of future EID potential								███	███	███		
Objective 4												
Create Tethys app for display and analysis of EID database	███	███										
Integrate HiClimR to Tethys			███	███								
Present preliminary system to MAL-ED community								███				
Refine and operationalize system								███	███	███	███	███

Risks

COVID has slowed our research and communication with health system partners

Slower feedback might limit the number of pathogens we can address by the end of the project

ARL

Current: ARL 5 for Shigella / ARL 4 for others

Expectation: ARL 7 for some pathogens by early 2021, but further COVID-related slippage is possible

Goal: ARL 7
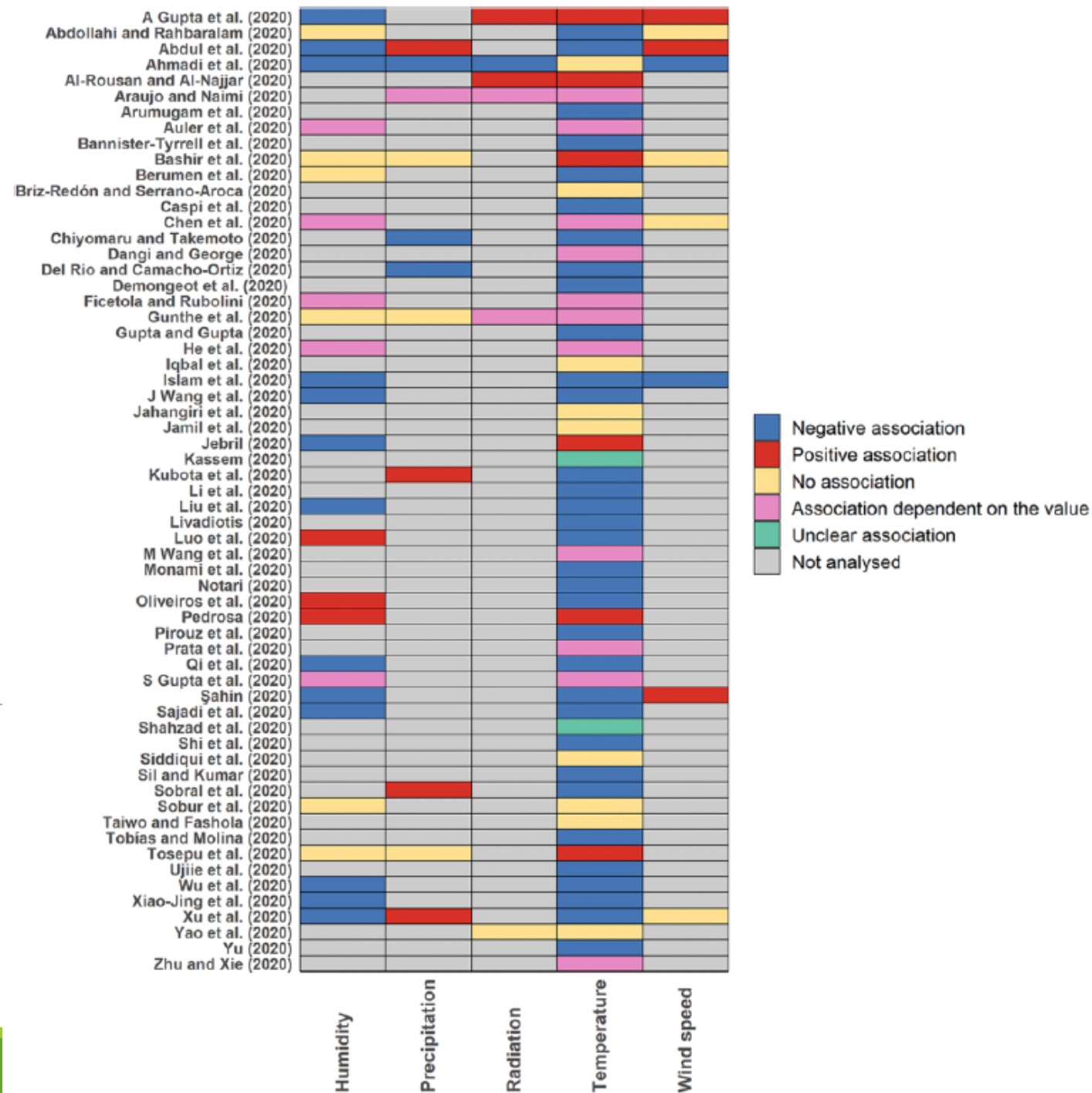
Why?

A short and unreliable COVID-19 data record

Inconsistent and sometimes inappropriate definition of response variable

Inconsistent and sometimes inappropriate scales of analysis

Difficulty of accounting for non-meteorological predictors: behavior, policy, demographics, cultural practices, etc.

Differences between climate zones

Diverse and sometimes questionable methodologies

Challenge of isolating climate influence early in the pandemic

Creating a unified, reliable data record

United States:

US	36	061	10476
Admin 0	Admin 1	Admin 2	Admin 3
Country	State	County	District
ISO 3166 1	**FIPS**	**FIPS**	**ZCTA**
2 letters	+ 2 digits	+ 3 digits	+ 5 digits

Europe:

DE	2	1	H
Admin 0	Admin 1	Admin 2	Admin 3
Country	State *	County **	District
ISO 3166 1	**NUTS 1**	**NUTS 2**	**NUTS 3**
2 letters	+ 1 digit/letter	+ 1 digit/letter	+ 1 digit/letter

Global:

AU	ACT		
Admin 0	Admin 1	Admin 2	Admin 3
Country	Province/State	County	District
ISO 3166 1	**ISO 3166 2**	**Local 2**	**Local 3**
2 letters	principal divisions	country specific	country specific

*   NUTS 1 level represents groups of subregions (or equivalent) for some European countries (e.g., Italy).
** NUTS 2 level represents subregions (or equivalent) for some European countries (e.g., Italy).

https://github.com/hsbadr/COVID-19

- Maps all geospatial units globally into a unique standardized ID.
- Standardizes administrative names and codes at all levels.
- Standardizes dates, data types, and formats.
- Unifies variable names, types, and categories.
- Merges data from all credible sources at all levels.
- Cleans the data and fixing confusing entries.
- Integrates hydrometeorological variables at all levels.
- Optimizes the data for machine learning applications.

Creating a unified, reliable data record

red = Admin0, blue = Admin1, green >= Admin2

Selecting a response variable
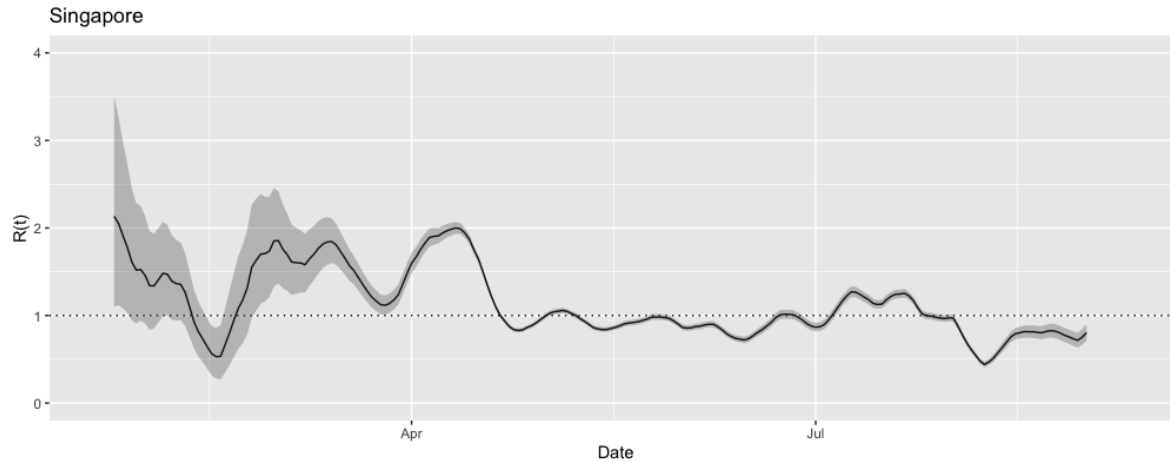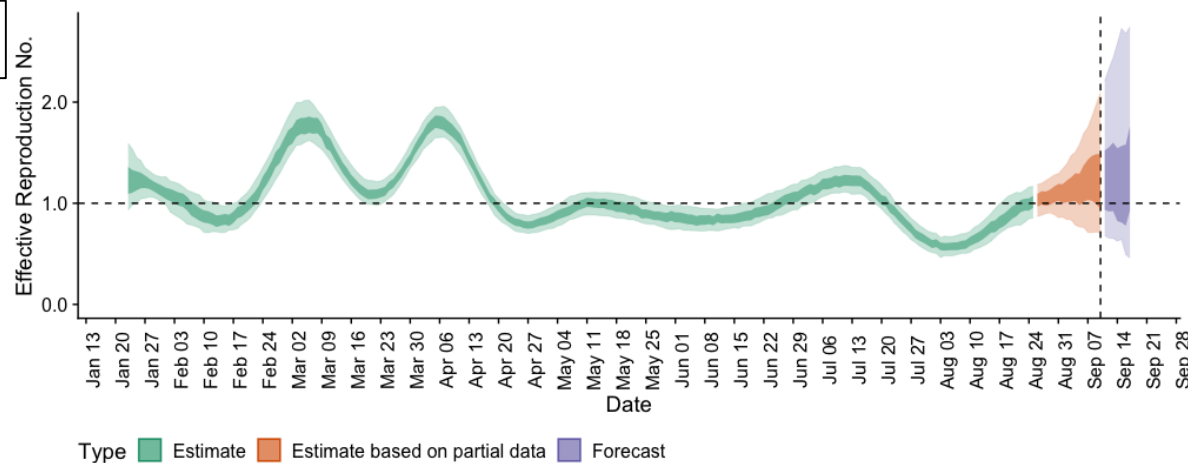
Following graphs were generated in R using the following epidemiological parameters:

- Daily new case data (JHU CSSE)
- Serial interval, incubation period, generation time
- Reporting delay
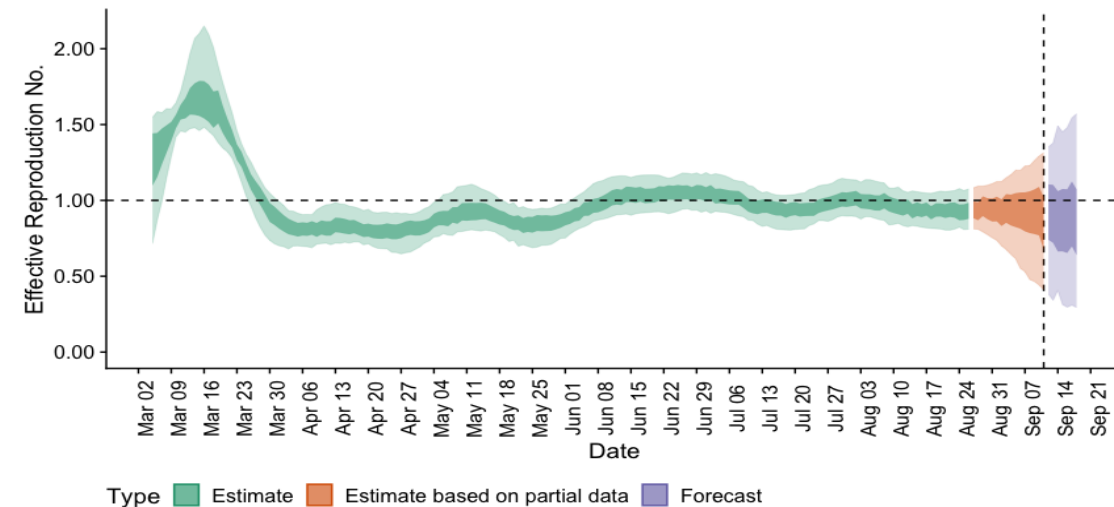
Hydrometeorological data

We have a "long list" of EO to incorporate, but for now we are focused on hydrometeorology drawn from reanalysis (NLDAS, MERRA2/FIPS and ERA5) and satellite-derived data (GPM)

Daily, spatially averaged to unit of COVID-19 case data in the harmonized database

Non-environmental covariates

Demographics

	cases	deaths	excess deaths	population	pop. Density	incidence score	Accessibility	intervention data	Facebook Mobility	Google mobility
Colombia	2			2	2	2	2	*1	2	1
Peru	3	3	0*	3	3	3	3	*0	2	1
Chile	2		1*	2	2	2	2	*~1	1	1
Ecuador	1	1	1	1	1	1	1	*1	1	1

District-level Population Density per square meter in Peru

ECUADOR

Pacific Ocean

N

0 55 110   220   330   440
                              Kilometers

Population density
(per square meter)
- 13 - 300
- 301 - 600
- 601 - 1000
- 1001 - 2000
- 2001 - 3000
- 3001 - 5000
- 5001 - 2900000

Source 1: The Humanitarian Data Exchange
Source 2: WorldPop

Made by: Yen Ting, Chen
M.Ed student in Department of Kinesiology
University of Virginia

Esri, HERE, Garmin, (c) OpenStreetMap contributors, and the GIS user community

Non-environmental covariates

Demographics

Non-pharmaceutical interventions

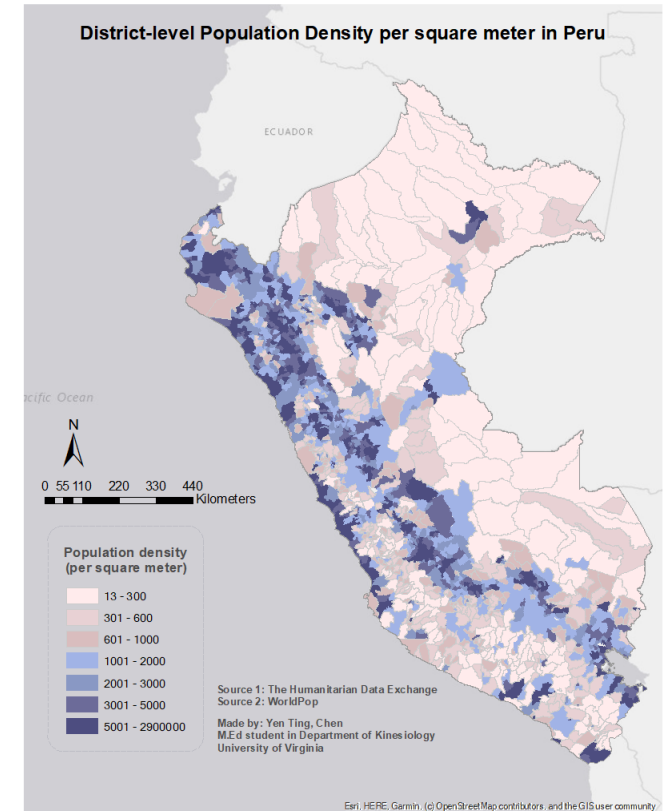## scientific data

Explore our content ⌄   Journal information ⌄

nature > scientific data > data descriptors > article

Data Descriptor | Open Access | Published: 27 August 2020

HIT-COVID, a global database tracking public health interventions to COVID-19

Qulu Zheng, Forrest K. Jones, Sarah V. Leavitt, Lawson Ung, Alain B. Labrique, David H. Peters, Elizabeth C. Lee, Andrew S. Azman ✉ & HIT-COVID Collaboration

Scientific Data **7**, Article number: 286 (2020) | Cite this article

District-level Population Density per square meter in Peru

Population density (per square meter)
- 13 - 300
- 301 - 600
- 601 - 1000
- 1001 - 2000
- 2001 - 3000
- 3001 - 5000
- 5001 - 2900000

Source 1: The Humanitarian Data Exchange
Source 2: WorldPop

Made by: Yen Ting, Chen
M.Ed student in Department of Kinesiology
University of Virginia
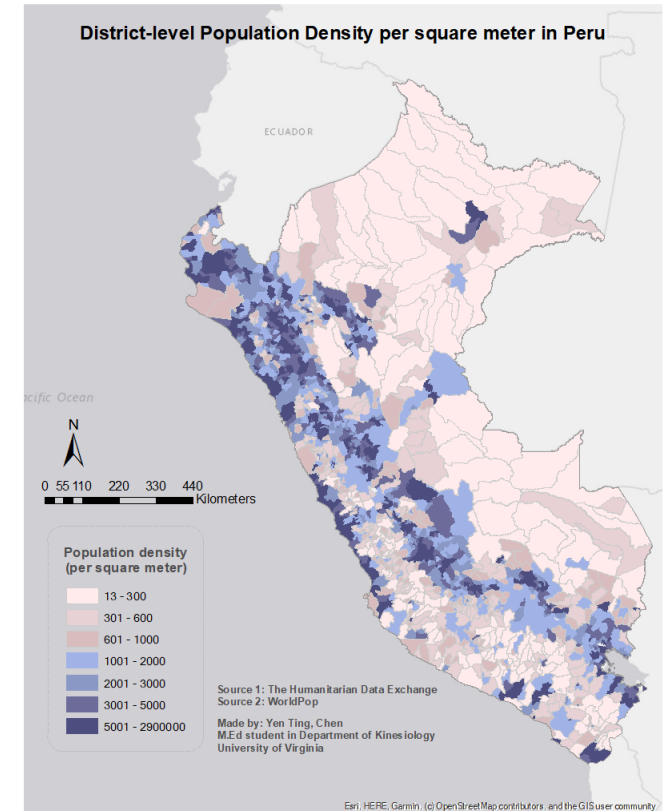
Non-environmental covariates

Demographics

Non-pharmaceutical interventions

Comorbidity:
◦ diabetes, obesity, HIV, hypertension, smoking, COPD, cardiovascular disease index all compiled at Admin0 globally and Admin1 for the US

Mobility
◦ High resolution mobility data for the US and selected European countries
◦ International air travel data

District-level Population Density per square meter in Peru

ECUADOR

Pacific Ocean

N

0 55 110   220   330   440
                          Kilometers

Population density (per square meter)
- 13 - 300
- 301 - 600
- 601 - 1000
- 1001 - 2000
- 2001 - 3000
- 3001 - 5000
- 5001 - 2900000

Source 1: The Humanitarian Data Exchange
Source 2: WorldPop

Made by: Yen Ting, Chen
M.Ed student in Department of Kinesiology
University of Virginia

Esri, HERE, Garmin, (c) OpenStreetMap contributors, and the GIS user community

Current analyses:

1. Global analysis at national level, covering ~187 countries

2. Higher-resolution global analysis at Admin 1

3. Detailed analysis for selected countries: US, Germany, Italy, Colombia, Peru, Ecuador, Chile

Collaboration across the AST!

Our EO database has been leveraged for:

◦ COVID-19 stay at home orders and heat-related illness by Suwei Wang (Julia Gohlke's group)

◦ County-level analysis of COVID-19 risk for the US by Bill Pan's group