SEASONALITY AND HYDROMETEOROLOGICAL PREDICTORS OF ROTAVIRUS
INFECTION IN AN EIGHT-SITE BIRTH COHORT STUDY:
*IMPLICATIONS FOR MODELING AND PREDICTING PATHOGEN-SPECIFIC ENTERIC
DISEASE BURDEN*

by
Josh Michael Colston, MSc.

A dissertation submitted to Johns Hopkins University in conformity with the
requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland
May 2018

Abstract

Improving understanding of the pathogen-specific seasonality of enteric infections is critical to informing policy on the timing of preventive measures and to forecasting trends in the burden of diarrhoeal disease. Longitudinal and time-series analyses are needed to characterize the associations between hydrometeorological parameters and pathogen-specific enteric infectious disease (EID) outcomes. Data obtained from active surveillance of cohorts can capture the underlying infection status as transmission occurs in the community. However, there is a need for an approach that can be systematically applied to evaluate the combined impact of multiple meteorological exposures at a level of spatiotemporal disaggregation sufficient to characterize potential lag effects, interactions and non-linearity. Earth Observation (EO) climate data products derived from satellites and global model-based reanalysis have the potential to be used as surrogates in situations and locations where weather-station based observations are inadequate or incomplete. However, these products often lack direct evaluation at specific sites of epidemiological interest.

The first aim of the research presented here was to characterize rotavirus seasonality in eight different locations (the MAL-ED study sites) to demonstrate the feasibility of applying an adapted Serfling model approach to data on EID from a multi-site cohort study. The second was to select climate data products and assess their performance both in characterizing meteorological conditions at those specific eight locations and as predictors of a known climate-sensitive outcome – namely

rotavirus infection episodes. The third aim was to characterize the associations between a suite of nine EO-derived hydrometeorological variables and rotavirus infection status ascertained through community-based surveillance in a way that can be used to predict future trends in disease burden.

In all seven of the eight study sites where seasonality in rotavirus infection was identified, the primary annual peak occurred outside of the rainy season. In all except two of these, a smaller, secondary annual peak was identified occurring during the rainier part of the year. The patterns predicted by this approach are broadly congruent with several emerging hypotheses about rotavirus transmission and are consistent for both symptomatic and asymptomatic rotavirus episodes. These findings have practical implications for programme design, but caution should be exercised in deriving inferences about the underlying pathways driving these trends, particularly when extending the approach to other pathogens.

The availability and completeness of weather station-based meteorological data varied depending on the variable and study site. The performance of the two gridded EO climate models varied considerably within the same location and for the same variable across locations, according to different evaluation criteria and for the peak-season compared to the full dataset in ways that showed no obvious pattern. They also differed in the statistical significance of their association with the rotavirus outcome. For some variables, the station-based records showed a strong association while the EO-derived estimates showed none, while for others, the opposite was true.

Numerous hydrometeorological parameters – including several that are not commonly measured by weather stations – were found to exhibit complex, non-linear associations with rotavirus infection that differ by infection episode type and may be independently, and highly statistically significant over multiple consecutive or non-consecutive lags, including as short a period as two days. The results show evidence for the hypothesis that the effect of climate on rotavirus transmission is mediated by four independently-operating mechanisms: waterborne dispersion via rainfall and surface runoff, airborne dispersion in humidity-sensitive aerosols, virus survival on soil and fomites, and host factors.

Researchers wishing to utilize publicly available climate data – whether EO-derived or station based - are advised to recognize their specific limitations both in the analysis and the interpretation of the results. Epidemiologists engaged in prospective research into environmentally driven diseases should install their own weather monitoring stations at their study sites whenever possible, in order to circumvent the constraints of choosing between distant or incomplete station data or unverified EO estimates.

Since the EO datasets from which the predictors used in this analysis were extracted are available at global scale and sub-daily resolution and updated continually, as new studies using similar methods are carried out at different locations, they can be added to the MAL-ED data to derive more precise predictions for more diverse conditions. Furthermore, emerging tools for objective climate regionalization can be combined with the results of these models to divide extensive

geographic zones into smaller regions that are homogenous with respect to

important climate characteristics and to which predictions from these models can

be applied.

Table of Contents

List of Figures

List of Tables

<u>Dissertation Committee Members:</u>

Margaret N. Kosek (Advisor), MD, Associate Professor, International Health

Benjamin Zaitchik (Chair), PhD, Associate Professor, Earth and Planetary Sciences

David Sack, MD, Professor, International Health

Abhirup Datta, PhD, Assistant Professor, Biostatistics

<u>Alternate Committee Members</u>

Kawsar Talaat, MD, Assistant Professor, International Health

Justin Lessler, PhD, Associate Professor, Epidemiology

Genee Smith, PhD, Assistant Professor, Environmental Health and Engineering

Acknowledgements

I would like to thank my advisor, Dr. Margaret Kosek, who has been my mentor and advocate within the PhD program, has shown great faith in me, directed countless opportunities my way and conceived of much of the research described here. Dr. Benjamin Zaitchik of Earth and Planetary Sciences is principal investigator on the NASA-funded project "Environmental Determinants of Enteric Infectious Disease: a GEO platform for analysis and risk assessment" of which this research forms a part. Dr. Zaitchik was also instrumental in carrying out this research by extracting the data from the climate models, and providing close consultation on the climate data sources, analysis plan, interpretation and reporting of the results. I am grateful to the many co-authors on the three manuscripts on which this thesis is based, in particular professors Lawrence H. Moulton here at JHSPH and Elena Naumova of Tufts University. I would also like to thank the other members of my exam and thesis committees for their insightful comments and feedback - including Professors Abhirup Datta, William Checkley, Justin Lessler, Roger Peng and David Sack and Dr. Elizabeth Colantuoni.

Within the International Health department, Pablo Yori and the other members of the MAL-ED working group (Ruthly François, Saba Rouhani, Francesca Schiaffino and Nora Pisanic) as well as Cristina Salazar, Karla McCarthy, Samantha Almozara and Karen Charron, and the many students on my PhD program have all supported me in innumerable ways. As a trainee on the National Science Foundation's IGERT Water, Climate and Health fellowship, I have benefitted greatly from the financial support and other opportunities that that program provided, as

Abbreviations

AIC	Akaike information criterion
BGD	Dhaka, Bangladesh
BRF	Fortaleza, Brazil
CHIRPS	Climate Hazards Group Infrared Precipitation with Stations
COD	Coefficient of Discrimination
DD	Decimal Degree
DISC	Data and Information Services Center
DLM	Distributed Lag Model
EAEC	Enteroaggregative Escherichia coli
EE	Environmental Enteropathy
EID	Enteric Infectious Disease
EIEC	Enteroinvasive Escherichia coli
ELISA	Enzyme-Linked Immunosorbent Assay
ENSO	El Niño–Southern Oscillation
EO	Earth Observation
EPEC	Enteropathogenic Escherichia coli
ETEC	Enterotoxigenic Escherichia coli
FPR	False Positive Rate
GDAS	Global Data Assimilation System
GES	Goddard Earth Sciences
GLDAS	Global Land Data Assimilation System
GLM	Generalized Linear Model
INV	Vellore, India
IPCC	Inter-Governmental Panel on Climate Change
IRB	Institutional Review Board
LMIC	Low- and Middle-Income Countries
LSM	Land Surface Model
MAL-ED	The Interactions of Malnutrition & Enteric Infections: Consequences for Child Health and Development project
MBE	Mean Bias Error
MCAR	Missing Completely at Random
NAO	North Atlantic Oscillation
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NEB	Bhaktapur, Nepal
NOAA	National Oceanic and Atmospheric Administration
NSE	Nash-Sutcliffe Efficiency Coefficient
PEL	Loreto, Peru
PKN	Naushero Feroze, Pakistan
qPCR	Quantitative Polymerase Chain Reaction
RMSE	Root Mean Square Error
RV	Rotavirus
RV	Pearson's Correlation Coefficient

SAV	Venda, South Africa
STH	Soil-transmitted Helminthiasis
TPR	True Positive Rate
TZH	Haydom, Tanzania
UTC	Coordinated Universal Time
UV	Ultra-Violet
WHO	World Health Organization

1. Chapter 1: Background

Diarrheal disease is the second leading infectious cause of mortality in children under 5 years of age, accounting for 8.9% (526,000) of all such deaths globally in 2015.[1] Infectious causes of diarrhea are numerous, encompassing a variety of viral, bacterial, protozoan and macroparasitic agents each with their own distinct and complex modes of transmission and interactions with ecological processes and host and reservoir biologies. Enteric viruses – chiefly rotavirus, but also calicivirus (mostly norovirus), astrovirus and adenovirus – account for the largest proportion of diarrheal disease mortality in under 5s and morbidity in the first year of life (Table 1). In the second year of life, bacterial infections take over as the largest contributors to diarrheal disease morbidity. Bacterial agents include the enterobacteriaceae - *Salmonella* spp., *Shigella* spp., *Yersinia enterocolitica* and the various pathotypes of *Escherichia coli* – as well as the water-dwelling *Plesiomonas shigelloides, Vibrio cholerae* and *Aeronomonas* spp. and the zoonotic *Campylobacter* spp.. Protozoan parasites are the etiological agents that account for the smallest proportion of the global burden of childhood diarrhea. Cryptosporidiosis and giardiasis are the most prevalent diarrheagenic protozoa infections, while amebiasis – infection by parasites of the *Entamoeba* and *Iodamoeba* genera, chiefly *E. histolytica* – is a relatively minor contributor and the burden of cyclosporiasis and balantidiasis is so low that it has yet to be quantified. The soil-transmitted helminthiases – principally ascariasis, trichuriasis and hookworm infection – rarely cause death, but they exact a large public health toll in the under-5 age-group by

inhibiting statural growth and cognitive development.[2] Long-term sequelae of repeated enteric infections from viruses and other pathogens include cumulative damage to the structure of the mucosal lumen surface, which simultaneously blunts the villi tips while elongating the crypts leading to disruption of normal intestinal function and eventually, it is hypothesized, chronic secondary malnutrition, a process known as "environmental enteropathy" (EE).[3–5]

Table 1: Estimated percentage of diarrheal deaths (global, 2011) and adjusted attributable fraction of diarrhea (eight MAL-ED sites 2009 - 2014) by individual pathogen

		Percent of deaths[6]	Attributable fraction of diarrhea[7]	
		Age 0-59 months	Age 0-11 months	Age 12-24 months
Viruses	Rotavirus	27.8%	4.8%	4.9%
	Norovirus	9.9%	5.2%	5.4%
	Astrovirus	2.1%	2.7%	4.2%
	Adenovirus	3.1%	1.6%	0.9%
	Total	**42.9%**	**14.3%**	**15.4%**
Bacteria	Aeromonas	-	-	-
	Campylobacter spp	3.2%	3.5%	7.9%
	Enteroaggregative *E. coli*	-	-	-
	Enteroinvasive *E. coli*	-	-	0.8%
	Enteropathogenic *E. coli*	11.1%	1.3%	-
	Enterotoxigenic *E. coli*	6.0%	3.2%	5.1%
	Plesiomonas shigelloides	-	-	-
	Salmonella spp	2.5%	-	0.3%
	Shigella spp	3.9%	0.4%	4.0%
	Vibrio cholerae	1.3%	-	-
	Yersinia enterocolitica	-	-	-
	Total	**28.0%**	**8.4%**	**18.1%**
Protozoa	*Balantidium coli*	-	-	-
	Cryptosporidium spp	2.0%	2.0%	3.8%
	Cyclospora spp.	-	-	-
	Entamoeba histolytica	0.2%	-	0.7%
	Entamoeba coli	-	-	-
	Giardia lamblia	2.3%	-	-
	Iodamoeba butschilii	-	-	-
	Total	**4.5%**	**2.0%**	**4.5%**
Unknown etiology		**24.5%**	**75.3%**	**62.0%**

1.1. Diarrheagenic enteropathogens

1.1.1. Enteric Viruses

Table 2 summarizes the structure, classification and clinical features of the four main enteric viruses. The virus species vary slightly in their genomic structure, including one DNA virus (adenovirus) and one with a segmented genome, which is also, unusually, a double-stranded RNA virus (rotavirus). However, they are similar with respect to their morphology, all four having icosahedral capsids and, importantly, lacking envelopes, a feature which enhances their ability to survive outside the host. Indeed, so environmentally resilient are they that rotavirus can persist for days on surfaces and for weeks in fecal matter, while norovirus can survive on surfaces for up to 2 weeks and is resistant to heating, freezing and many chemical disinfectants.[8,9]

Table 2: Structure, classification and clinical features of four enteric viruses

			Adenovirus[10–12]	Astrovirus[13,14]	Norovirus[15,16]	Rotavirus[17–20]
Structure and classification		Family	*Adenoviridae*	*Astroviridae*	*Caliciviridae*	*Reoviradae*
		Human strains	57 genotypes[i]	8 genotypes	31 genotypes	>40 genotypes
	Genome	Type	DNA	RNA	RNA	RNA
		Strands	Double	Single	Single	Double
		Linearity	Linear	Linear	Linear	Linear
		Segmentation	Monopartite	Monopartite	Monopartite	Multipartite
		Polarity	+/-	+	+	+/-
		Capsid morphology	Icosahedral	Icosahedral	Icosahedral	Icosahedral
		Envelope	Non-enveloped	Non-enveloped	Non-enveloped	Non-enveloped
Clinical features		Incubation period	3 – 10 days	4.5 days	1 – 2 days	1 – 3 days
		Diarrhea	✓	✓	✓	✓
		Vomiting	✓	✓	✓	✓
		Fever	✓	✓	✓	✓
		Abdominal pain	✓	✓	✓	✓
		Loss of appetite	-	✓	-	✓
		Dehydration	✓	✓	✓	✓

The pathogeneses of these viruses are also broadly similar. The dominant route of transmission is fecal-oral with the primary source of environmental contamination being the copious numbers of virions that are shed in the stool of infected individuals in concentrations of 10^5 to 10^{13} copies per gram of feces, often long after clinical symptoms have resolved.[9,13,15,17] These virions enter the body through the mouth on fecally contaminated food, water, fomites or, it is hypothesized, airborne respiratory droplets (rotavirus[18] and adenovirus[11]), aerosolized particles of vomitus (norovirus[15]) or dried infective fecal and dust particles (rotavirus[21]). Only a tiny dose is required to establish a full infection (<100

---

[i] Only two adenovirus genotypes (40 and 41) infect the gastrointestinal tract. The others tend to infect the upper respiratory tract.

particles of rotavirus[18], <20 of norovirus[15]), making them some of the most highly contagious pathogens. Once they have migrated to the gastrointestinal tract, these viruses adsorb onto the cells of the epithelium of the small intestine, where they replicate and establish an IgA-driven infection, triggering a cell-mediated inflammatory mucosal immune response.[13,17] Immunity to any one strain of a given enteric virus does not necessarily confer immunity to others and repeated infections with the same virus type are common.[17,22]

The clinical features of infection differ little among the enteric viruses and include non-specific symptoms characteristic of gastroenteritis of any etiology (including bacterial) – watery diarrhea, vomiting, fever, dehydration. For this reason, differential diagnosis of viral diarrheal disease is only possible by laboratory analysis of stool specimens to detect viral particles or antigens using techniques such as enzyme-linked immunosorbent assay (ELISA) or quantitative polymerase chain reaction (qPCR).[17] The impaired absorption of water, sodium and glucose through the gut lumen is what causes the diarrheal symptoms, dehydration, metabolic acidosis and electrolyte imbalance.[17]

For children raised in communities that lack adequate sanitation, enteric viral diseases are a near ubiquitous feature of childhood. Prior to the introduction of the vaccine around 95% of children globally could expect to experience at least one rotavirus infection before 5 years of age.[23] A study in pre-vaccine era Mexico observed a similar cumulative probability of first rotavirus infection by just 2 years of age, while the equivalent probability for norovirus across the eight MAL-ED sites

was 89%.[22,24–26] It is estimated that deaths due to diarrheal disease declined by 51.0% between 1990 and 2013, an achievement that has been attributed in large part to the introduction of the rotavirus vaccine coupled with widespread improvements in water and sanitation access.[27] Since this live-attenuated oral vaccine against rotavirus was endorsed by the World Health Organization (WHO) in 2009, 86 countries have introduced it into their national routine childhood immunization schedules[26], a policy that is credited with large reductions in disease burden - in Mexico, a documented 41.5% reduction in diarrhea-related child mortality and, in El Salvador, a 70 - 80% reduction in hospitalizations in children under age 5.[24,25]

1.1.2. Bacterial Diarrheal Disease

Table 3 summarizes the physiology, structure, primary reservoirs and pathogeneses of the main diarrheagenic bacteria. The many strains of *E. coli* that populate the gastrointestinal tract as normal microbial flora are grouped into 5 major groups or pathotypes based on their pathogeneses and the clinical features of the disease they cause. Four of these are of public health importance for pediatric populations in low- and middle-income countries (LMICs) and are transmitted by consumption of contaminated food or water. The enteroaggregative pathotype (EAEC) adheres to the mucosa of the small intestine provoking an inflammatory response and leading to hemorrhage and decreased fluid absorption.[28,29] Enteroinvasive *E. coli* (EIEC), so called because it invades and destroys the epithelial cells of the large intestine, is phenotypically related to *Shigella* spp.[28,30] Enteropathogenic *E. coli* (EPEC) particularly effects neonates and infants under 6

7

months of age and is divided into typical, and atypical strains, the latter capable of zoonotic transmission from mammalian reservoirs.[28,30] Enterotoxigenic *E. coli* (ETEC) is so called because it produces one of two enterotoxins – heat labile or heat stable – which induces hypersecretion of fluids and electrolytes.[28,30]

Food contamination is the principal mode of transmission for campylobacteriosis, the major reservoir being poultry, but with other meats and raw milk also being important sources of infection.[31,32] However, there are numerous other, complex routes of infection including anthropozoonotic transmission from livestock and domestic pets. Salmonellosis is also mainly transmitted by food contamination.[31] No non-human animal reservoirs for *Shigella* exists and person-to-person transmission primarily occurs through contact between contaminated hands, and therefore thrive in communities that lack adequate clean water to support hygiene practices. Cholera is caused by a species of the water-borne *Vibrio* bacteria that is part of the natural ecosystem in aquatic environments, and thrive in brackish waters.[33] *Aeromonas* spp and *P. Shigelloides* are also a common feature of the microflora of aquatic ecosystems, but survive at lower levels of salinity and infection often occurs as a result of ingesting contaminated drinking water as well as undercooked fish or shellfish.[28,30,34] *Y. enterocolitica*, because it can grow at colder temperatures than most other bacteria, is a common cause of entercolitis in the northern regions of Europe and North America, but of much less public health importance in tropical regions.[28,30]

Table 3: Clinical features, primary reservoirs, physiology and structure of the highest burden enteric bacteria[28,30]

		Aeromonas spp	Campylobacter spp	EAEC[29]	EIEC	EPEC	ETEC	Plesiomonas Shigelloides[34]	Salmonella spp	Shigella spp	Vibrio cholerae	Yersinia enterocolitica
				E. coli pathotypes								
Clinical features[28,30]	Infectious dose	$>10^{10}$	500	10^{10}	$>10^6$	10^6	10^8	Unknown	$>10^3$	10 – 200	10^6 - 10^{11}	10^8
	Incubation period (days)	1 - 2	1 - 10	<1.5	<2	<2	<2.5	<2	<2	1 - 3	4 - 6	1 - 10
	Duration of illness (days)	Unkn.	2 - 10	$>14^†$	Unkn.	~7	3 - 5	<14	2 - 7	1 - 7	<5	7 - 14
	Diarrhea	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Bloody stool	✓	✓	-	✓	-	-	✓	-	✓	-	-
	Vomiting	-	-	✓	-	✓	✓	✓	✓	-	✓	-
	Fever	-	✓	✓	✓	✓	✓	✓	✓	✓	-	✓
	Abdominal pain	✓	✓	-	✓	-	✓	✓	✓	✓	✓	✓
	Dehydration	-	-	✓	-	✓	-	-	-	-	✓	-
	Myalgia, headache	-	-	-	-	-	-	-	✓	-	-	-
	Primary reservoir	Fresh/ brackish water	Animals (esp. birds)	Humans				Fresh/ brackish water	Animals	Humans	Estuarine/ marine water	Mamm- alian hosts
Physiology & Structure[28]	Gram stain	Negative	Negative	Negative				Negative	Negative	Negative	Negative	Negative
	Oxidase test	Positive	Positive	Negative				Positive	Negative	Negative	Positive	Negative
	Catalase test	Positive	Positive	Positive				Positive	Positive	Positive	Positive	Positive
	Morphology	Bacillus	Helical	Bacillus				Bacillus	Bacillus	Bacillus	Vibrio	Bacillus
	Metabolic requirements	Facultative anaerobe	Micro- aerophile	Facultative anaerobe				Facultative anaerobe	Facultative anaerobe	Facultative anaerobe	Facultative anaerobe	Facultative anaerobe
	Fermentative	Yes	No	Yes				Yes	Yes	Yes	Yes	Yes

† Chronic cases

1.1.3. Enteric Parasites

Two types of parasites may establish infections in the human gastrointestinal tract: unicellular protozoa, and parasitic intestinal worms known as helminths. Table 4 summarizes the classification, clinical features and primary reservoirs of the main enteric parasite species that infect humans. Protozoan parasites are excreted in the stool and transmitted in the form of cysts (or oocysts for some species) which contaminate recreational water bodies (a means by which they are dispersed through the environment), drinking water sources, food and fomites or are carried by insect. Following ingestion, these cysts hatch upon contact with gastric acid releasing pathogenic trophozoites (or sporozoites in the case of oocysts) in a process called excystation.[35] The principal protozoan infections that are of public health concern are cryptosporidiosis (mainly caused by the *C. parvum* and *C. hominis* species) and giardiasis (*G. lamblia*), while cyclosporiasis can reach prevalence rates of up to 18% in endemic areas and balantidiasis (*B. coli*) is endemic globally, but mostly asymptomatic.[35] All of these parasite species exhibit zoonotic transmission between a variety of animal hosts. In the case of *B. coli*, this is limited to swine kept as livestock (and sometimes monkeys) and pig farming is a major risk factor for transmission, but for *G. lamblia*, sylvatic transmission cycles can be maintained indefinitely between wild mammalian hosts in the absence of humans.[28] *Entamoeba histolytica* is the most common agent behind pathogenic amebiasis, and in tropical and subtropical regions, prevalence of infection can reach 15% though infection is often asymptomatic.[35] Amebiasis due to *Entamoeba coli* and *Iodamoeba Bütschilii* infection is believed not to be pathogenic in humans, and is only of public health

concern insofar as the presence of these amebae in the stool is an indicator of fecal contamination of food or water.[35]

The helminthiases are some of the most common infections in humans, particularly those caused by the soil-transmitted helminths (STH) – *A. lumbricoides*, *T. trichuria*, *S. stercoralis* and the two hookworm species *A. duodenale* and *N. Americanus* – with which around 1.5 billion people are thought to be infected by one or more species.[36] These worms inhabit the gastrointestinal tract where they survive for up to four years reproducing sexually. The two cestodes or tapeworm species *Hymenolepis nana* and *diminuta* as well as several species of trematode *Schistosoma* also infect the intestine, but these require intermediate hosts to complete their lifecycles.[28,35] A female helminth may, depending on the species produce between a few dozen and 200,000 eggs per day, which, like protozoan cysts, are passed by the host in the feces and, in the absence of adequate sanitation, contaminate the environment.[28,37] For ascariasis and trichuriasis these eggs remain viable for up to several years or until they are ingested directly into the gastrointestinal tract in drinking water or vegetables grown in contaminated soil, where they establish new infections. *S. stercoralis* and hookworm eggs hatch into larvae, which develop in the feces or soil and upon contact with a human host, penetrate the skin to the circulatory system and on via the heart and lungs or connective tissue to the small intestine.[35] Tapeworm eggs must first infect an arthropod insect, which, in turn is ingested by the primary human or mammal host, whereas schistosome eggs hatch into miracidia when they reach open water bodies, penetrating first into freshwater snail hosts to convert into infective cercariea, and

then the skin of the human host as they swim or bathe.[35] Helminthiasis rarely causes

death, although acute *A. lumbricoides* infection in very young children can cause

fatal obstruction of the gut. Instead its public health consequences are manifest in

the chronic, insidious effects that the condition has such as malnutrition, anaemia,

impeded growth and increased susceptibility to other infections.[38]

		Classification	Duration of illness/infection	Incubation period	Reservoir for infection	Clinical features											
						Diarrhea	Bloody stool	Vomiting	Nausea	Fever	Abdominal pain	Loss of appetite	Dehydration	Myalgia	Headache	Fatigue	Weight loss
-Protozoa	***Balantidium coli***	Ciliate	Unkn.	3 – 4 days	Swine	✓	✓	✓	✓	-	✓	✓	-	-	✓	-	✓
	***Cryptosporidium* spp.**	Sporozoa	10 - 21 days	7 – 10 days	Humans, animals	✓	-	✓	✓	✓	-	-	-	✓	-	✓	-
	***Cyclospora* spp.**	Sporozoa	Weeks	2 – 11 days	Surface water	✓	-	-	✓	-	✓	✓	-	-	-	-	-
	Entamoeba histolytica	Amebae	Unkn.	>3 days	Humans	✓	✓	-	-	✓	✓	-	-	-	-	-	-
	Entamoeba coli	Amebae	Unkn.	-	Humans	-	-	-	-	-	-	-	-	-	-	-	-
	Giardia lamblia	Flagellate	10 – 14 days	1 – 2 weeks	Humans, sylvatic mammals	✓	-	-	✓	✓	✓	-	-	-	-	-	-
	Iodamoeba bütschlii	Amebae	Unkn.	-	Humans	-	-	-	-	-	-	-	-	-	-	-	-
Helminths	***Ascaris Lumbricoides* (roundworm)**	Nematode	<2 years	6 – 8 weeks	Contaminated soil	✓	-	-	✓	-	✓	✓	-	-	-	-	-
	***Hymenolepis diminuta* (tapeworm)**	Cestode	Unkn.	20 days	Humans, arthropods	-	-	-	-	-	-	-	-	-	-	-	-
	***Hymenolepis nana* (tapeworm)**	Cestode	Years	-	Humans, arthropods	✓	-	-	-	-	✓	✓	-	✓	✓	✓	-
	Hookworm (*A. duodenale & N. Americanus*)	Nematode	<5 years	3 – 5 weeks	Contaminated soil	✓	✓	-	✓	-	✓	-	-	-	✓	✓	-
	***Strongyloides stercoralis* (threadworm)**	Nematode	Years	Unkn.	Contaminated soil	✓	-	-	-	-	✓	-	-	-	-	-	-
	***Schistosoma* spp. (blood fluke)**	Trematode	<30 years	4 – 6 weeks	Water-dwelling snails	✓	✓	-	-	✓	-	-	-	-	✓	-	-
	***Trichuris Trichuria* (whipworm)**	Nematode	<5 years	Unkn.	Contaminated soil	✓	-	✓	-	-	✓	-	-	-	✓	-	-

---

[iii] Clinical features listed here apply only to the intestinal infections with the parasite. "Unkn." = unknown.

1.2.Seasonality of infectious diseases

A variable is said to be "seasonal" if it exhibits repeating, periodic

fluctuations over calendar time, usually within the course of a year.[31,39] When a

seasonally varying outcome is plotted against calendar time it may take the

appearance of an undulating sinusoidal curve with a regular interval between the

maximum and minimum values occurring within the same year – the peaks and

troughs.[16] The amplitude (or magnitude) of the seasonality is the difference

between these two values (or the ratio of one to the other).[31,39] There are countless

communicable diseases that display some kind of seasonal pattern and in certain

cases, the annual curve can explain up to 60% of the variability in the outcome.[40]

While it may be tempting to attribute this phenomenon merely to cyclical changes in

the weather constraining the transmission and survival of the infectious agents, this

only gives a partial picture.[41] In fact, the true mechanisms underlying infectious

disease seasonality include multifarious environmental, behavioral and

immunological drivers that are specific to particular pathogens, their human and

non-human hosts and their locations.[42] These can interact to produce subtle

periodic fluctuations in either the reproductive number or the fraction of the

population susceptible to infection at a given time of the year.[16] For some diseases,

the influence of the weather may indeed be fairly direct, like in Bangladesh, where

variation in ambient temperature and sunshine hours over the annual cycle, interact

to produce spikes in cholera incidence in both the winter and summer.[43] However,

correlations between environmental factors and seasonal peaks should not

automatically be construed as direct causation, since they may be mediated through changes in behavior of the human host. For example, rainfall may alter patterns of contact between infected and susceptible individuals as they congregate indoors[44], while a preference for cooking outdoors during clement weather may put people at a greater risk of food-borne illness during summer months.[41] Seasonal fluctuations in tuberculosis incidence have been explained by annual variations in rates of contact, reactivation and tuberculosis-induced mortality, that are largely independent of whatever effects the weather may have on *Mycobacterium* survival.[45] Outbreaks of zoonotic diseases may occur as the result of seasonal animal husbandry practices while higher rates of international travel at certain times of the year may facilitate the importation of pathogens from high- to low-endemicity settings.[32,46,47] Summertime increases in recreational water activities and water consumption may explain the coincident peaks in waterborne pathogens.[31,48,49] Seasonal fluctuations in immune competence are another hypothesized driver, but have yet to be conclusively identified in humans.[42] However, immune status can play a role in seasonality of pediatric infections, since surges in birth rates at certain times of the year can introduce annual waves of susceptible individuals into the population at risk.[50] A historical example that illustrates this point is that a major driver behind seasonality in measles infections that long perplexed epidemiologists turned out to be the sudden congregation of immune susceptible children at the beginning of school terms.[42]

Most enteric infectious diseases (EIDs) exhibit some kind of seasonality, their incidence peaking at a particular point in the annual cycle and receding at others

year after year.[51] These patterns vary with latitude and climatic zone, while diseases caused by the same type of pathogen and that share common routes of transmission and exposure sources, tend to converge in their seasonal patterns suggesting an overlapping environmental influence.[31] At least part of this observed seasonal variation can be explained by annual changes in the weather conditions that in some seasons favor transmission and in others inhibit it, however host factors no doubt also play a role. The relative influence of these co-seasonal social, behavioral and immunological determinants of transmission risk is poorly understood due to the methodological challenges of adjusting for them.[16,42] Furthermore, for diseases with multiple routes of transmission, different mechanisms may come to dominate their relative contribution to overall burden at different points in the annual cycle. For example, for rotavirus in south Asia, the large winter peak may be accounted for by airborne transmission on dried dust substrates, while the smaller monsoon peak may be due to wider dispersal of the virus in floodwater and runoff promoting water-borne transmission.[21]

1.2.1. Seasonality of Enteric Viruses:

That viral diseases demonstrate seasonality is reflected in the everyday lexicon in phrases like "flu season" and the colloquial names for rotavirus and norovirus – respectively "winter vomiting disease" and "wintery diarrhea". Of the highest burden enteric viruses at least three – astrovirus, norovirus and rotavirus – show clear seasonality following a similar pattern. Broadly speaking, transmission peaks in winter and recedes to a sustained lower level in the offseason across all latitudes but most pronouncedly in temperate climes, tropical regions having more

sustained, year-round patterns with discernable, but less pronounced relative peaks.[13,19–21,48,52–55] Human astrovirus disease, while circulating throughout the year, peaks annually in winter in temperate climates – thought to be due to lower temperatures conferring stability of the virus – and in the rainy season in the tropics.[13]

For rotavirus, this broad pattern may be something of an oversimplification.[19,21] Several studies in tropical South Asia have found biannual peaks in rotavirus diarrhea in children, with a large winter peak followed six-months later by a smaller, secondary peak coinciding with the monsoon rains.[20,53,56–58] There are conflicting results in studies attempting to link particular climatic variables to temporal variations in rotavirus burden.[53] Temperature and humidity are known to be inversely associated with rotavirus transmission risk, which may relate to the conditions conducive to the virus's survival, or, in the case of humidity, to its areal transport in dried, infective fecal and dust particles.[16,19–21,53,59] Drier environments have been shown to support transmission of the virus, and the association observed between wind speed and rotavirus hospitalizations also lends credibility to the latter putative mechanism.[16,19,21] Higher atmospheric pressure and solar radiation, in addition to lower temperature, have been associated with increases in rotavirus outcomes.[19] One study in Bangladesh, by contrast, found that, above a 29°C threshold, higher temperature was associated with an increase in rotavirus diarrhea hospital admissions, although it was also associated with higher river levels, so this could have been the result of confounding with waterborne transmission of the virus during the monsoon season.[59] A study of three Australian

17

cities showed increased temperature and humidity to be associated with decreased rotavirus diarrhea hospitalizations.[60]

Ambient temperature (absolute as well as relative), humidity and rainfall have all been associated with norovirus risk and as such are potential candidate drivers of the disease's seasonality although most studies aggregated climatic variables to monthly averages and are thus limited in how accurately they can quantify the effect measure and eliminate residual confounding.[48,53] A direct effect of temperature is suggested by the observation that norovirus has increased durability in cold water.[16] Water-borne transmission is a secondary route of infection and in several studies, higher measures of rainfall predicted greater norovirus seasonality measures.[48] The results of one laboratory study supported the hypothesis that lower absolute (as opposed to relative) humidity was favorable to the survival of norovirus in the environment and its prolonged infectivity.[61] A similar relationship is observed with influenza, the seasonality of which is influenced by the virus being able to survive longer in the air when humidity is low.[42] However, the evidence for such an inverse association is not substantial enough to deter some scholars from hypothesizing a direct association, mediated by increased humidity facilitating the airborne transport of virus-containing aerosols.[55] It is conceivable that conditions of low humidity might promote transmission via the fecal-oral route and contamination of surfaces and fomites, while high humidity enables transmission of aerosols by vomitus. In addition to seasonal climatic drivers, it has been speculated that an increased proportion of time spent indoors in crowded conditions could be a factor increasing human-to-human transmission of

norovirus in winter.[48] Reviews have been unable to confirm or refute the claim that norovirus peaks in warmer months in the southern hemisphere compared to the northern, due to paucity of published studies from south of the equator.[53]

Much less evidence has been published on the seasonality of adenovirus. While one study in Bangladesh reported a peak in prevalence coinciding with the May to July rainy season[62], and in Japan a winter-spring peak has been reported[63], several other studies have failed to discern a clear seasonal pattern.[64–66] Outbreaks of waterborne adenovirus are associated with recreational water activities, a behavioral exposure that may increase in summer months.[67]

1.2.2. Seasonality of Bacterial Diarrheal Disease:

In contrast to the seasonal patterns exhibited for enteric viruses, almost all bacterial enteric infections tend to peak between spring and late summer. Many foodborne bacteria thrive at temperatures close to that of the human body so environmental temperatures that approach 37°C would be expected to be most conducive to propagation of these agents, but their survival in soil can be prolonged by low temperatures at which their metabolic processes slow.[32]

The multiple routes of transmission for campylobacteriosis make it difficult to isolate the effects of individual potential seasonally varying predictors.[32,49] In most locations, the annual peak in cases occurs in the summer months, coinciding, according to the findings of a time series analysis, with the peak in temperature and following the ambient temperature trend to a low in winter.[31,32,46] Elsewhere, it has been suggested that campylobacteriosis incidence tends to peak in spring (or earlier

in areas with milder winters), roughly three months prior to the annual maximum temperature in mid-latitudes, with equatorial regions experiencing sustained, year-round transmission, a geographical pattern that is strongly suggestive of climatic drivers.[68] It is also possible that the spring peak of campylobacteriosis is associated with annual changes in avian and mammalian reservoir behavior and foraging or animal husbandry activity following the winter, that promotes increased contact between human and animal hosts.[49] While a direct association between temperature and *Campylobacter* spp survival has been observed, it has also been suggested that, in surface waters, increased UV radiation reduces survival of the bacterium.[69] It has yet to be establish the extent to which the seasonality is determined by the bacteria's biology, in particular, its sensitivity to oxygen, moisture and pH and its inability to survive in the open environment and to propagate at temperatures below 30°C.[32] It has been speculated that milder winters promote the proliferation of *Campylobacter* spp in the environment, however, the precise effect of climate on seasonality remains unclear, and the evidence for the presumed direct effect of rainfall or temperature changes on transmission is weak.[49]

Incidence of salmonellosis peaks in summer, around two weeks after the maximum temperature with a low in winter according to time series analyses.[31,46] A similar analysis of data from ten European countries found a clear association between average environmental temperature and salmonellosis incidence with a one-week lag effect, however whether the mechanism behind this association is direct or indirect – for example, seasonal changes in eating and food preparation habits, undercooked barbecue meat etc. – is unclear.[32,70,71] In Ontario, infections

peaked in the spring.[47] *Salmonella* are one of the more resilient bacteria species surviving outside the host in a range of temperatures and environmental conditions, and gardening has been highlighted as a possible seasonally varying behavioral risk factor, since contamination of soil from stool of domestic pets may be a pathway.[32,71]

Diarrheagenic *E. coli* infection also peaks in summer, and ambient temperature may be the main environmental driver. A meta-analysis of 18 studies found a 1°C increase in mean monthly temperature to be associated with an 8% increase in incidence.[72] Monthly incidence of bloodstream infection due to *E. coli* was predicted to increase by 7% with every 10°F increment in mean monthly temperature, although the mechanism is unclear and seasonal changes in bacterial virulence or host susceptibility are not known to occur.[73] After being shed in the feces of animal reservoirs, the bacterium can persist in the soil for long periods before migrating to groundwater or contaminating fresh produce so seasonally varying agricultural or food-borne exposure cannot be ruled out.[32]

The seasonality of *Vibrio cholerae* infection is relatively well studied. One review paper found that cholera outbreaks exhibited more pronounced seasonality the higher the latitude often coinciding with the wettest months, but that this pattern broke down in the tropics where outbreaks are more common, occurring year-round with no discernible peak.[74] A time series analysis of cholera data from Bangladesh concluded that upper tropospheric humidity, cloud cover and top-of-atmosphere absorbed solar radiation were likely drivers of the temporal variability.[75] Others have emphasized the role of surface water temperatures, sea

surface height (which transports *V. cholerae* into inland waters), water salinity and turbidity[33] and the link with warm El Niño events.[74,76]

Little has been published on the seasonality of *Aeromonas* spp, shigellosis, *Y. enterocolitica*, and *P. shigelloides*. Seasonality of shigellosis is not as marked as for other bacterial enteropathies, but a modest peak seems to occur later in summer, around one month after the annual peak in temperature.[31,77] Gastrointestinal carriage of *Aeromonas* spp has been shown to occur in the warmer months.[28] *Yersinia enterocolitica*, uniquely among the bacterial enteric diseases, exhibits a seasonal peak in winter, which is thought to be a direct result of the psychrotrophic nature of the bacterium.[32] As a result, it is perhaps the pathogen for which there is the most convincing evidence for climate as a direct factor influencing seasonality.

1.2.3. Seasonality of Enteric Parasites:

Among the diseases caused by protozoan parasites, the cryptosporidiosis season occurs shortly before autumn, around a month after the annual temperature high and peak incidence of giardiasis follows shortly after that.[31] This varies considerably by location however, with the UK and New Zealand experiencing spring peaks and North America, summer-autumn peaks in cryptosporidiosis.[46] In some settings, animal husbandry practices can supersede climatic drivers, such as in New Zealand, where the annual peak in cryptosporidiosis coincides with the spring lambing and calving season.[46,78] Temperature, precipitation and the Normalized Difference Vegetation Index (NDVI), a remotely sensed proxy measure for the combined effect of these two variables on foliage abundance, have all been shown to

be predictive of cryptosporidiosis incidence, particularly in warm, wet locations like the humid mid-latitude climate zones.[79] For amebiasis caused by *Entamoeba histolytica*, there is some published evidence for a seasonal peak occurring in summer and autumn though little is understood about the drivers of this trend.[80]

The timing of annual peaks in helminthiasis transmission varies by location and species, but in many settings coincides with the rainy season.[81] Survival and viability of helminth eggs in the environment and the development of worm larvae are highly dependent on ambient temperature. In temperate regions, there is a direct relationship between temperature and most indicators of helminthiasis transmission up to an upper limit of 35-45°C (depending on the worm species) above which, eggs are no longer able to develop.[81] The role of precipitation is unclear. In general moist and humid conditions are conducive to helminth proliferation, but it is speculated that rainfall at larger volumes may rinse helminths from the environment in soil runoff.[82]

Table 5: Transmission routes and seasonally varying drivers of transmission for EIDs of public health importance[12,13,47–49,53,55,56,59,61,65,68,16,70,71,73,74,77,79–83,19,84–92,20,21,28,31,32,34] [iv]

		Food-borne	Water-borne	Zoonotic	Fecal-Oral	Airborne	Surfaces/fomites	Timing of seasonal peak	Rainfall	Temperature	Humidity	Soil moisture	Wind speed	Atmospheric pressure	Solar radiation	Travel	Water exposure	Eating/food habits	Indoor crowding	Animal contact	Immune function
Viral	Adenovirus	-	✓	-	✓	✓	✓	Unknown	-	-	-	-	-	-	-	-	↗	-	-	-	(✓)
	Astrovirus	✓	✓	-	✓	-	✓	Winter	-	(↘)	-	-	-	-	-	-	↗	-	-	-	(✓)
	Norovirus	✓	✓	-	✓	✓	✓	Winter	↗	↘	(↘)	-	-	-	-	↗	-	-	(↗)	-	(✓)
	Rotavirus	✓	✓	-	✓	✓	✓	Winter	↘	↘	↘	(↘)	(↗)	↗	-	-	-	-	-	-	(✓)
Bacterial	*Aeromonas* spp	✓	✓	-	✓	-	-	Unknown	-	↗	-	-	-	-	-	↗	-	-	-	-	(✓)
	Campylobacteriosis	✓	✓	✓	✓	-	-	Spring	-	↗	-	-	-	-	(↘)	↗	↗	(✓)	-	(↗)	(✓)
	Diarrheagenic *E. coli*	✓	✓	✓	✓	-	✓	Summer	-	↗	-	-	-	-	-	↗	↗	(✓)	-	(↗)	(✓)
	P. shigelloides	✓	✓	-	-	-	-	Summer	-	↗	-	-	-	-	-	↗	↗	✓	-	-	-
	Salmonellosis	✓	✓	✓	✓	-	-	Spring/summer	-	↗	-	-	-	-	-	(↗)	-	(✓)	-	(↗)	(✓)
	Shigellosis	✓	✓	-	✓	-	✓	Late summer	-	-	-	-	-	-	-	↗	↗	-	-	-	(✓)
	Cholera	✓	✓	-	✓	-	-	Rainy season	↗	↗	↗	-	-	-	↗	-	-	-	-	-	(✓)
	Y. enterocolitica	✓	✓	✓	✓	-	-	Winter	-	↘	-	-	-	-	-	-	-	-	-	(↗)	(✓)
Parasitic	Cryptosporidiosis	✓	✓	✓	✓	-	-	Late summer	↗	↗	-	-	-	-	-	↗	↗	-	-	(↗)	-
	Cyclosporiasis	✓	✓	✓	✓	-	-	Rainy season	↗	-	-	-	-	-	-	-	(↗)	(✓)	-	(↗)	-
	Giardiasis	✓	✓	✓	✓	-	-	Late summer	↗	↗	-	-	-	-	-	↗	↗	-	-	(↗)	-
	Amebiasis	✓	✓	-	✓	-	-	Summer/autumn	-	-	-	-	-	-	-	(↗)	-	-	-	-	-
	Helminthiasis	-	-	-	✓	-	✓	Rainy season	(↘)	↗	↗	(↘)	-	-	-	-	-	-	-	-	-

---

[iv] "↗" = directly associated, "↘" = inversely associated, "-" = no evidence for an association. Parentheses indicate a hypothesized association for which evidence is as yet inconclusive.

1.3.Climate, Weather and Infectious Disease

The impact of climate on infectious diseases is closely related to, but distinct from seasonality. As explained in the previous section, cyclical annual weather patterns at a given location can be a key driver of seasonality for many infections, and assessing the relative contribution of non-environmental, host factors is a major methodological challenge. In contrast, the climate of a location – the average behavior of the weather over a longer term, multi-year period – determines the range of values within which hydro-meteorological parameters can be expected to vary both in annual cycles and from day to day. Since the growth, survival and dispersal of microorganisms and the distribution of their intermediary hosts and vectors is determined by environmental factors, hydrometeorological parameters being principal among them, climate is a key constraint on the geographic and temporal distribution of infectious diseases.[32,41]

While climatic drivers of some infections, principally vector-borne diseases have been relatively well-explored, currently little is understood of their role in enteric disease transmission and there is no unified theoretical framework through which to conceptualize the relative influence of individual drivers.[16,20,93] What exploration there has been has tended to focus on non-specific, morbidity-dependent outcomes such as hospitalizations for acute diarrheal disease and outbreaks of gastroenteritis and not on underlying, background endemicity of specific pathogens.[94] On the exposure side, there has been a disproportionate emphasis on extreme weather events and anomalous climate phenomena such as

the El Niño–Southern Oscillation (ENSO) and North Atlantic Oscillation (NAO) as predictors.[95–97] Less has been explored about small-scale meteorological variability within the normal range. A variety of hydrometeorological variables have been either shown or hypothesized to influence the transmission of enteropathogens and these are explained in detail in chapter 3. Many of these parameters are highly correlated so there may be complex interactions between these highly collinear variables and the transmission mechanisms that make it challenging to tease out the particular effect of single factors.[19]

1.4. Impact of Climate Change

The global proliferation of greenhouse gases in the atmosphere caused by human activities – primarily the burning of fossil fuels but also agriculture and deforestation – is leading to long-term and irreversible increases in global surface temperatures, shifts in the distributions of other climate variables and increased risk of extreme weather events.[98,99] Projections from the Inter-Governmental Panel on Climate Change (IPCC) indicate that, by the end of this century, average global surface temperature will have risen by more than 1.5°C compared with the period from 1850 – 1900, with some areas experiencing much greater warming due to regional variability.[98] As the effects of anthropogenic climate change gain momentum, tropical and sub-tropical regions are expected to see the largest seasonal increases in mean temperatures in the short term, while most areas will experience more frequent and longer lasting heatwaves and less frequent low temperatures.[32,98] Variability in precipitation and evapotranspiration is set to

increase due to climate change, exaggerating the pattern of high rainfall at the equator and polar fronts and low rainfall at the subtropical heights.[98] This is due to a precipitation feedback mechanism, which, in the subtropics will lead to increased entrainment and elevated lifting condensation level. At the equator, drought precipitation feedbacks will cause a cycle of increases in soil moisture, evapotranspiration and precipitation. The direction of the change in rainfall will likely vary on smaller geographic scales as well, but in ways that are more difficult to predict.[100,101]

Among the many ways in which these trends will make human life on earth more precarious, is in their impacts on public health generally and diarrheal and infectious diseases specifically. The geographical ranges of disease-transmitting insects have already expanded poleward and regions in which vector-borne disease transmission had previously been eliminated, such as southern Europe, are again seeing autochthonous cases.[102] Enteric pathogen transmission will likely be impacted too, but in ways that are difficult to predict. Incidence of food- and water-borne enteropathies may increase as larger swathes of the globes surface spends a greater proportion of the year within the temperature range that is optimal for bacteria and protozoa proliferation. A review of the literature on this topic found a 1°C increase in temperature to be associated with a risk in diarrheal disease incidence of 3 – 11%.[99] Specific humidity is predicted to increase with climate change while relative humidity – the ratio of the specific humidity to the saturation specific humidity - will remain constant.[98,103,104] This may engender a suppression of enteric virus transmission, or a shift from transmission via dried airborne fecal

particles to droplet, food- or water-borne transmission. The effects of changes in rainfall patterns on diarrheal disease will likely not be linear but felt most pronouncedly at the extremes, with both heavy precipitation events and more frequent droughts making water sources more precarious.[93,99,105]

Under a likely greenhouse gas emissions scenario, the year 2030 may see an additional 50,000 additional deaths due to diarrheal disease attributable to climate change, with 60,000 for malaria, 258 for dengue and 95,000 for undernutrition.[106] However, there is considerable uncertainty around these attempts to quantify the impacts of climate change on human health in general and infectious diseases in particular.[106,107] One of the major sources of this uncertainty is the scarcity of empirical data linking climate and health data spatiotemporally.[107] Due to lack of collaboration between the relevant disciplines climate models are only just beginning to be used to make inferences about health outcomes.[101,108] The outputs from time series regression models need to be combined with projected medium-term changes in climatological parameters, the kind of modeling exercises that are commonplace in the study of ecology and biodiversity but have been slow to be adapted for infectious diseases.[100]

1.5.Study Rationale

Despite widespread recognition of the seasonal nature of the epidemiology of enteric diseases, the mechanisms underlying this phenomenon are poorly understood.[42] Improving understanding of the seasonality of these diseases is methodologically challenging, but essential in order to inform policy on the timing

of preventive measures and to forecast how climate change might influence trends in the burden of diarrhea.[21,42,73] With some notable exceptions, most studies of seasonality of infectious diseases have been hindered by data being aggregated up to weekly, monthly or quarterly cumulative incidence, rather than daily estimates.[31,53] There is a need for more studies in which both the outcome and exposure are matched by precise timing (i.e. daily estimates) and specific geographic areas with relatively homogenous meteorological and population profiles.[48] They need to adjust for confounding by behavioral and socio-demographic factors such as access to improved water and sanitation, socio-economic status, fertility rates and contact with animals.[53] There is also a paucity of studies on seasonality of enteric diseases coming from developing countries – particularly Africa and South America - and those in the tropics and the southern hemisphere.[48]

The extent to which climatic processes and mechanisms drive the transmission of enteric infections is a similarly underexplored issue. This may be a consequence of the fact that it is a public health burden that is primarily borne by underserved and marginalized populations in LMICs. However, given the toll in terms of child mortality that these diseases exact and in the context of a rapidly changing climate, the unanswered questions surrounding this phenomenon are lent a renewed urgency. The data and methods exist to greatly improve predictions of how shifts in climatic variables will affect disease incidence so that the public health community can be better prepared for future changes, but the will to carry out these studies has so far been somewhat lacking. What efforts there have been have

disproportionately focused on vector-borne disease, for which the viable habitat for the insect host is the main environmental constraint on transmission.[93] To address this knowledge gap for a wider range of communicable diseases, multiple regression models are needed to isolate statistical interactions between the numerous, collinear climatic variables and longitudinal analysis approaches are necessary to assess trends over annual cycles and longer-term calendar time, to incorporate lag effects and to adjust for cumulative population immunity.[19,53]

2. Chapter 2: Study context and outcome variable

The three analyses presented in this document all used the same outcome variable from the same study population.

2.1. Study population

In 2009 a unique coordinated cohort study was established by the Interactions of Malnutrition & Enteric Infections: Consequences for Child Health and Development project (MAL-ED), with the aim of investigating the risk factors for enteric infection, diarrheal disease, undernutrition and other related adverse outcomes. This network of institutions has recruited and monitored birth cohorts of 227 – 303 newborns from eight communities, each in a different low- and middle-income country (LMIC) across three continents – Dhaka, Bangladesh (BGD), Fortaleza, Brazil (BRF), Vellore, India (INV), Bhaktapur, Nepal (NEB), Naushero Feroze, Pakistan (PKN), Loreto, Peru (PEL), Venda, South Africa (SAV); and Haydom, Tanzania (TZH). Subjects were monitored continuously over their first 2 years of life using molecular diagnostics and standardized surveillance protocols and assays to track data on EID incidence, nutritional and anthropometric outcomes, cognitive development and biological markers.[109,110] Stool samples were routinely collected as part of active surveillance on, or within two days of the monthly anniversary of each child's birth.[110] Additional, off-monthly samples of diarrheal stool were collected in between the monthly assessments on days in which the caregiver reported that the child was having a diarrheal episode.[109] In this way, the study was able to detect

mild and even asymptomatic cases, a more proximal indicator of endemic pathogen transmission and a less rare outcome than clinical manifestations.[109] The resulting longitudinal dataset contains, among other things, data on infection status for more than 30 enteric pathogens in 50,000 stool samples collected from 2,199 infants from ages 0 to 24 months between November 2009 and March 2014.

The MAL-ED project offers a unique opportunity for a long called-for comparative analysis that can apply the same approach to data from multiple locations and on different pathogens in order to elucidate the patterns and determinants of their seasonality.[42,56,109] The fact that the project has sites in multiple countries situated in different climatic zones means that data on the same infections can be compared longitudinally at different locations to test hypotheses about environmental drivers of their transmission and seasonality.[42] Although the study subjects were only followed up until 24 months of age, because of the staggered enrollment of participants, the MAL-ED data spans a total period of over four years, a sufficient span of time to detect recurring annual patterns and within-season variation. Furthermore, because the precise dates of each assessment are known, the health outcome data can be linked to environmental exposures by the day on which they occurred, a temporal resolution that is rarely seen in published literature.

Furthermore, previous studies have been forced to focus on surveillance or hospitalization data which, by their nature, constrains them to morbidity-related outcomes which manifest only as severe, symptomatic cases. Often this also means

that the outcome is rare, necessitating aggregation to weekly, monthly or seasonal summary statistics.[16] Moreover, underreporting of cases that are less severe, or do not present to the health system may introduce bias.[100] By using routine, active surveillance of the underlying infection status as they occur in the community, the MAL-ED study was able to detect mild and subclinical cases, a more proximal indicator of endemic pathogen transmission and a less rare outcome, which lends greater statistical power to analyses.[109] Differential detection of multiple pathogen species is a further asset, since environmental factors that are detrimental to one pathogen's proliferation may be propitious to another's (the example of water temperature inactivating viruses, but enabling bacterial growth is a case in point).[100] Furthermore, it allows for the exploration of distributed lag effects which may differ by pathogen.[94] Finally, the wealth of baseline and longitudinal data on socio-demographic variables collected by the study will allow for adjustment for common confounders, effect modifiers and mediation of any identified associations.

Table 6 summarizes the Köppen-Geiger climate classifications, precipitation and temperature patterns and other features of each of the MAL-ED study sites and Figure 1 shows their locations in relation to the Equator and Tropics of Cancer and Capricorn. While the sites were originally selected to be characteristic of a variety of epidemiological contexts, they also vary in the type of climate that they experience, offering a representative range of the kinds of weather patterns that prevail across the developing regions of the world. Half of the sites lie in the Southern hemisphere with three of these – Brazil, Peru and Tanzania – situated close to the Equator and the fourth close to the southern edge of the tropics. Of the sites in the Northern

hemisphere, one (Vellore, India) is located in the tropics, and the remaining three are within the mid-latitudes. Because of their different locations they also experience their rainy seasons and annual peaks in temperature at different times of the year and at different intensities. Similarly, the type of settlement and the altitude and topography of their locations – factors which may either have a direct effect on the weather they experience, or mediate the effect on EID incidence – all vary between sites. Three of the sites are located in tropical savannah climate zones (Bangladesh, India and Tanzania), while two share a humid subtropical climate (South Africa and Nepal). The maximum distance over which any site extends is 12.5 km (South Africa east to west) with most sites extending less than 5km in any direction and, according to accounts from site staff, none of the sites exhibit large topographic contrasts. It can therefore be assumed that within-site climate variation will be small, though localized environmental risk factors for specific households— e.g. small depressions or local water bodies—cannot be ruled out. At the time of data collection (2009 – 2014), three of the countries in which the study sites are located – Brazil, Peru and South Africa – had introduced the rotavirus vaccine into their routine childhood immunization schedules.

Table 6: Köppen-Geiger climate classifications, precipitation and temperature patterns and other features of the locations of each MAL-ED study site [111–120]

Site	Main Climate	Precipitation	Temperature	Topo-graphy	Geographic extent of site (km)		Altitude (m)	Distance to weather station (km)	Settlement type	Hemi-sphere
					North-south	East-west				
Dhaka, Bangladesh	Tropical savanna	Dry Nov - Feb, Monsoon Jun - Oct	Mar – May peak	Alluvial plain	0.8	0.6	12	4.6	Urban	Northern
Fortaleza, Brazil	Tropical monsoon	Dry Aug - Dec, rainy Jan - July	Hot year-round	Coastal	1.5	1.2	28	5.3	Urban	Southern
Vellore, India	Tropical savanna	Dry Jan - May, Monsoon Jun - Dec	Mar – Jun peak	Hilly	1.5	1.1	231	1.0	Urban	Northern
Bhaktapur, Nepal	Humid subtropical	Dry Oct- Mar, Monsoon May - Aug	Hot during Monsoon, Apr – Jun	Hilly	2.2	2.8	1,317	7.5	Peri-urban	Northern
Naushero Feroze, Pakistan	Desert/ Arid	Dry, short Monsoon Jul - Sep	Very hot Mar – Oct	Flat	8.4	4.6	44	21.9	Rural	Northern
Loreto, Peru	Tropical rainforest	Fully humid, year-round rain	Hot year-round	Flat	2.1	1.5	89	3.3	Rural	Southern
Venda, South Africa	Humid subtropical	Dry May - Sep, rainy Oct - Mar	Hot Sep – Feb	Hilly	3.8	12.5	657	36.9	Peri-urban	Southern
Haydom, Tanzania	Tropical savanna	Dry Jun - Sep, rainy Nov - May	Temperate year round	Hilly	3.0	5.5	1,650	31.8	Rural	Southern
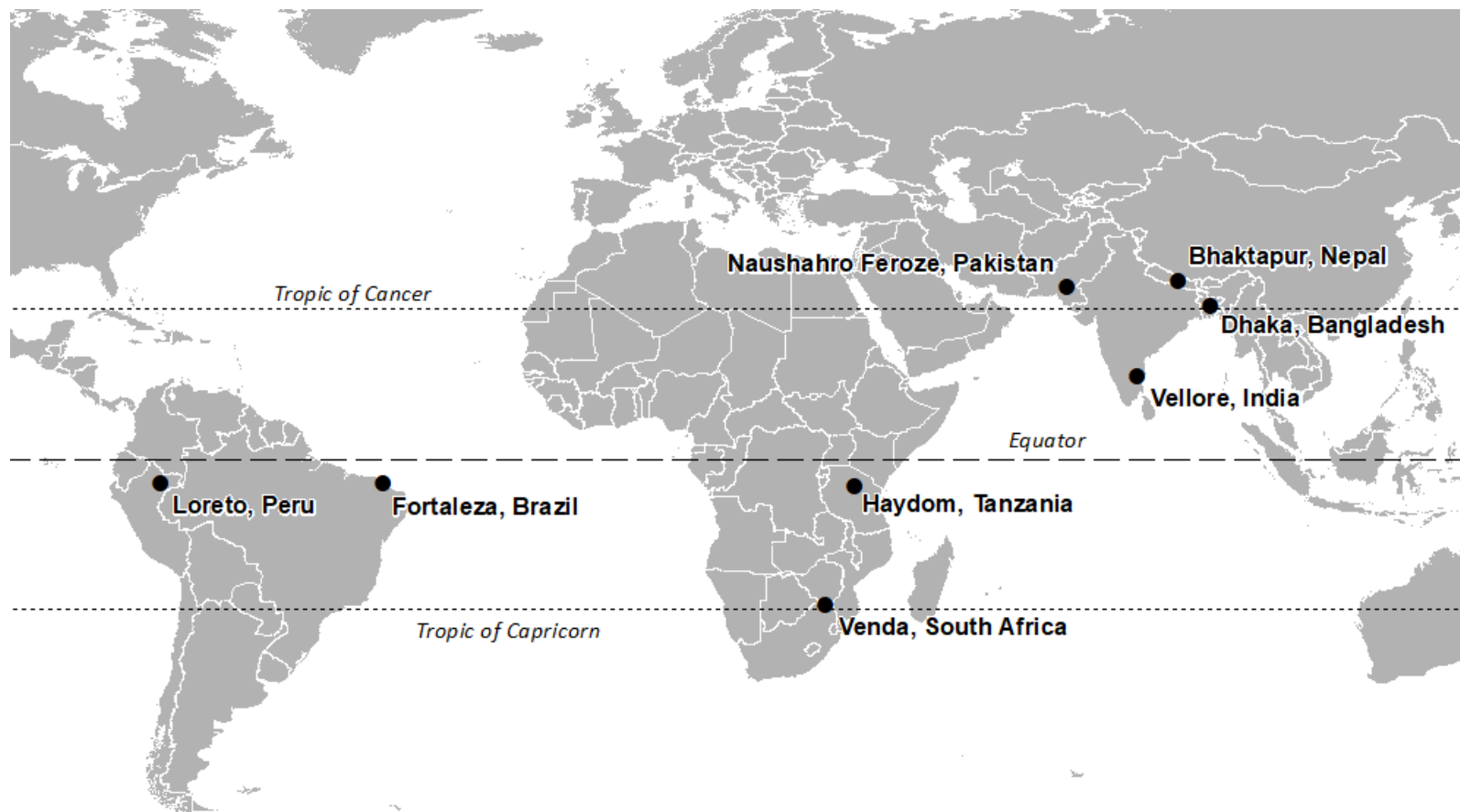
Figure 1: Locations of the eight MAL-ED study sites in relation to the Equator and Tropics of Cancer and Capricorn

2.2.Sample selection and enrollment

Communities were selected at each of the eight study sites and fieldworker teams enumerated a sample frame of all women of reproductive age living within each community's catchment area. If an enumerated woman then gave birth during an enrolment period that started in November 2009, they were approached by field staff to gain consent and determine eligibility for their newborn to participate in the study, with the aim of recruiting at least 200 study infants per site. Enrollment of each child took place within 17 days of birth.

2.3.Data collection

Baseline data on demographic and anthropometric characteristics were obtained at enrollment, while daily data on morbidity, care-seeking and infant feeding were ascertained by care-giver report from twice-weekly household visits, and anthropometric and clinical data were collected during monthly assessments.[109] Stool samples were collected routinely from the infants at roughly monthly intervals from birth to 24 months of age. In addition to the routine collections, some stool samples were collected between scheduled assessments on days in which the child's care-giver reported that the child had experienced a diarrheal episode.[109] The samples were collected by field workers who stored them for processing at a temperature of -70°C without fixative.[121]

2.4. Quality control

All procedures, questionnaires, checklists and guidelines used in MAL-ED were produced by a technical subcommittee assembled within the project study with the purpose of maintaining quality standards across all eight sites of the study.[122] Prior to digital data entry, all forms were reviewed by field supervisors, who also carried out monthly surveillance on 10% of sampled households to validate data from the field workers. In addition, data from all sites was screened by a central Data Coordinating Center that was established to implement rigorous quality control measures (which have been described elsewhere).[122]

2.5. Ethical considerations

Prior written informed consent from a parent and/or guardian was a prerequisite for all participants' inclusion in the study. Study staff read the consent form to the caregiver and then a copy was left in their possession. Ethical approval was given by the Johns Hopkins Institutional Review Board (IRB) as well as from the local partner organization at each site.

2.6. Outcome variable

Stool samples were tested for the presence of shed rotavirus using two methods that were standardized across sites. Samples from children who remained in the study up to 24 months of age were retrospectively tested for rotavirus (and 28 other enteropathogens) using probe-based qPCR assays on custom-developed TaqMan Array Cards (ThermoFisher, Carlsbad, CA, USA).[123] Samples from subjects

who did not complete follow-up were assessed for rotavirus positivity prospectively by ELISA but only for quarterly samples in the second year of life.[110] The outcome of interest in these analyses was therefore the time-varying, stool sample-level, binary infection status variable, for which a value of 1 signified that a stool sample was found to be positive for rotavirus and 0, negative. Rotavirus-positive stool samples were excluded from the dataset if they were collected within 14 days of a previous positive sample, without being separated by an intermediate negative sample or within the subsequent 14 days of the subject receiving the rotavirus vaccine (this being a plausible maximum duration of viral shedding in immunocompetent persons).[17,22]. This was to ensure that a single infection episode was not counted multiple times and that stools containing shed vaccine virus were not incorrectly classified as positive for wild virus.[58]

Table 7 summarizes the number, percentage and incidence of rotavirus episodes detected by sample type as well as the number length of follow-up time, number of subjects and the proportion that were lost to follow-up in each of the eight sites and by vaccine category. At all sites, in both vaccine categories and in the sample overall, the percentage of diarrheal stool samples that were rotavirus positive was consistently higher than that observed in the monthly samples, but by ratios that varied from 1.8 times at the Pakistan site, to 4.5 at the Bangladesh site. While the rate of rotavirus positivity in both sample types was lower in sites situated in countries that had introduced the rotavirus vaccine, the incidence rates of the two types of rotavirus episode exhibited more variation by site and vaccine category, with incidence in the Peru site exceeding that observed in several of the

non-vaccine country sites. Loss to follow-up was defined as the proportion of the total number of participants for whom no observation was recorded after age 23.5 months and exceeded 10% in all sites except India and Nepal. In the Brazil and Peru sites, these levels of attrition were considerable with more than a third of the initial participants lost to follow-up by 24 months of age. Exploratory analysis of the missingness patterns (not reported) indicated that the probability that an observation was missing did not depend on the subjects' history of observed rotavirus episodes, suggesting an underlying mechanism of Missingness Completely at Random (MCAR).

Figure 2 shows needle plots of the observed daily distribution of rotavirus infections (included, positive samples) for each site along with the overall length of follow-up time. With the possible exception of the Peru site, infections appear to be sparsely distributed at the beginning and end of the follow-up period and occur with more density during the middle period suggesting a non-linear association with age. Some clustering of infections at certain times of year are discernible in most sites with the exception of Brazil.

Table 7: Number of study subjects, number, percentage and incidence rate of rotavirus-positive samples in each of the MAL-ED study sites and by vaccine category and sample type[v]

		Diarrheal samples				Monthly samples				Follow-up time		Subjects	
		RV positive	Total	% RV positive	RV incidence rate	RV positive	Total	% RV positive	RV incidence rate	Person-years	Calendar time (months)	Total	% Lost to follow-up
Vaccine introduced	BRF	4	91	4.4	1.1	38	3,624	1.0	10.2	371.7	42.3	227	37.0
	PEL	103	1,892	5.4	21.3	178	6,328	2.8	36.9	482.7	49.7	303	35.6
	SAV	5	108	4.6	1.0	104	5,764	1.8	20.6	505.3	51.9	290	20.0
	Total	**112**	**2,091**	**5.4**	**8.2**	**320**	**15,716**	**2.0**	**23.5**	**1,359.7**	**51.9**	**820**	**30.9**
Vaccine not yet introduced	BGD	339	1,519	22.3	74.1	264	5,282	5.0	57.7	457.6	47.7	265	22.6
	INV	65	486	13.4	14.1	374	5,570	6.7	80.8	462.6	46.8	243	9.5
	NEB	95	822	11.6	20.5	163	5,519	3.0	35.2	462.6	44.3	238	7.6
	PKN	117	2,112	5.5	22.7	172	5,676	3.0	33.3	516.3	48.1	275	10.9
	TZH	17	101	16.8	3.6	239	5,147	4.6	51.2	467.0	47.6	259	23.6
	Total	**633**	**5,040**	**12.6**	**26.8**	**1,212**	**27,194**	**4.5**	**51.2**	**2,366.0**	**50.3**	**1,280**	**14.8**
	Overall total	**745**	**7,131**	**10.4**	**20.0**	**1,532**	**42,910**	**3.6**	**41.1**	**3,725.7**	**51.9**	**2,100**	**20.9**

[v] BGD = Dhaka, Bangladesh; BRF = Fortaleza, Brazil; INV = Vellore, India; NEB = Bhaktapur, Nepal; PKN = Naushero Feroze Pakistan; PEL = Loreto, Peru; SAV = Venda, South Africa; TZH = Haydom, Tanzania. RV = rotavirus, Incidence rates are per 100 person-years.
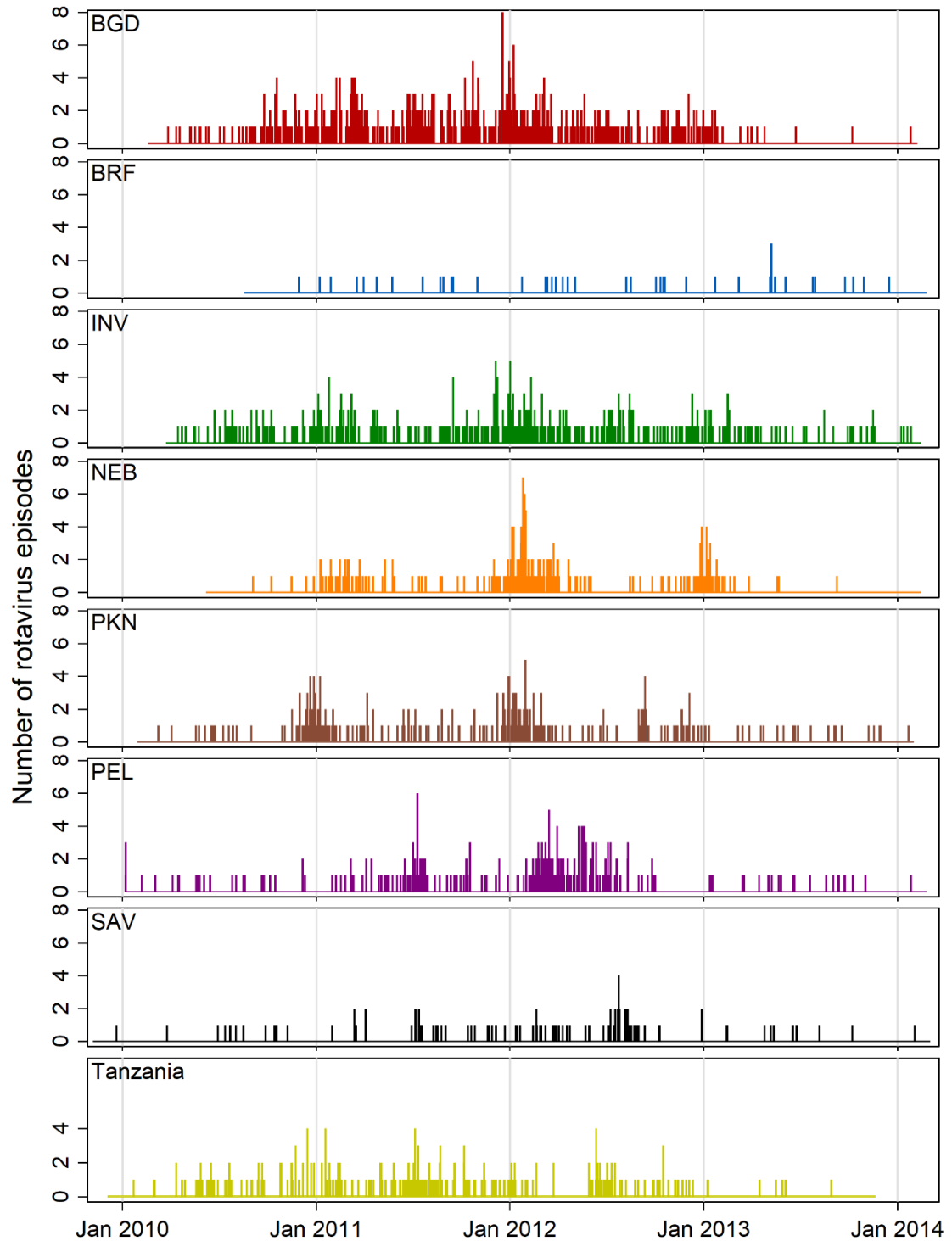
Figure 2: Needle plots of the daily distribution of rotavirus-positive stool samples recorded at each MAL-ED site (rotavirus-negative samples not shown). BGD = Dhaka, Bangladesh; BRF = Fortaleza, Brazil; INV = Vellore, India; NEB = Bhaktapur, Nepal; PKN = Naushero Feroze Pakistan; PEL = Loreto, Peru; SAV = Venda, South Africa; TZH = Haydom, Tanzania.

2.7.Financial support

3. Chapter 3: Exposure variables

The first analysis (chapter 4) is an assessment of seasonality so the day of the year is the main exposure of interest. The second and third (chapters 5 and 6) used hydrometeorological data as the main exposures of interest.

3.1.1. Statistical Methods to Assess Seasonality

A commonly used method for assessing seasonality in health outcomes is to include the month (or season) of the year in which the outcome occurred as indicator variables in a regression model with one month (usually January) as a reference category for comparison.[124] A problem with this approach is that it amounts to taking a continuous predictor – cyclic calendar time – and categorizing it into 12 bins, which are in some sense arbitrary, in order to characterize an underlying smooth, wave-like trajectory and thereby discard potentially explanatory information.[125] Furthermore, the method treats the fixed effect of the average value of a seasonally varying outcome for a given month as independent of the previous month's value, whereas in reality we would expect them to be highly correlated. An alternative method for characterizing continuous, non-linear relationships is to use restricted cubic spline models, however, these are not able to allow for the cyclical nature of seasonal patterns, in which there is continuity between December 31st and January 1st.[126]

A flexible, parametric method for assessing the seasonality of health outcomes that overcomes these limitations is the harmonic method, (sometimes known as a cosinor model[127] or Fourier series[128]) which involves including in the

regression model sinusoidal transformations of the time metric – both sine and cosine functions – to capture the oscillating pattern characteristic of a seasonally varying outcome when plotted over time.[129] This method can be used either to test for the statistical significance of seasonality as a primary exposure and calculate the timing, peak value and amplitude of any annual peaks identified, or to control for seasonality as a potential confounder of a main association of interest.[130] The sinusoidal terms can be introduced in pairs, to capture increasing complexity of a seasonal pattern.[129] The first harmonic terms (sine and cosine pair) are annual, oscillating over a twelve-month period and can detect a basic rotationally symmetrical pattern with a single peak and trough separated by approximately six months. The second, biannual harmonic oscillates over six months and can contribute to capturing the seasonal pattern in two ways. When it is in alignment with the first, it can account for a steeper, more amplified peak than is possible with a single harmonic. In other situations, it can detect the presence of a secondary annual peak (as has been observed, for example, with rotavirus in certain contexts).[56] However, in cases where there is a single annual peak of moderate or low amplitude, the addition of the biannual harmonic may not improve upon an annual-only model and can therefore by excluded on grounds of parsimony.

3.2. Hydrometeorological data

A list of hydrometeorogical variables was compiled that either have been demonstrated or are hypothesized to be associated with EID transmission in general and rotavirus incidence in particular. The following variables were selected based

on a review of the literature: ambient temperature [20,21,60,131–133], rainfall [20,21,133], humidity [21,59,60], atmospheric pressure [19], wind [19,133], UV radiation [19,60,134], soil moisture [135] and water runoff.[132] Next, available data sources were reviewed for daily estimates of these parameters for the eight MAL-ED study site locations over the period of follow-up.

3.2.1. EO-derived data

Estimates of all the above variables are available from the Global Land Data Assimilation System (GLDAS – version 1). GLDAS derives meteorological fields from the Global Data Assimilation System (GDAS), an operational atmospheric analysis system that merges a global climate model – a numerical representation of the physical processes and energy fluxes occurring in the earth's atmosphere, oceans and land surfaces — with a diverse suite of *in situ* and satellite-derived observations.[98,136] The system applies bias correction to GDAS precipitation and radiation estimates and employs the adjusted surface meteorology fields to drive advanced land surface models (LSMs) that simulate surface hydrological conditions. The GLDAS ensemble of LSMs includes the Noah LSM[137], which is implemented in GLDAS at a horizontal resolution of 0.25 decimal degrees (DDs) and parameterized with globally gridded maps of land surface parameters such as topography, land cover and soil texture classifications to produce near-real-time predictions available with global coverage and a temporal resolution of 3 hours.[136,138] Its products have been applied in numerous studies of climate, hydrology, agriculture, and ecology, as well as, more recently, public health outcomes.[76,108] It is internally consistent across locations and between variables (although GLDAS version 1 can suffer from

temporal discontinuities as input datasets change over time).[139] GLDAS data is disseminated as part of the mission of NASA's Earth Science Division and archived and distributed by the Goddard Earth Sciences (GES) Data and Information Services Center (DISC).

Although GLDAS does offer precipitation estimates, it employs a standard correction for bias in the GDAS precipitation field, whereas the Climate Hazards Group Infrared Precipitation with Stations (CHIRPS, version 2) product, which was developed solely to estimate rainfall, calibrates cloud-top temperature estimates and gauge-satellite data by interpolating observation data and weighting it according to proximity to the five closest weather stations.[140] CHIRPS daily data has a resolution of 0.05 DDs (~5km^2) and has the potential to offer greater information content in poorly monitored areas and in tropical regions than alternative, entirely gauge-based products.[140] Precipitation estimates from both sources were evaluated to determine the better-performing estimate. Neither GLDAS or CHIRPS products come with error estimates.

For both GLDAS and CHIRPS, a script was run to extract all variable values from the gridded files during the period 2009 – 2014 for the grid cells corresponding to the coordinates of the eight MAL-ED site locations. For the GLDAS variables, the 3-hourly estimates were aggregated to daily averages, totals or maximum and minimum as appropriate, while the daily estimated rainfall totals were taken from the CHIRPS product.

The following variables were extracted from the two gridded products.

- **Maximum and minimum daily temperatures in degrees Celsius –** Air

  temperature varies as a function of latitude, elevation, and large scale

  atmospheric circulations and is sensitive to prevailing weather patterns and

  differences in local surface conditions. Pathogens are only able to propagate

  within certain temperature ranges, so associations between ambient

  temperature and infectious disease risk could be related to the agents' ability

  to survive in the environment.[31] Laboratory studies have found that various

  non-enveloped enteric viruses including rotavirus and adenovirus survive

  longer at lower temperatures.[141–143] As temperatures increase within the

  range 7 - 37°C, reproduction of *Salmonella* and other bacteria increases,

  outcompeting *Campylobacter*, perhaps causing a shift in the relative burden

  of these diseases.[41] A study comparing daily maximum and minimum land

  surface temperatures in GLDAS with data from weather stations across the

  globe found broad agreement both globally and separately for the regions in

  which the MAL-ED sites are located.[144]

- **Daily total precipitation volume in millimeters –** Precipitation patterns

  vary on a very localized scale due to interactions between energy and water

  fluxes and features of land-sea geometry and topography.[103] Following heavy

  precipitation events, microorganisms that are able to survive in aquatic

  environments may be dispersed over large geographical areas in water

  runoff.[32] Conversely, periods of drought and decreased rainfall have also

  been shown to be associated with increases in rates of diarrheal disease,

  thought to be due to water scarcity that leaves people reliant on unprotected

water sources and unable to maintain hygiene-promoting behaviors.[145] Drought may also lead to an increase in surface water salinity, making aquatic environments less hospitable to most foodborne pathogens, but more favorable to *Vibrio* and *Aeromonas* species.[32] CHIRPS rainfall estimates have been shown to correlate well with *in situ* precipitation records across South America and West Africa for annual rainfall totals[146], across East Africa for springtime averages[147,148], in Mozambique for dekadal totals[149] and in Cyprus for monthly totals.[150]

- **Surface pressure (Pa) -** High surface pressure is frequently associated with still, clear sky conditions, in which mixing of air masses is slow, while low pressure occurs under convective, stormy conditions when winds are high. Such conditions may favor rapid diffusion of airborne particles, including infectious agents, though they may also be associated with rainfall events that scavenge airborne particles from the air.[19] Pressure belts and other mesoscale air circulation patterns may contribute to seasonal patterns in some diseases by restricting or promoting the dispersion of airborne particles including infectious agents such as rotavirus.[19,21]

- **Wind speed (m/s) -** Wind is a possible means of dispersal of virally infective dried fecal particles and for spore-forming bacteria.[32] Using a meteorological standard, GLDAS reports estimates of 10m height winds as an indicator of broader conditions.

- **Humidity** – Humidity is thought to be inversely associated with transmission

  risk for numerous viruses, which may relate to the conditions conducive to

  the survival of the shed virions outside the host, or to their areal transport in

  dried, infective fecal and dust particles.[21] Published laboratory evidence

  about the survival of viruses in the environment under conditions of varying

  relative humidity is contradictory.[142] While one study found that extremes of

  humidity were most propitious for rotavirus survival,[141] another reported

  humidity in the mid-range to be more favorable,[151] while a third found only

  conditions of high humidity to prolong survival[142] a relationship similar to

  that confirmed for norovirus[152] and poliovirus.[153,154] Adenovirus has been

  shown to be able to survive for at least eight weeks at room temperature

  under conditions of low relative humidity.[155] Health facility- and

  surveillance-based studies tend to be in agreement that, in settings in which

  there is a single annual peak in rotavirus incidence, transmission increases as

  the weather becomes less humid.[133,156–158] In settings where there is a

  secondary annual peak, the relationship may be more complicated - the two

  peaks coinciding with the annual extremes of humidity, but in a way that is

  difficult to disentangle from the effects of other, closely-related

  meteorological factors like precipitation and temperature.[56,159] A hospital-

  based study of norovirus case counts found a similar relationship with

  humidity to that of rotavirus[160], but one of the few such studies that looked at

  astrovirus reported a peak that occurred in the season of highest humidity.[161]

Two measures of humidity are widely used in climate science and so were included in this analysis:

- **Specific humidity in kilograms of water per kilogram of air (kg/kg)** - Near-surface specific humidity – the moisture content of the air – is closely related to temperature, distribution of surface water, soil moisture, and evapotranspiration, and so seasonal and zonal variations vary closely alongside these parameters. GLDAS-derived estimates of specific humidity have been demonstrated to have near-perfect correlation (r = 0.98) with *in situ* meteorological observations for a particular location in Northeast China.[162]
- **Relative humidity (%)** – Expressed as a percentage, relative humidity is the ratio of the specific humidity to the saturation specific humidity and is the measure more commonly used in research on determinants of pathogen survival and transmission.[151,163]

- **Solar radiation (W/m²)** – The typical amount of radiation received at a particular location is influenced by the earth's position in its elliptical orbit around the sun and the tilt in the rotation axis.[104] Higher levels of solar radiation may reduce the number of pathogens in the outdoor environment due to the effect of UV radiation inactivating some viruses[60,164], and impeding the survival of some bacteria.[41] Published site-specific evaluation of GLDAS solar radiation estimates are limited, though one study did show high correlation between GLDAS estimates at 3-hour intervals with recordings

from a ground-based flux measurement instrument in South Korea (r = 0.81).[165]

- **Soil moisture (%) –** The moisture content of soil is hypothesized to influence the survival of enteric pathogens in the environment.[46,135] Evaluation of GLDAS soil moisture estimates is also limited, particularly in subtropical and tropical areas. A study of GLDAS estimates of soil moisture using the Noah LSM showed high correlation with station-averaged surface soil moisture data for a for 20–40 cm layer on the central Tibetan Plateau, and another showed excellent correlation with data from a station in South Korea (r = 0.94).[165,166]

- **Surface runoff in millimeters –** The rate at which water drains following precipitation events may affect how microorganisms are dispersed over the landscape.[32] Increased sewage outflows and runoff volumes – particularly following droughts - increase water turbidity causing pathogens from the sediment to re-suspend in surface water bodies, processes that may explain seasonal upticks in waterborne diseases.[32,46] Modelled estimates of surface runoff are challenging to validate since field measurements of this parameter are sparse. However, one study used GLDAS runoff estimates and a source-to-sink river routing scheme to model river discharge at river gauge locations for major basins across the globe with performance varying by region and by LSM.[167]

3.2.2. Weather station data

In the next stage, sources of ground-based observational data were sought that contained equivalent variables to the EO-derived measured at the nearest weather station to each MAL-ED site and covering as much of the MAL-ED follow-up period (2009 – 2014) as was available. Only the one nearest weather station to each site was considered. The data were either retrieved from the National Oceanic and Atmospheric Administration's (NOAA) Climate Data Online repository[168], if there was a National Climatic Data Center-contributing station close to the site, or otherwise were acquired from local meteorological authorities in coordination with site staff. Six of the sites had data available from NOAA for a nearby weather station, and for four of these – Bangladesh, Brazil, Nepal and Peru – the station was located within 7km of the study site (table 6). The nearest weather station to the study site in Pakistan was situated 22km away, the equivalent distance for the site in South Africa was 37km, a scale that is likely to introduce error which will need to be taken into account when interpreting the results. The following variables were available from the NOAA database:[169]

- Maximum and minimum temperature for the day (degrees Fahrenheit to tenths)
- Total precipitation (rain and/or melted snow) reported during the day (inches to hundredths)
- Mean station pressure for the day (millibars to tenths)
- Mean wind speed for the day (knots to tenths)

While NOAA data was not available for the India site during the period of interest, similar data were obtained from the India Meteorological Department from a weather station located approximately 1km from the study site, which included maximum and minimum daily temperature (°C), rainfall (mm) and relative humidity at 8:30am and at 17:30pm Indian Standard Time, but did not include pressure or wind speed. In addition to the variables in the NOAA data, estimates of relative humidity (%) were obtained from the Pakistan Meteorological Department from a station at the same location as the NOAA-contributing station at 0:00am and 12:00pm UTC (5:00am and 5:00pm Pakistan Standard Time). Similarly, site staff in South Africa were able to provide hourly estimates of relative humidity from local authorities for the same station used by NOAA. In Tanzania there were no NOAA weather stations within 260km of the study sites, and the only daily weather data that site staff were able to obtain for nearby were hand-written daily rainfall records from a farm located 32km from the site, which, because they were used in routine monitoring of crop pests, only covered the armyworm moth season running from November to May coinciding with the rainy season. Qualitative reports from site staff indicated that conditions at this farm are slightly drier than at the study site itself. Despite its limitations, this information was digitized and included in the validation exercise.

Temperature and humidity at the South Africa weather station were measured using a Vaisala HUMICAP probe HMP45 D, which has an accuracy at 20°C

of ±0.2°C and ±2%[vi] while precipitation was measured using a tipping bucket rain gauge. No information was available regarding the equipment used at the other weather stations or the limits of uncertainty, distance at which they are believed to be accurate or expected variograms of the climate parameters. All NOAA-contributing weather stations are required to use equipment that conforms to the World Meteorological Association's general meteorological standards and recommended practices.[170]

[vi] Field calibration against references for 0 – 90% relative humidity. For 90-100% relative humidity, ±3%, and against factory references, ±1%.[209]

4. Chapter 4: Seasonality of rotavirus infections in the MAL-ED study

4.1. Background

Diarrheal disease is the second leading global infectious cause of under 5 mortality and can be caused by a variety of viral, bacterial, protozoan and macroparasitic agents.[6,171] Most EIDs exhibit some kind of seasonality, their incidence peaking at a particular point in the annual cycle and receding at others each year.[51] These patterns may vary with latitude and climatic zone and, for diseases with multiple routes of transmission, different mechanisms may come to dominate their relative contribution to overall burden at different points in the annual cycle.[56] Improving understanding of the pathogen-specific seasonality of EID is methodologically challenging but critical to informing policy on the timing of preventive measures and to forecast the impact of climate change on future disease burden trends.[21,42]

As described in chapter 3, seasonality of health outcomes can be assessed using Serfling models, regression models fitted to time series data with a harmonic component, wherein sinusoidal transformations of the time metric are introduced in pairs to capture the oscillating pattern characteristic of seasonally varying outcomes when plotted over time.[129] This approach has the advantage of treating cyclical calendar time as a continuous predictor, rather than categorizing it into 12 bins representing each month of the year.[125] Furthermore, harmonic pairs can be introduced in a stepwise manner, to capture increasing complexity of a seasonal

pattern.[129] This approach was first developed for use on passive surveillance or health-facility based data sources, which tend to cover long time periods but are, by their nature, restricted to reportable morbidity- or mortality-related outcomes which manifest only as severe, symptomatic or fatal cases.[128] Often this also means that the outcome is rare, necessitating aggregation to weekly, monthly or seasonal summary statistics.[16]

For EID episodes, which may be mild, self-limiting or sub-clinical, a promising alternative source of data is active surveillance of community-based cohorts, which can capture the underlying infection status as transmission occurs in the community as well as potential time-varying covariates.[56,109] While such observational studies tend to span a shorter length of follow-up time and are not guaranteed to detect sufficient numbers of cases, Sarkar and colleagues have demonstrated that it is still possible to make statistically valid inferences about pathogen-specific seasonality of EID – namely, rotavirus diarrhoea episodes - by applying the harmonic method to data from a small birth cohort with follow-up spanning less than 5 years.[56] Such analyses must take careful account of the interaction between age and risk of infection, loss to follow-up and within-subject correlation, and, since a consensus has yet to be reached on the precise methodology for doing so, results should be interpreted cautiously and in light of prior findings.

The objective of this chapter is to apply an adapted Serfling approach to data on rotavirus infections from a multi-site cohort study in order to characterize this

EID's seasonality in eight different locations. Infection with this segmented, double-stranded RNA virus is a near ubiquitous feature of infancy, with around 95% of children globally experiencing at least one rotavirus infection before 5 years of age prior to the introduction of the vaccine in, as of the time of writing, 92 countries.[23,26] Recent meta-analyses and review articles have concluded that, while the long-recognized pattern of sharp winter peaks receding to negligible levels in the offseason holds in temperate, mid-latitude regions, tropical regions exhibit more sustained, year-round rotavirus transmission with discernable, but less pronounced relative peaks.[19–21,48,52–54] In tropical South Asia, moreover, biannual peaks in rotavirus are observed, with a large spike in winter followed around six-months later by a smaller, secondary uptick coinciding with the monsoon rains.[20,53,56,57] Questions remain about how consistently these patterns hold across different climatic zones, about the relative contribution of environmental drivers and host-related factors such as asymptomatic infections, and whether year-round transmission in tropical regions acts as a reservoir for the seasonal reintroduction of the virus to temperate zones.[133,172] In this chapter we attempt to address these knowledge gaps using data from MAL-ED.

4.2. Statistical methods

This analysis applied a Serfling model to the binary rotavirus infection status data to test for and quantify the effects of seasonal patterns in prevalence at each MAL-ED site. Because this is a multi-site study, with cohorts selected from communities located in different parts of the globe each with their own seasons and

climate, the analysis was performed separately for each site to model the rotavirus

isolation rate - the proportion of positive stools – over time.[133] This was

approximated by the probability of a stool sample being positive for rotavirus,

which in turn was estimated from the regression coefficient values for equation 1

(adapted from Stolwijk and colleagues):[173]

1) $logitP(Y_{it} = 1) = \beta_0 + \beta_1 \times sin\left(\frac{2\pi t}{365.25}\right) + \beta_2 \times cos\left(\frac{2\pi t}{365.25}\right) + \beta_3 \times$

$sin\left(\frac{4\pi t}{365.25}\right) + \beta_4 \times cos\left(\frac{4\pi t}{365.25}\right) + \sum_j(\gamma_j \times time_t) + \sum_k(\theta_k \times age_{it})$

Where:

t = the date of follow-up

$P(Y_{it} = 1)$ = The probability of a stool from individual i being positive

for rotavirus on date t

β_0 = the mean log odds of positivity over the entire follow-up period

β_1 and β_2 = sine and cosine coefficients for the first harmonic

β_3 and β_4 = sine and cosine coefficients for the second harmonic (to be

included based on a comparison of the Akaike information criterion

(AIC) for the model compared to a null model that includes only the

first harmonics)

γ_j = a series of j cubic spline terms for calendar time in months on

date t (centered at the mid-point of follow-up)

θ_k = a series of k cubic spline terms for individual i's age in months on

date t (centered at 12 months of age)

The model was fitted using generalized linear models (GLM) with robust variance

estimation to account for within-subject clustering of the outcome. The primary

exposures of interest in this longitudinal analysis were the four Fourier series

functions (first and second harmonic sine and cosine transformations of the day of

the year ranging from 0 to 365), which were added as terms to the model with

stepwise selection of the second harmonic pair based on the AIC statistic. The time

metric was continuous calendar time in days, a variable that was included in the

model as a covariate centered on the mid-point between the earliest and latest

observation in the dataset, to adjust for potential secular trends in rotavirus

transmission over the course of follow-up. Restricted cubic spline terms for this

variable were included with degrees of freedom and knot positions determined

separately for each study site using a multivariable regression spline fitting

algorithm.[174] This was to account for the potential confounding effects of isolated

outbreaks, which might give the appearance of seasonality, due to being heavily

clustered over a short period within a single year. The infants' age in continuous

months at the time of the stool sample (centered on the first birthday) was also

included using site-specific cubic splines to account for the non-linear association

between rotavirus risk and age in this cohort.[175] By including this smooth function

for age the cumulative acquisition of immunity within the study population could

also be adjusted for.[130] Infection by a given rotavirus genotype confers only partial,

homotypic immunity that diminishes the severity of, but does not prevent,

subsequent infections.[176] Repeated rotavirus episodes within the same individual

are therefore common, but may be less likely to become clinically apparent in older

infants. Since the cumulative incidence of rotavirus infection necessarily rises

steadily over the first two years of life, age can be used as a proxy for acquired

immunity.

The timing, amplitude and number (single or double) of the annual peaks in

the rotavirus isolation rate were estimated for each study site from the output of the

model. The primary peak was defined as the highest probability predicted by the

model and its amplitude was calculated as the difference between this value and the

lowest predicted probability over the year. The secondary peak was defined as the

peak with the lower maximum value and its amplitude was also calculated relative

to the lowest predicted probability (the global, as opposed to the local minimum).

The overall statistical significance of the seasonal pattern was assessed based on the

Wald test for the combined contribution to the model of all included harmonic

terms. Finally, the seasonal pattern was visualized by plotting the predicted

probability of rotavirus positivity for each study site against the day of the year. The

shapes of these plots were compared to those obtained from lowess smoothed

averages and restricted cubic spline models to assess their fit. The combined

significance of the age terms, the calendar time terms and the harmonic terms were

each assessed using the Wald test while the overall model fit was assessed by

calculating AIC statistics for the final models. To compare the seasonality patterns

between symptomatic and asymptomatic infection episodes, the analysis was

repeated first on only the stool samples obtained according to the monthly schedule,

and secondly on those collected during diarrheal episodes, to approximate the rate of, respectively, symptomatic and asymptomatic transmission. Analyses were carried out using Stata 13.1.[177]

4.3. Results

Table 8 summarizes the Wald test chi squared statistics for the combined significance of the terms for each of the covariate predictors (i.e. the cubic spline terms for age and time and the harmonic terms for seasonality) as well as the seasonality parameters predicted by the logistic model fitted with GLM. The test statistic for the combined contribution of the harmonic terms to the model was highly statistically significant in all sites with the exception of Brazil where it was not significant at the α =0.05 level. In all seven of the eight study sites where the model terms for seasonality attained significance, the primary annual peak occurred outside of the rainy season. In all except two of these (Nepal and Peru), a smaller, secondary annual peak was identified occurring during the rainier part of the year. In all 8 sites, the model that included both the first and second harmonics (4 degrees of freedom) was a better fit for the data than an equivalent that only included the first harmonic according to stepwise selection based on the AIC statistic. The highest predicted probability of rotavirus infection in absolute terms were the single annual peaks in Nepal and Peru and the primary peak in Bangladesh, which all had a predicted maximum value of 16%. The highest amplitude – the largest difference between the highest and lowest annual values – was also seen in Nepal (15 percentage points' difference) followed by Peru (14 percentage points).

Figure 3 shows a graphic visualization of the annual seasonal pattern in rotavirus prevalence predicted by each of the eight site-specific models with the approximate period of the year with the highest precipitation shown in blue-grey. In all four of the South Asian sites, the primary peak in rotavirus prevalence occurred between the end of December and the very beginning of February coinciding with the dry season. In addition, at the Bangladesh, India and Pakistan sites, the models predict a smaller secondary peak during the second half of the year, whereas in Nepal, transmission appears to recede to a very low level during the off-season. In Brazil, for which only a very small number of episodes were recorded in the data set, biannual peaks were also predicted but with very small amplitude values, and without attaining any level of statistical significance. With the data from the Peru site, the model predicted the probability of infection rising from the start of the year to a prolonged peak from April through June that then recedes to its lowest level just prior to the start of the rainier part of the year. The two African sites – Venda, South Africa and Haydom, Tanzania – also have highly statistically significant seasonal patterns which predict mid-year, dry season primary peaks and smaller secondary peaks during the wetter part of the year. Of the sites that exhibited seasonality, Venda, South Africa showed the lowest amplitude for its primary peak.

Table 8: Wald test chi squared statistics for covariate predictors (with degrees of freedom) and seasonality parameters predicted by logistic model fitted with GLM[vii]

		BGD	BRF§	INV	NEB	PKN	PEL§	SAV§	TZH
Calendar time		27.56*** (1)	0.28 (1)	16.63*** (3)	4.46 (2)	0.27 (1)	59.08*** (4)	6.84* (2)	21.42*** (1)
Age		95.45*** (3)	0.11 (1)	21.22*** (4)	25.43*** (4)	0.07 (1)	6.92* (2)	0.02 (1)	29.93*** (3)
Seasonality		42.09*** (4)	5.44 (4)	49.73*** (4)	178.95*** (4)	132.66*** (4)	70.92*** (4)	32.61*** (4)	53.68*** (4)
1st Peak	**Timing**	Jan 4	Apr 10	Jan 22	Jan 25	Jan 5	Jun 8	Aug 3	Jul 4
	Peak value	0.16	0.02	0.14	0.16	0.11	0.16	0.06	0.10
	Amplitude	0.09	0.01	0.09	0.15	0.09	0.14	0.05	0.08
2nd Peak	**Timing**	Jul 20	Oct 4	Aug 6	-	Jul 9	-	Feb 16	Dec 20
	Peak value	0.08	0.02	0.08	-	0.03	-	0.03	0.07
	Amplitude	0.01	0.01	0.03	-	0.02	-	0.02	0.05
AIC statistic		3,897.04	466.35	3,070.77	1,860.22	2,336.29	2,301.73	1,054.29	1,943.40

Figure 3: Predicted probability of rotavirus infection by day of the year with 95% confidence intervals, Wald test chi squared statistics and degrees of freedom (d.f.) for harmonic terms (*** p <0.001, ** p = 0.001 – 0.01, * p = 0.01 - 0.05). Local rainy seasons are shaded blue-grey. BGD = Dhaka, Bangladesh; BRF = Fortaleza, Brazil; INV = Vellore, India; NEB = Bhaktapur, Nepal; PKN = Naushero Feroze Pakistan; PEL = Loreto, Peru; SAV = Venda, South Africa; TZH = Haydom, Tanzania (age held constant at 1 year and time at the mid-point of follow-up).

Table 9 presents the equivalent seasonality and model statistics resulting from restricting the analysis first to only the monthly samples, and then to the diarrheal samples. In all sites, the primary peak in rotavirus prevalence in diarrheal stools was more pronounced than in monthly stools. However, the overall pattern, timing and number of peaks was broadly similar across the three analyses, with some notable exceptions. In the Bangladesh site, the secondary, monsoon season peak that was evident in the pooled and diarrheal stools-only analysis was not apparent in the monthly stools. In Tanzania, the pattern in diarrheal stools only demonstrated one annual peak and was not statistically significant.

Table 9: Wald test chi squared statistics for harmonic terms (with degrees of freedom) and seasonality parameters predicted by logistic model fitted with GLM by sample type[viii]

			BGD	BRF[§]	INV	NEB	PKN	PEL[§]	SAV[§]	TZH
Monthly samples only		**Seasonality**	18.72*** (2)	4.76 (4)	34.67*** (4)	125.00*** (4)	59.38*** (4)	36.30*** (4)	28.19*** (4)	56.83*** (4)
	1st Peak	**Timing**	Jan 4	Apr 9	Jan 21	Feb 1	Jan 8	Apr 12	Aug 3	Jul 5
		Peak value	0.07	0.02	0.10	0.10	0.08	0.04	0.04	0.10
		Amplitude	0.04	0.01	0.06	0.10	0.07	0.03	0.04	0.08
	2nd Peak	**Timing**	-	Oct 6	Aug 3	-	Jun 30	-	Feb 18	Dec 17
		Peak value	-	0.01	0.06	-	0.02	-	0.02	0.06
		Amplitude	-	0.01	0.02	-	0.01	-	0.01	0.05
		AIC statistic	2,021.10	429.59	2,680.83	1,289.09	1,482.28	1,557.54	1,020.53	1,837.92
Diarrheal samples only		**Seasonality**	36.85*** (4)	0.31 (2)	17.17** (4)	86.57*** (4)	66.50*** (4)	34.96*** (2)	19.21*** (4)	4.47 (2)
	1st Peak	**Timing**	Jan 3	Jul 30	Jan 24	Jan 14	Dec 28	May 26	Jul 28	Apr 30
		Peak value	0.43	0.07	0.48	0.43	0.20	0.22	0.53	0.28
		Amplitude	0.26	0.04	0.35	0.40	0.19	0.20	0.53	0.23
	2nd Peak	**Timing**	Jul 18	-	Aug 8	-	Jul 15	-	Jan 5	-
		Peak value	0.29	-	0.22	-	0.08	-	0.09	-
		Amplitude	0.13	-	0.10	-	0.07	-	0.09	-
		AIC statistic	1,504.57	42.30	362.23	493.72	818.07	713.21	37.14	96.14

[viii] *** p <0.001, ** p = 0.001 – 0.01,* p = 0.01 - 0.05. Numbers in parentheses indicate degrees of freedom – the number of model terms selected for that covariate. § Country in which the rotavirus vaccine had been introduced at the time of the study. BGD = Dhaka, Bangladesh; BRF = Fortaleza, Brazil; INV = Vellore, India; NEB = Bhaktapur, Nepal; PKN = Naushero Feroze, Pakistan; PEL = Loreto, Peru; SAV = Venda, South Africa; TZH = Haydom, Tanzania.

4.4. Discussion

In this analysis, we applied a flexible statistical approach to characterizing cyclical patterns in prevalence to data on rotavirus infection status in the eight MAL-ED birth cohorts and obtained results that are consistent with those reported in prior literature, while also being suggestive of novel insights that merit further exploration. In the four South Asian sites, our model predicted marked primary peaks in the December to January dry season, congruent with a hypothesized inverse relationship with air humidity.[178] Furthermore, at the Bangladesh, India and Pakistan sites, secondary monsoon-season peaks were predicted, in line with similar findings documented in the literature.[56] In two of the sites in rotavirus vaccine countries – the urban community in Brazil with relatively widespread access to improved water and sanitation[116], and the remote, rural location in South Africa[114] – the amplitude of the seasonality predicted by the model was much smaller than elsewhere (and, in the case of Brazil, not statistically significant) in a way that was proportionate to the lower background endemicity of the virus in those locations. This is consistent with the documented "blunting" of the annual rotavirus peaks following the introduction of the vaccine.[24] While the peak at the Peru site occurred during the drier part of the year, this apparent correlation should be interpreted with caution, since that location is subject to year-round rainfall, which means that the rainy season is far less marked than at the other sites.[120] In Tanzania, the biannual peaks predicted by the model both had a similar amplitude

68

with the slightly larger of the two coinciding with the mid-year dry winter and the smaller with the November to May rainy season [111].

Seasonal patterns of rotavirus positivity were broadly similar when the analysis was repeated on both the monthly and the diarrheal stool subsets (with the exception of the diarrheal stools at the Tanzania site, which may be explained by their representing by far the smallest proportion of the overall data at any site). The amplitude of the peaks in diarrheal stools appear to have a greater magnitude, but this analysis did not adjust for the underlying seasonal variation in diarrheal episodes of any etiology. Our results suggest that symptomatic and sub-clinical rotavirus episodes do not differ substantially in their seasonal patterns and contribute roughly equally to the overall annual trend.

Since human rotavirus has no known animal reservoir, it has been speculated that tropical regions with year-round endemic transmission are the main reservoir from which the virus spreads poleward each year at the start of the rotavirus season when conditions favor its reintroduction into temperate zones.[172] Findings from Török and colleagues that the start of the annual rotavirus season in the United States occurs earliest in the southwest and progresses sequentially northeastward over the course of the winter, lend weight to this theory.[179] Some of the results presented here are consistent with this hypothesis. In the higher latitude sites of Nepal, Pakistan and South Africa, off-season transmission appears to recede to negligible levels, while the more tropical Bangladesh, India and, to a lesser extent, Tanzania sites experience more sustained low season transmission. The exceptions

to this pattern is the Brazil site (perhaps due to the success of the vaccine in reducing transmission there) a tropical location that nonetheless has low transmission throughout the year.

Although we have identified statistically significant seasonal patterns, caution should be exercised in deriving inferences about the underlying pathways driving these trends, particularly when extending the approach to other pathogens. While these results may seem highly suggestive of an overlapping environmental influence, such as cyclical changes in the weather constraining the transmission and survival of the disease agent, the influence of co-seasonal social, behavioral and immunological determinants of transmission risk cannot be ruled out on the basis of these findings alone.[16,41,42] In reality, the true mechanisms underlying infectious disease seasonality include multifarious environmental, behavioral and immunological drivers that are specific to particular pathogens, their human (and non-human) hosts and their locations.[42] These can interact to produce subtle periodic fluctuations in either the reproductive number or the fraction of the population susceptible to infection at a given time of the year.[16] For some diseases, the influence of the weather may indeed be fairly direct, whereas, for others they may be mediated through changes in behavior of the human host such as rainfall altering patterns of contact between infected and susceptible individuals as they congregate indoors.[44] Furthermore, even for pathogens for which weather has a direct influence on transmission, competing mechanisms may come to dominate their relative contribution to overall burden at different points in the annual cycle. The example of rotavirus in south Asia is illustrative of this. The large winter peak

noted here and elsewhere in the literature may be accounted for by airborne transmission on dried dust substrates, while the smaller monsoon peak may be due to wider dispersal of the virus in floodwater and runoff promoting water-borne transmission.[21] Analysis of the influence of specific seasonally varying climate variables (beyond basic comparisons with the timing of the site-specific rainy seasons presented here) are beyond the scope of this paper, but have the potential to provide further insights into rotavirus transmission dynamics and will be explored in subsequent MAL-ED publications (including in chapters 5 and 6).

The approach presented here is sufficiently flexible to be adapted to other pathogens or outcomes in the MAL-ED data set or to similar studies in other locations, to further characterize and explore seasonal patterns in their occurrence. The Serfling approach may be applied to other ways of modeling these outcomes, such as log incidence[56], case counts (modeled as a Poisson distribution)[31] or a hazard function in a survival analysis. Furthermore, the method can be used not just to test for the statistical significance of seasonality but also to control for it as a potential confounder of a main association of interest.[130] Investigations into the association between climate and seasonally varying health outcomes may incorporate harmonic terms to assess whether they retain their statistical significance in the presence of specific hydro-meteorological parameters, a sign that host factors and other non-environmental drivers may be playing a causal role. Future research into the relative contribution of these causal pathways may also adjust for confounding by behavioral and socio-demographic factors such as access to improved water and sanitation, socio-economic status, dietary intake, fertility

rates and contact with animals as well as highly temporally disaggregated climate data.[53]

The findings presented here have practical implications for the evaluation of programmes and interventions that aim to reduce EID incidence. Attempts to quantify the impact of such measures should take into account the point in the seasonal cycle at which the outcome is ascertained in order to avoid misattributing to the intervention trends that are in fact consistent with the normal annual patterns. As this study demonstrates, these patterns can vary considerably between neighboring countries – as is evident from comparing the results for the Peru and Brazil MAL-ED sites - or show similarities across large geographic areas – as seen in the four South Asian sites – so knowledge of the local context is critical.

5. Chapter 5: Evaluating meteorological data from weather stations, and from satellites and global models for a multi-site epidemiological study

5.1. Background

Climate and weather influence population health through a number of interrelated pathways. Extreme weather events such as heatwaves, coastal floods and storm surges can both cause mortality directly and can compromise water sources and the crop production, leading to widespread food and water insecurity, illness, undernutrition and other morbidities.[106] Moreover, climate is one of the primary constraints on the geographic and seasonal distribution of pollutants[180] and infectious agents.[41] The growth, survival and dispersal of microorganisms and the viable range of their intermediary hosts and vectors is determined by environmental and hydrometeorological conditions.[32] An increased awareness of the knowledge gaps surrounding these relationships, as well as the urgency of the climate change threat and greater understanding of its likely impact on public health has spurred calls for a research agenda to elucidate the interactions and biological mechanisms through which weather influences health.[101,181] A major barrier to this is the scarcity of empirical data linking climate and health at a sufficient level of spatiotemporal disaggregation for use in longitudinal and time series regression analyses.[107] To isolate interactions between the numerous, collinear climatic variables, quantify annual cycles and long-term trends, and incorporate lag effects, the health outcome and environmental exposure must be matched by their precise timing.[19,48,53,107] Until recently, such analyses were hindered by the difficulty of

accessing accurate and complete data on hydrometeorological predictors at high temporal resolution. The increased accessibility of Earth Observation (EO) climate data products – those derived from satellites and model-based reanalysis - is beginning to change this, but uptake has been slow due to a lack of interdisciplinary collaboration between the planetary sciences and public health fields.[76,101,108,182]

Researchers wishing to include climate variables as predictors in analyses of health outcomes generally have two options: to use either EO-derived or station-based data. The former have the advantage of completeness, both temporal and spatial. Estimates may be available at a daily or even sub-hourly resolution[183] without gaps and can be extracted for any location for which the geographical coordinates are known or a relevant geographic area can be mapped. Many also offer a larger suite of mutually consistent variables than are typically available from weather stations, and the data are often freely available to access online. Disadvantages include the wide variation in the uncertainty of the estimates.[184]

Weather conditions recorded at ground-based stations may be considered the gold standard for meteorological data, insofar as one exists, but are also subject to limitations. Lack of capacity to maintain routine record keeping may lead to significant data gaps, forcing researchers either to exclude outcome data for which no coincident exposure measures are available thus reducing statistical power, or to rely on summary measures such as moving mean values or binned aggregates, reducing variability and temporal resolution. Furthermore, weather stations are often situated in locations key to their primary uses in aviation or in monitoring

weather for large population centers (i.e. cities and airports) and may be more geographically representative of some areas than others. Epidemiological surveillance sites may lie many kilometers from their nearest weather stations, distances greater than those over which localized meteorological conditions vary, introducing further error. Accessing data may be a challenge and, while NOAA offers a substantial online repository of historical data for some 9,000 stations around the globe, for less well-served locations coordination with local meteorological agencies and organizations on the ground may be required.[168] Finally, weather stations vary in the accuracy and generally only record a small subset of variables – often only temperature, rainfall, pressure and wind speed - and more technically demanding measures, such as humidity and solar radiation, may be lacking.

The aim of this chapter is to report on an exercise in selecting climate data products and assessing their performance both in characterizing meteorological conditions at the specific locations of epidemiological study sites and as predictors of a known climate-sensitive outcome – namely rotavirus infection episodes. The hypothesis that we aim to test is that gridded, EO-derived climate data products can be used as valid surrogates in longitudinal analyses where ground-based measurements are unavailable or incomplete to predict health outcomes at particular locations. As an illustrative case study, we use the eight study sites of the MAL-ED project and focus on variables that we hypothesize to be associated with EID transmission.[109] The hydrometeorological variables used in this analysis were described in chapter 3.

5.2. Statistical methods

Numerous statistical metrics are commonly used to assess the performance of EO-derived estimates of hydrometeorological parameters relative to ground- or station-based measurements. The four metrics chosen for this study were:

- Pearson's correlation coefficient (R) – This measure of linear dependence between two variables is widely used and easily interpreted, taking a value between -1 and 1 with 1 indicating perfect positive linear correlation.[185]

- Nash-Sutcliffe efficiency coefficient (NSE) – This normalized indicator of model efficiency corresponds to the statistical agreement or skill of the estimates relative to the observed measurements and takes a value ranging from minus infinity to one, with one being a perfect fit and negative values meaning that the station mean offers a better estimate.[144,149,186–188]

- Mean bias error (MBE) – This measures the extent to which the estimated value deviates from the observed value [189]. It can take any value, with negative values indicating systematic under-estimation and positive values, over-estimation, and zero indicating a lack of overall bias.[187,189]

- Root mean square error (RMSE) – This is an absolute measure of the overall error in the estimates relative to the observed values, expressed in the same units and scale as the data itself.[189,190] It can take any positive value with zero indicating a perfect lack of error[187]

In addition, to assess the ability of the EO data to characterize to extremes of each parameter, the following metrics were calculated with respect to days in which

the parameter value measured by the weather stations exceeded the 80th percentile of the distribution of that parameter over the 6-year period:

- True positive rate (TPR) – The proportion of days classified as extreme (>80th percentile) for a particular parameter by the weather station that were also classified as such by the EO data (equivalent to the sensitivity of a diagnostic test in epidemiology).
- False positive rate (FPR) – The proportion of days not classified as extreme by the weather station that were nevertheless classified as such by the EO data (equivalent to 1 – sensitivity).

All variables that were not in metric units were converted to their metric equivalents. The median daily temperatures were calculated from the maximum and minimum daily temperatures for both the station-based and gridded datasets. For Pakistan and India, the observed average daily relative humidity was approximated by taking the average of the station-based estimates for the two times that were available, whereas for South Africa, the daily averages of the hourly estimates were used. Surface pressure was expressed in millibars. The EO-derived data were found to be stable throughout the 2009-2014 period considered in this study at all sites and for all variables, with the exception of a slight discontinuity in the GLDAS surface pressure at the Tanzania site. This discontinuity in the original data was adjusted for in the data presented here by adding a simple offset to the second half

of the record to align it with the previous period. A small number of implausible

outlying values were dropped from the full dataset.[ix]

As an initial exploration of the data, each of the hydrometeorological

variables was plotted in time series alongside the station-based equivalents where

available for each site. Next, the EO-derived values were each plotted against their

station-based equivalents in scatterplots to visualize the fit between the two. Then,

the evaluation metrics were calculated separately for each variable and site, first for

the raw daily values and then for the mean of the values over seven days to

determine whether averaging over this period improved the performance of the

variables. The analysis compared EO-derived data extracted from the exact site

location with stations within varying proximity in order to reflect the scenario

realistically faced by epidemiologists in which a study site may be located some

distance from its nearest weather station. As a basic method for evaluating the

products in the absence of a seasonal cycle, these values were then recalculated

after restricting the data to only the site-specific season of peak rotavirus

transmission (for sites that experienced multiple peaks in transmission during the

annual cycle, the primary peak of highest amplitude was used). This was to control

for the sensitivity of the evaluation metrics to seasonal variation.

Finally, to test the relative ability of each data source and variable to predict

a climate-sensitive health outcome, logistic regression models were fitted to

rotavirus infections status across all sites combined using GLM with each of the

---

[ix] 11 observations, <0.0001% of the overall data

meteorological variables in turn as the main exposure, lagged by 3 days

(representing the estimated 2-day incubation period[191], plus 24 hours to report

symptoms). In all models, the main association was adjusted for age in continuous

months, seasonality - using annual and biannual Fourier series functions to account

for multiple peaks within the year[58] – and calendar time, each with site-specific

interactions. For each hydrometeorological variable, the model was fitted first using

the EO-derived data then, where available, the weather station data and third, a

combination of the two in which missing station data was substituted with its EO-

derived equivalent and compared results between daily and 7-day mean values. The

purpose of this was to assess the sensitivity of the prediction models to differences

in the data sources and period of aggregation. Odds ratios for these associations are

reported alongside their 95% confidence intervals. Potential non-linearity,

distributed lag effects, mediation and interaction among variables will be explored

in subsequent MAL-ED publications but are beyond the scope of this paper. Analyses

were carried out using Stata 13.1.[177]

5.3. Results

Figures 4 – 13 are time-series plots of each variable by source while table 16

in Appendix 1 summarizes the nine hydrometeorological variables for the eight

MAL-ED sites. The weather station data for Bangladesh had the most missing data of

the NOAA datasets, with just under 60% of the daily estimates for the period

available for each of the four variables - temperature, precipitation, pressure and

wind speed. The remaining NOAA datasets had fairly complete (>90%) data on

temperature and wind speed, while precipitation data were only below 90%

completeness in Nepal and South Africa. Daily data on surface pressure were

extremely sparse (~1%) for Nepal and Pakistan (estimates only available for 25 and

21 days respectively) and somewhat incomplete for Bangladesh and Brazil (59.1%

and 70.0% respectively). The data on relative humidity from the local

meteorological authorities were fully complete for Pakistan, and fairly complete for

India, but only somewhat complete for South Africa. As previously described, the

only station-based variable available for Tanzania was precipitation. These data

were only available for 36.4% of the days in the period of interest, representing only

the four November to May rainy seasons from 2010 to 2014. No *in situ* data on

specific humidity, solar radiation, soil moisture or surface runoff were available for

any of the 8 sites. These findings serve to underscore the fact that weather station

data varies widely in scope, completeness, accessibility and spatial resolution.
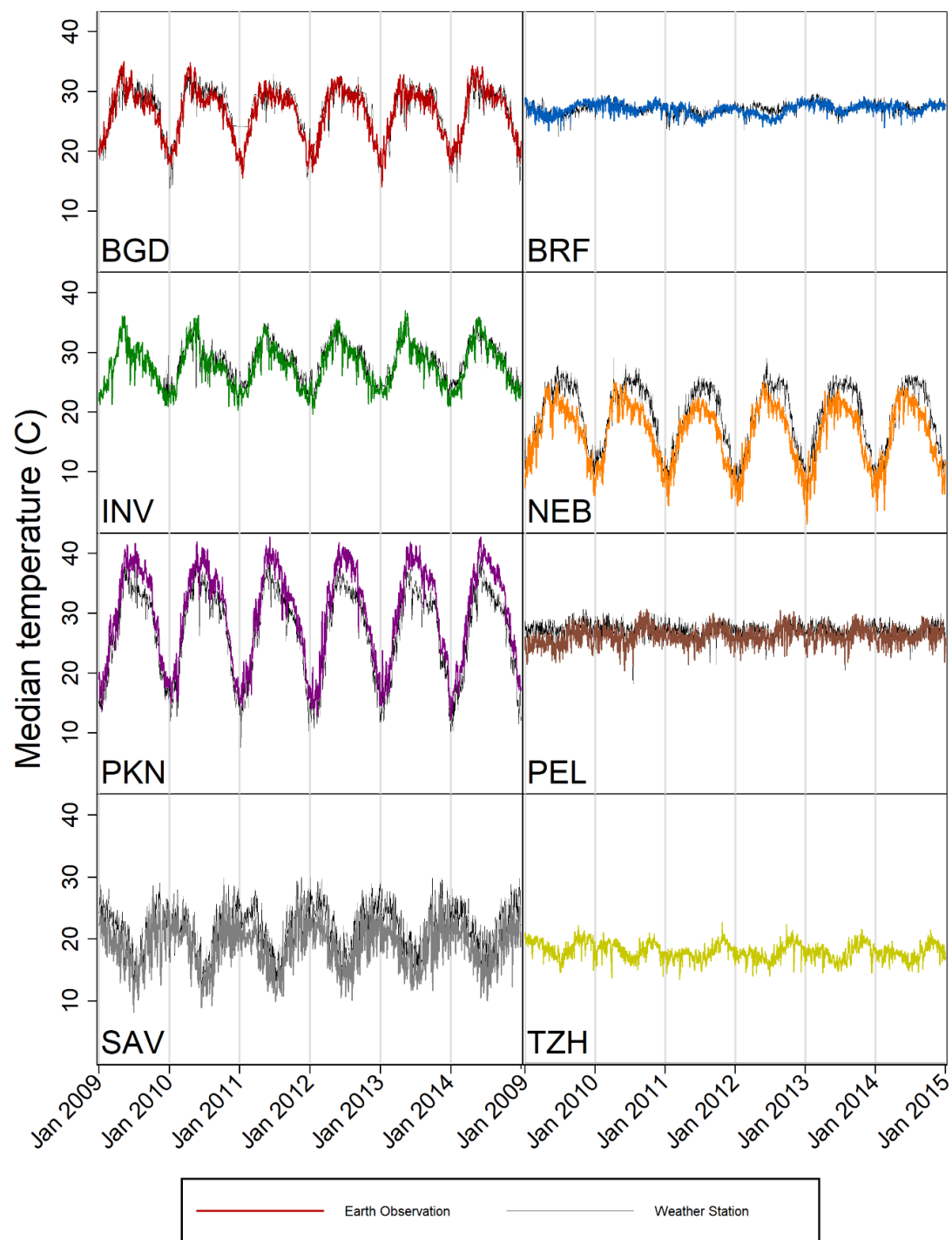
Figure 4: Daily median temperature (C) estimates by MAL-ED site, 2009 - 2014
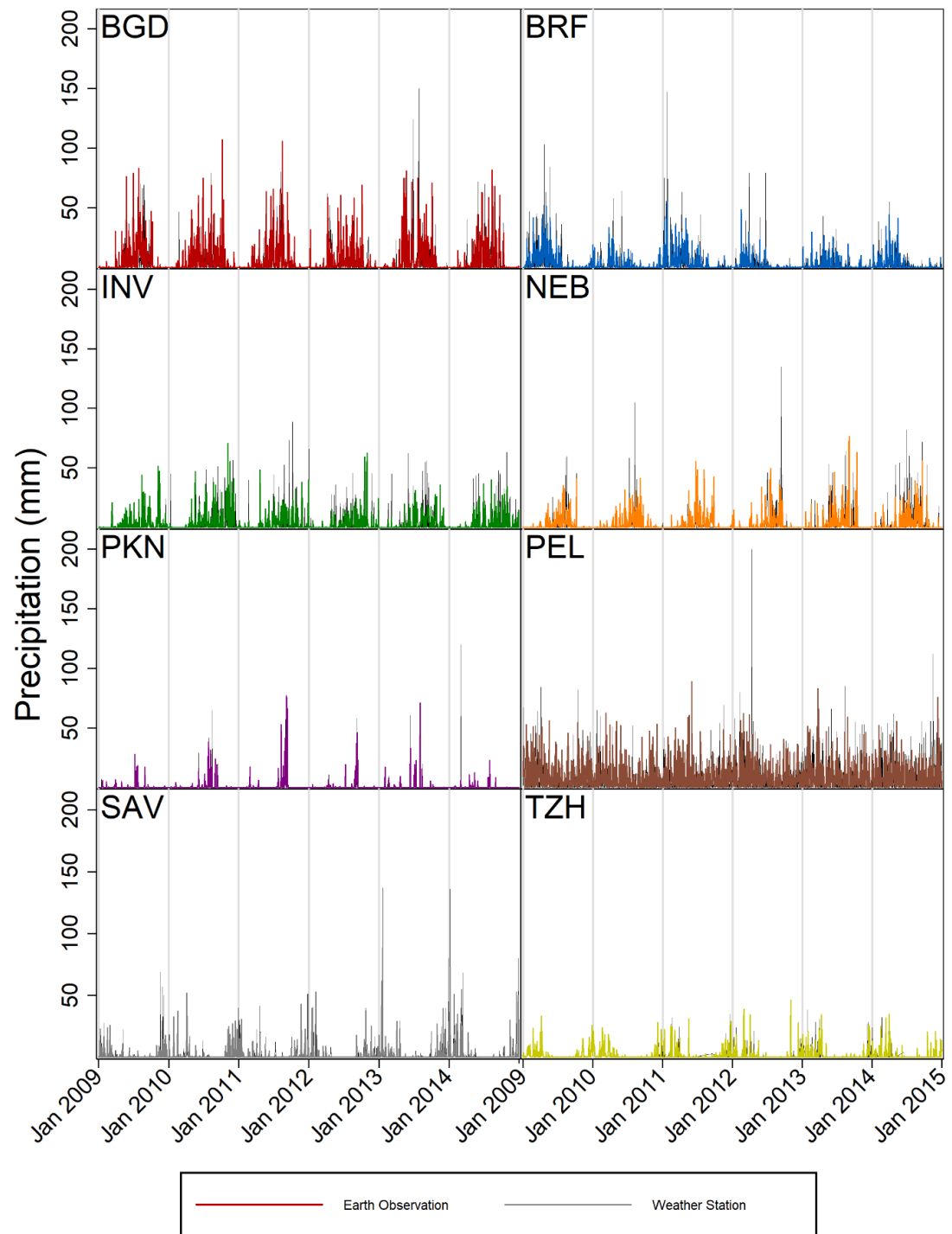
Figure 5: Daily precipitation (GLDAS - mm) estimates by MAL-ED site, 2009 - 2014
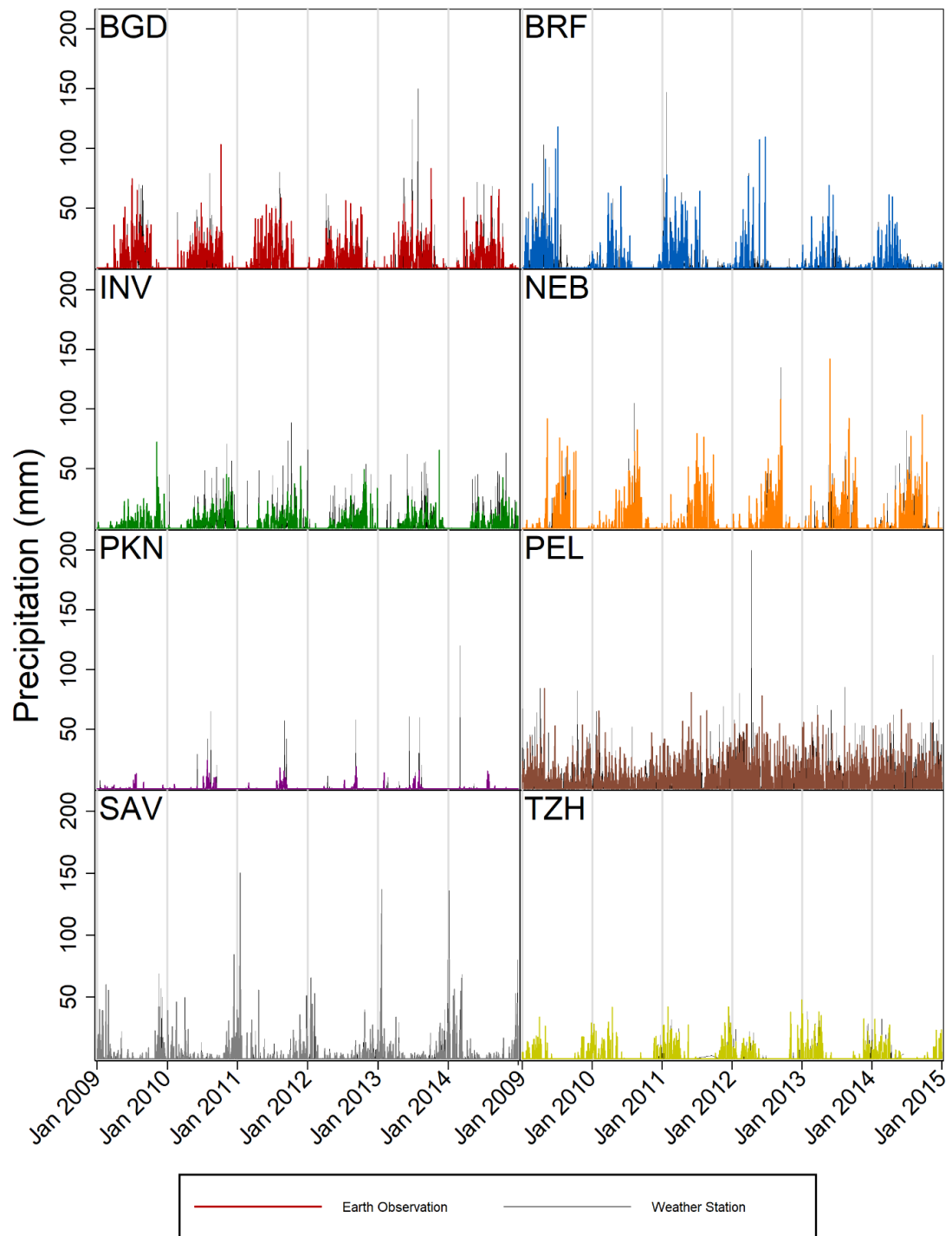
Figure 6: Daily precipitation (CHIRPS - mm) estimates by MAL-ED site, 2009 - 2014
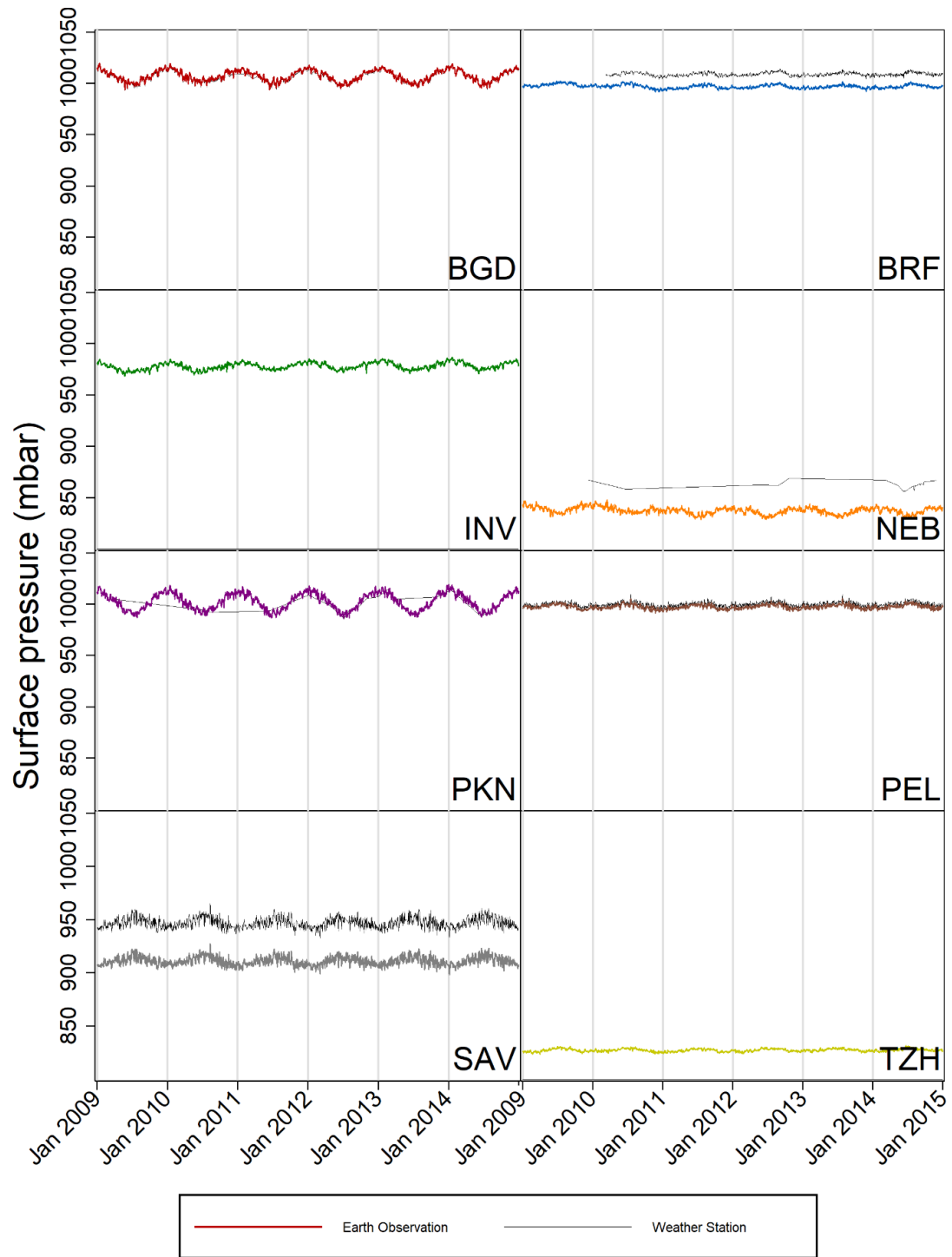
Figure 7: Daily surface pressure (mbar) estimates by MAL-ED site, 2009 – 2014 (with an offset applied to correct for a temporal discontinuity in Tanzania)
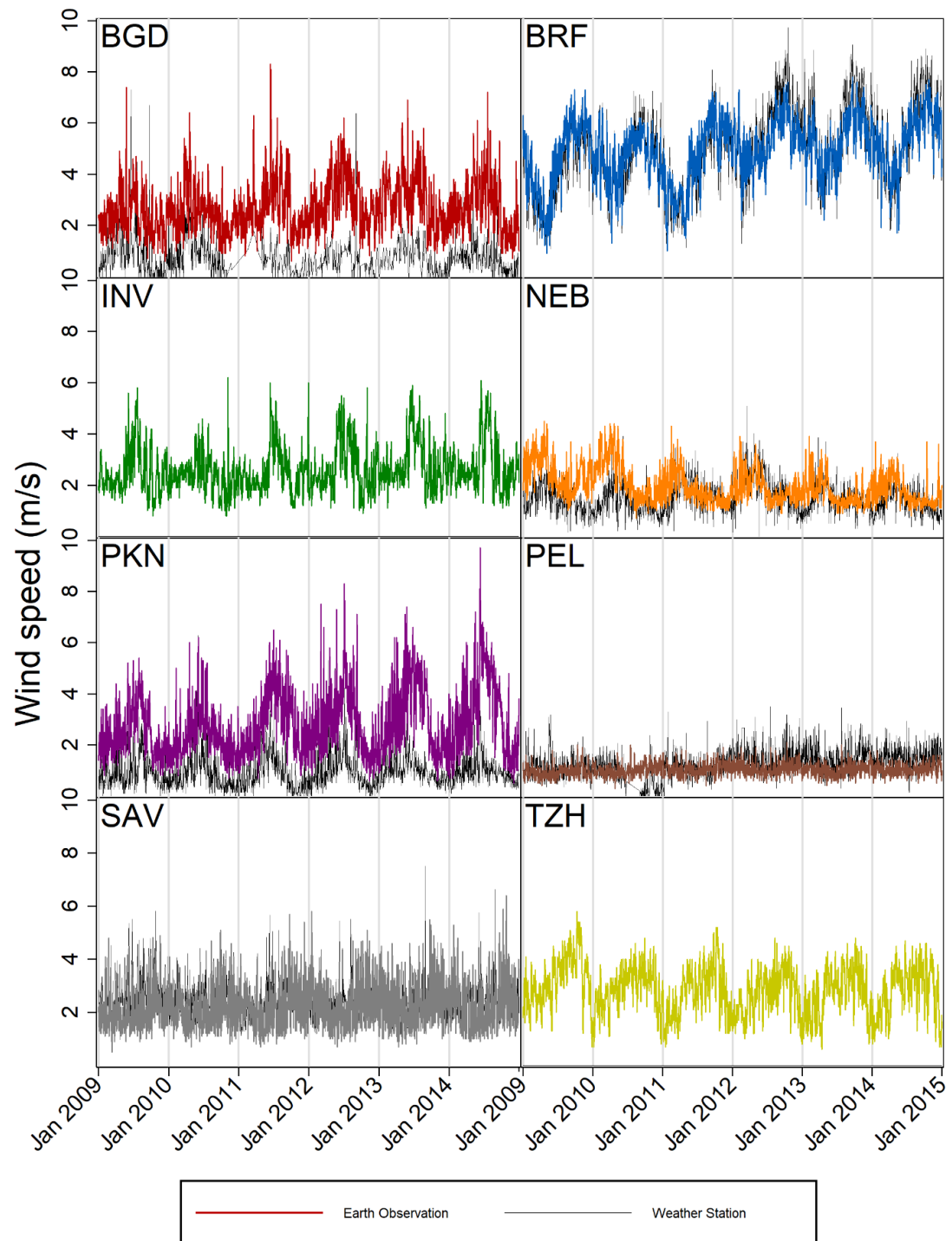
Figure 8: Daily wind speed (m/s) estimates by MAL-ED site, 2009 - 2014
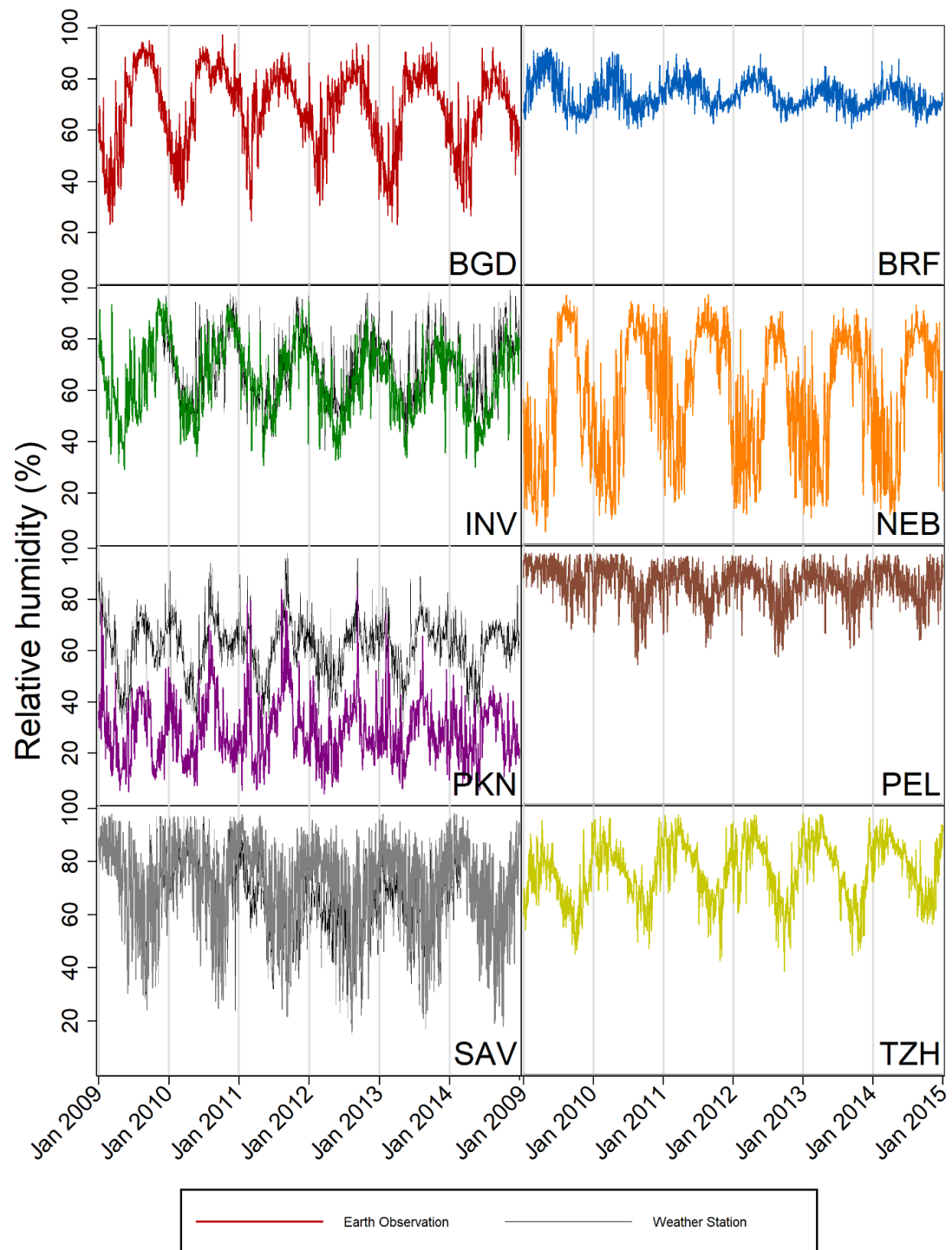
Figure 9: Daily relative humidity (%) estimates by MAL-ED site, 2009 - 2014
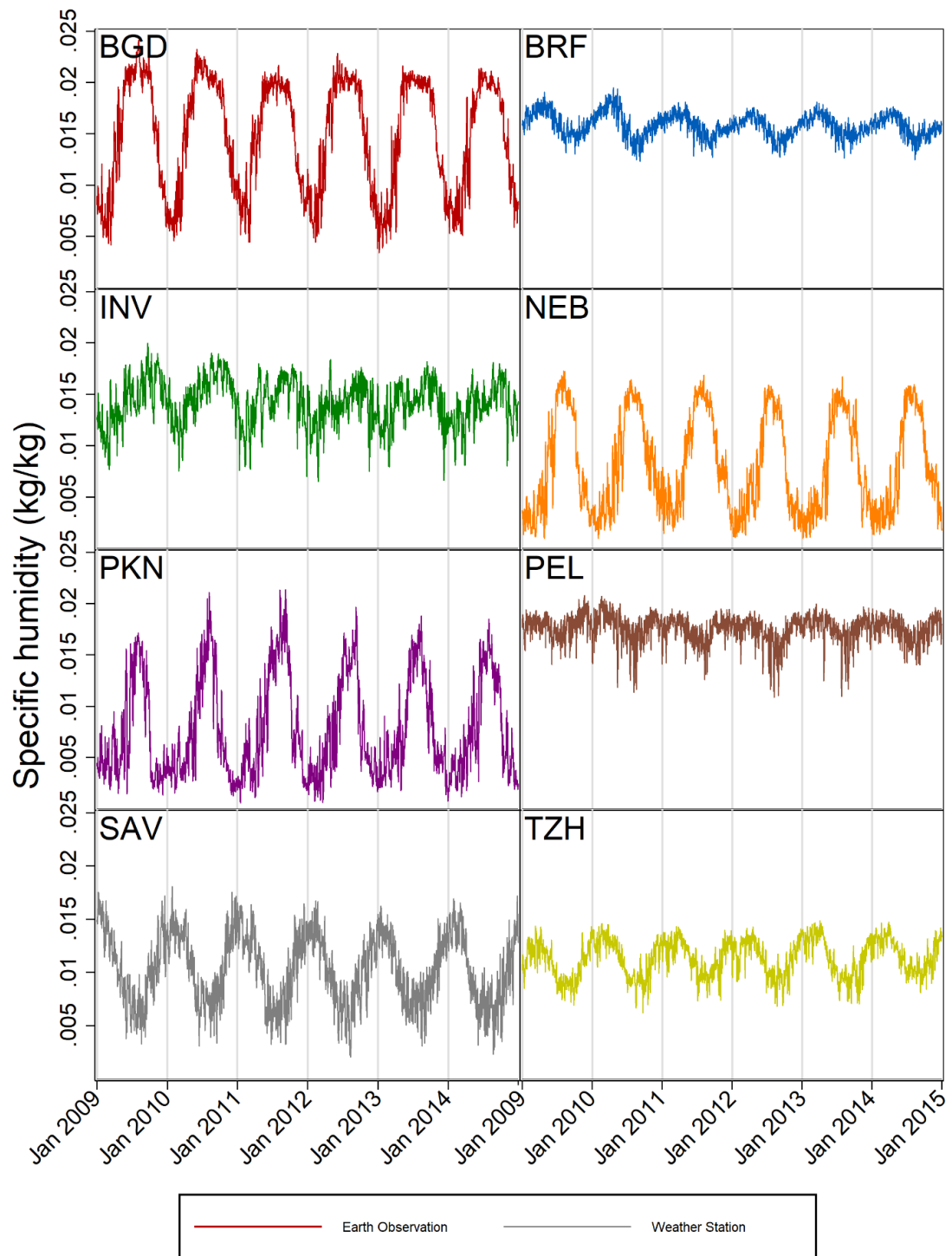
Figure 10: Daily specific humidity (kg/kg) estimates by MAL-ED site, 2009 - 2014
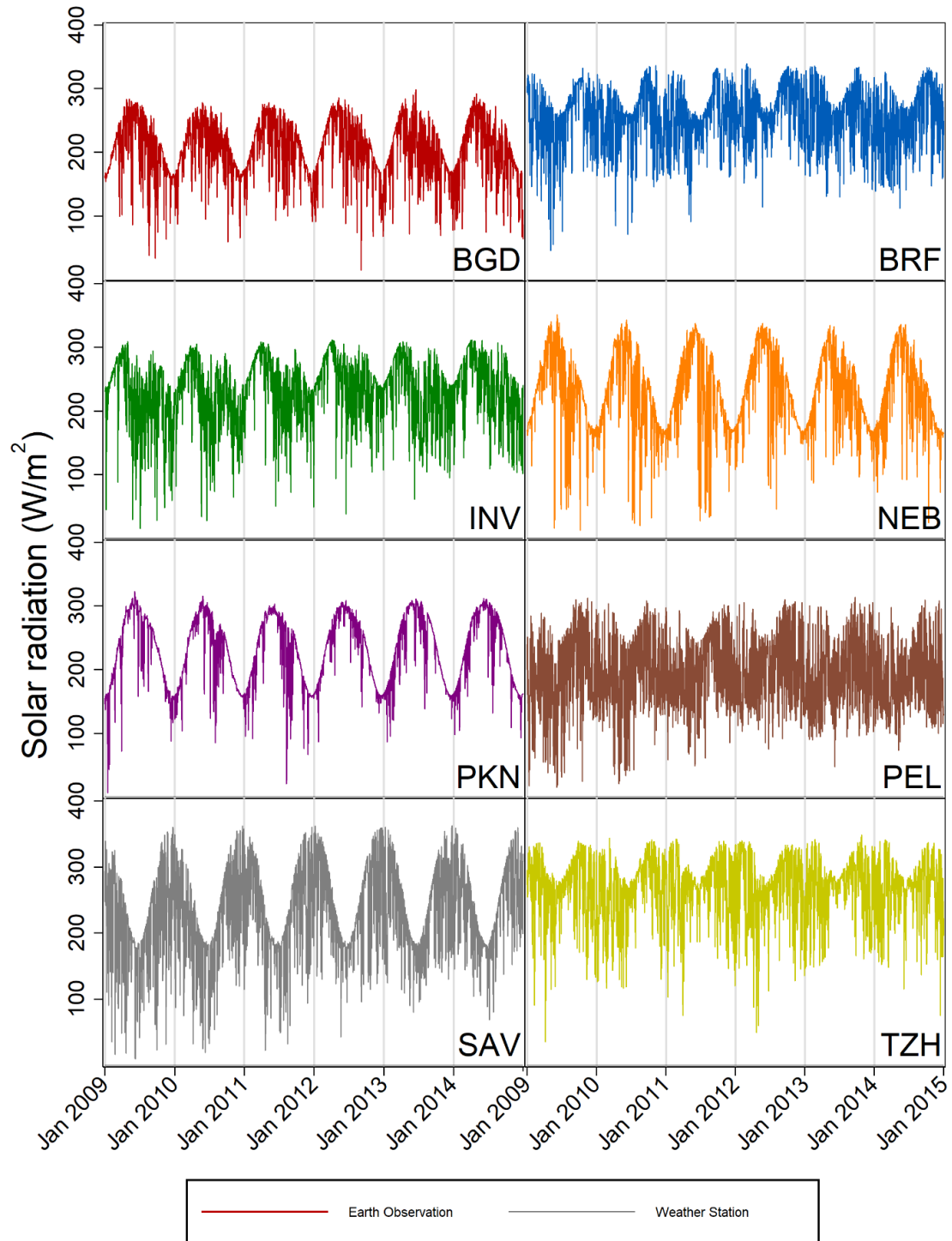
Figure 11: Daily solar radiation (W/m²) estimates by MAL-ED site, 2009 - 2014
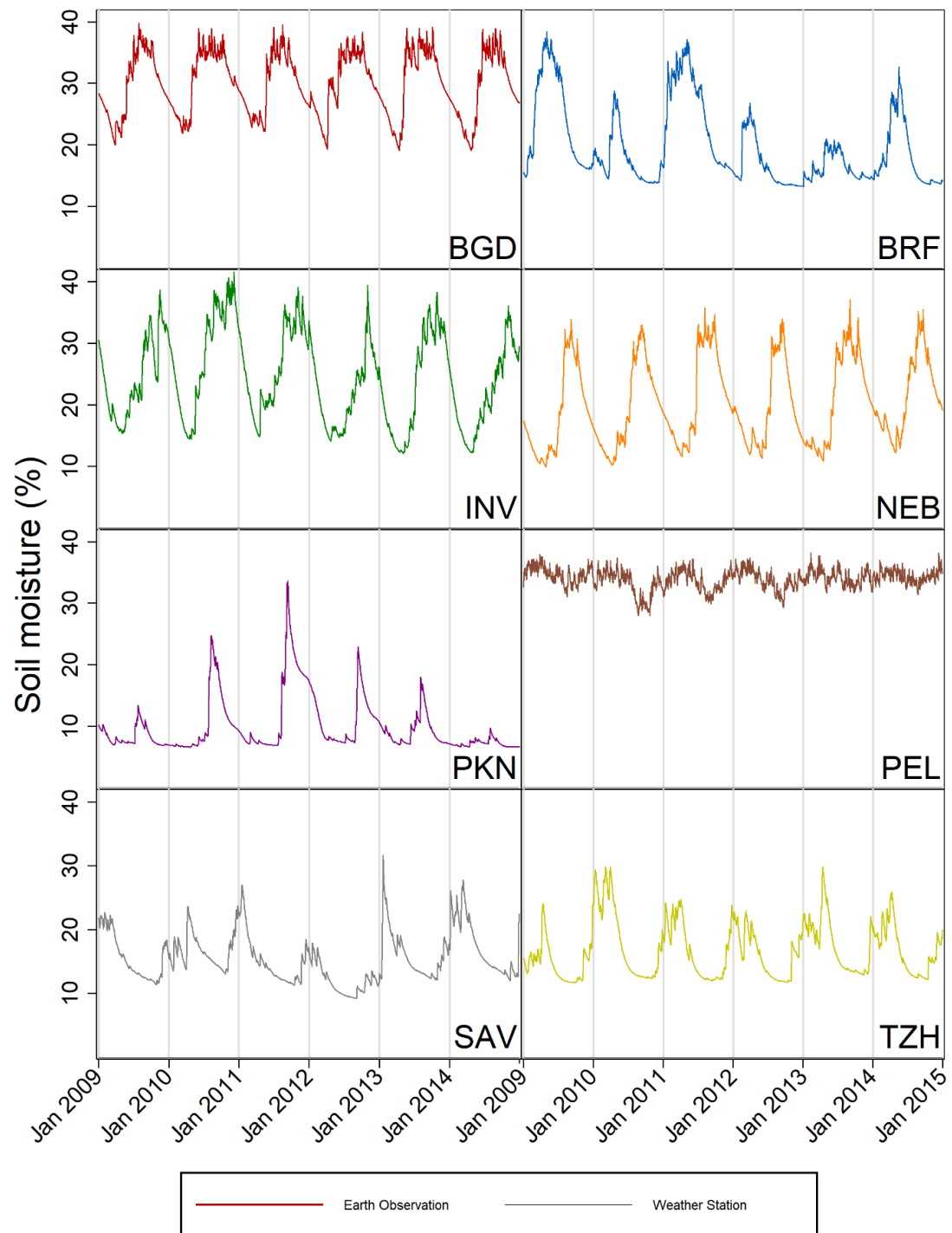
Figure 12: Daily soil moisture (%) estimates by MAL-ED site, 2009 - 2014
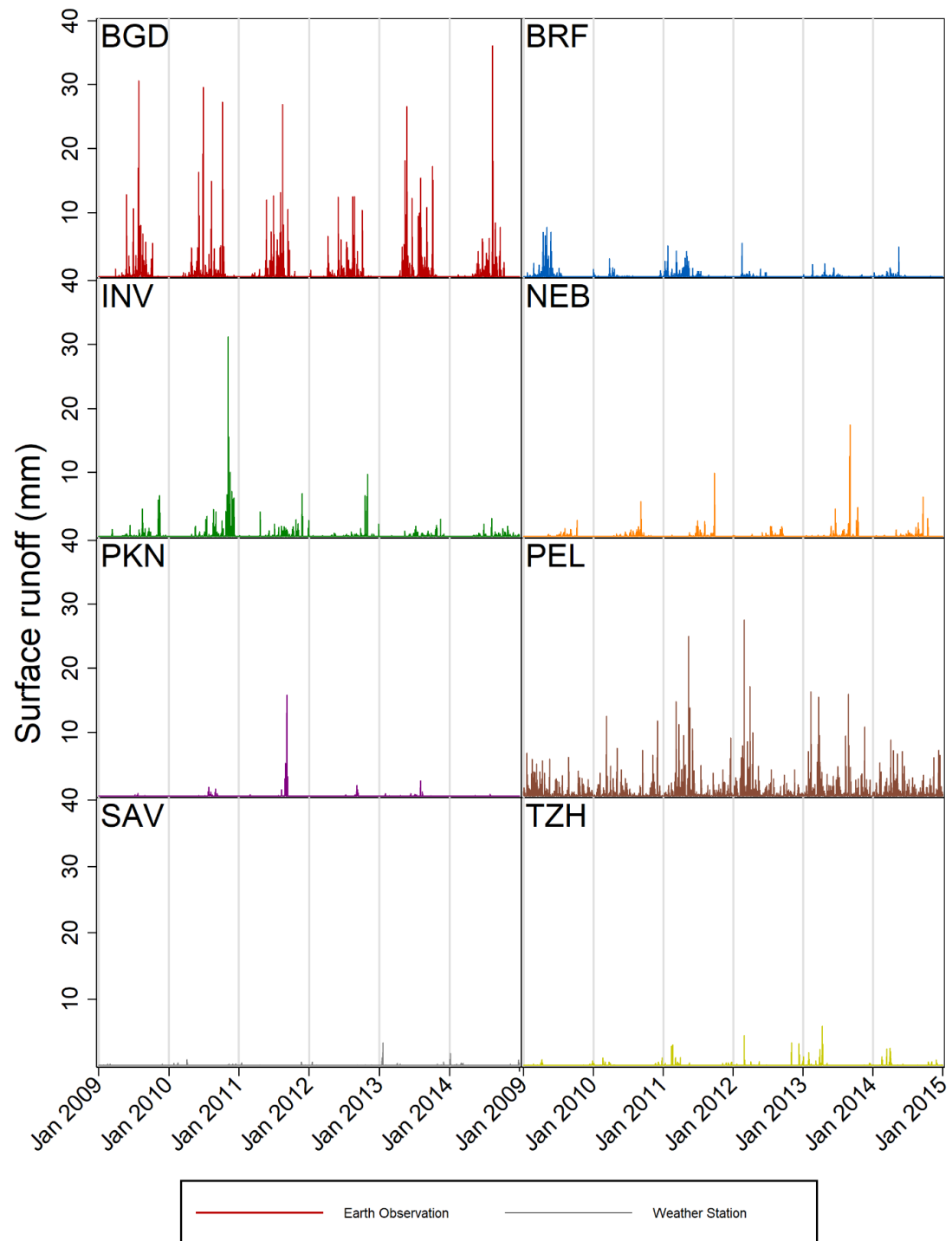
Figure 13: Daily surface runoff (mm) estimates by MAL-ED site, 2009 – 2014

Table 11 summarizes the evaluation metrics for both the daily estimates and the 7-day averages of all variables and sites for which both weather station and EO-derived data were available, and Figures 14a and b shows scatter plots of the daily variable values from the two sources plotted against each other, while figures 15a and b shows the same for the values aggregated to 7-day means. Table 12 reports the results of the same evaluation metric calculations when the analysis was restricted only to the months of the year during which rotavirus has been found to be highest.[58] Table 13 presents the odds ratios for the associations with rotavirus of each hydrometeorological variable calculated from logistic models fitted with GLM to the pooled (all-site) MAL-ED data over a 3-day lag adjusting for covariates. The columns indicate the source (weather station, EO or combined) and period of aggregation (daily or 7-day mean) of the meteorological predictor.

Table 10: Evaluation statistics for hydrometeorological variables in the eight MAL-ED sites[x]

	Variable	BGD Daily	BGD 7-day average	BRF Daily	BRF 7-day average	INV Daily	INV 7-day average	NEB Daily	NEB 7-day average	PKN Daily	PKN 7-day average	PEL Daily	PEL 7-day average	SAV Daily	SAV 7-day average	TZH Daily	TZH 7-day average
Median temperature (C)	n	1,307		2,186		1,851		2,138		2,170		2,190		2,106		-	
	R	0.93	0.95	0.58	0.68	0.93	0.97	0.93	0.95	0.98	0.99	0.45	0.59	0.88	0.95	-	-
	NSE	0.85	0.89	0.05	0.24	0.76	0.84	0.55	0.58	0.79	0.80	-1.22	-2.28	0.29	0.24	-	-
	MBE	0.27	0.28	0.13	0.13	0.86	0.86	2.89	2.89	-3.02	-3.02	1.11	1.11	2.63	2.63	-	-
	RMSE	1.54	1.27	0.88	0.65	1.58	1.25	3.51	3.32	3.37	3.22	1.82	1.43	3.23	2.83	-	-
	TPR	0.28	0.27	0.51	0.59	0.51	0.54	0.52	0.55	0.78	0.83	0.43	0.54	0.56	0.59	-	-
	FPR	0.12	0.12	0.12	0.10	0.05	0.04	0.12	0.10	0.05	0.04	0.14	0.12	0.09	0.08	-	-
Precipitation (mm) - GLDAS	n	1,306		2,168		1,857		1,829		2,167		1,991		1,933		798	
	R	0.50	0.68	0.50	0.78	0.24	0.64	0.27	0.61	0.18	0.58	0.34	0.42	0.39	0.67	0.22	0.67
	NSE	0.01	0.20	0.18	0.58	-0.30	0.22	-0.10	0.35	-0.72	-0.61	-0.24	-0.22	0.06	0.44	-0.84	-0.10
	MBE	-1.31	-1.44	-0.13	-0.10	-0.58	-0.57	0.12	0.03	-0.40	-0.42	-2.17	-2.10	0.28	0.27	-0.68	-0.67
	RMSE	12.06	6.13	7.64	2.87	9.71	3.46	9.23	4.12	5.61	2.28	14.07	5.71	7.58	2.86	7.03	2.44
	TPR	0.27	0.28	0.56	0.70	0.33	0.38	0.43	0.47	0.46	0.71	0.32	0.32	0.39	0.43	0.19	0.19
	FPR	0.12	0.11	0.12	0.07	0.13	0.11	0.09	0.06	0.11	0.12	0.15	0.16	0.12	0.10	0.22	0.22
Precipitation (mm) - CHIRPS	n	1,306		2,168		1,857		1,830		2,167		1,991		1,933		798	
	R	0.52	0.74	0.51	0.85	0.27	0.67	0.28	0.69	0.20	0.52	0.35	0.51	0.46	0.73	0.24	0.67
	NSE	0.17	0.50	-0.33	0.49	-0.15	0.36	-0.68	0.27	0.02	0.27	-0.25	0.09	-0.05	0.43	-1.41	-1.07
	MBE	-0.74	-0.90	-0.75	-0.71	-0.41	-0.41	-0.93	-1.07	0.10	0.10	-1.05	-0.94	-0.26	-0.33	-1.69	-1.68
	RMSE	11.07	4.86	9.69	3.17	9.12	3.14	11.42	4.37	4.25	1.54	14.11	4.92	7.98	2.91	8.04	3.34
	TPR	0.27	0.28	0.47	0.81	0.35	0.38	0.38	0.50	0.40	0.68	0.33	0.32	0.30	0.45	0.14	0.19
	FPR	0.12	0.11	0.09	0.04	0.13	0.11	0.11	0.05	0.06	0.12	0.15	0.15	0.07	0.09	0.21	0.23

[x] n = number of observations; R = correlation coefficient; NSE = Nash-Sutcliffe efficiency coefficient; MBE = mean bias error; RMSE = Root mean square error; TPR/FPR = True/false positive rate (for days exceeding the 80th percentile); BGD = Dhaka, Bangladesh; BRF = Fortaleza, Brazil; INV = Vellore, India; PKN = Naushero Feroze, Pakistan; PEL = Loreto, Peru; SAV = Venda, South Africa; TZH = Haydom, Tanzania.

Table 10: Evaluation statistics for hydrometeorological variables in the eight MAL-ED sites[x]

Variable	BGD Daily	BGD 7-day average	BRF Daily	BRF 7-day average	INV Daily	INV 7-day average	NEB Daily	NEB 7-day average	PKN Daily	PKN 7-day average	PEL Daily	PEL 7-day average	SAV Daily	SAV 7-day average	TZH Daily	TZH 7-day average
Surface pressure (Pa)																
n	1,295		1,534		-		25		21		2,059		2,080		-	
R	1.00	0.99	0.92	0.94	-	-	0.95	0.87	1.00	0.96	0.95	0.97	0.99	0.99	-	-
NSE	0.96	0.96	-52.71	-67.65	-	-	-62.80	-70.15	0.99	0.89	-0.47	-1.02	-56.58	-94.83	-	-
MBE	-0.89	-0.90	12.20	12.20	-	-	26.77	26.88	-0.59	-1.33	2.53	2.53	35.87	35.87	-	-
RMSE	1.02	1.07	12.22	12.21	-	-	26.79	26.93	0.99	2.75	2.62	2.57	35.88	35.87	-	-
TPR	0.38	0.37	0.42	0.42	-	-	0.20	0.20	0.20	0.20	0.68	0.71	0.78	0.78	-	-
FPR	0.01	0.01	0.03	0.03	-	-	0.00	0.00	0.00	0.06	0.04	0.03	0.02	0.02	-	-
Wind speed (m/s)																
n	1,300		2,186		-		2,138		2,170		2,072		2,106		-	
R	0.56	0.60	0.89	0.93	-	-	0.25	0.32	0.62	0.71	0.28	0.24	0.48	0.39	-	-
NSE	-7.19	-13.63	0.78	0.84	-	-	-1.32	-1.92	-8.77	-16.73	-0.16	-0.41	-0.49	-1.09	-	-
MBE	-1.89	-1.90	0.04	0.04	-	-	-0.46	-0.46	-1.84	-1.84	0.24	0.24	0.26	0.26	-	-
RMSE	2.08	2.01	0.75	0.58	-	-	0.92	0.76	2.11	2.01	0.58	0.44	0.87	0.52	-	-
TPR	0.26	0.27	0.69	0.72	-	-	0.31	0.36	0.48	0.54	0.32	0.33	0.44	0.35	-	-
FPR	0.10	0.13	0.06	0.06	-	-	0.16	0.14	0.13	0.11	0.15	0.14	0.12	0.15	-	-
Relative humidity (%)																
n	-		-		1,854		-		2,191		-		1,631		-	
R	-	-	-	-	0.79	0.86	-	-	0.68	0.75	-	-	0.87	0.85	-	-
NSE	-	-	-	-	0.40	0.47	-	-	-8.61	-10.48	-	-	0.56	0.37	-	-
MBE	-	-	-	-	4.62	4.61	-	-	33.23	33.24	-	-	-6.34	-6.37	-	-
RMSE	-	-	-	-	9.67	7.92	-	-	34.50	34.01	-	-	10.18	8.97	-	-
TPR	-	-	-	-	0.42	0.48	-	-	0.61	0.70	-	-	0.37	0.36	-	-
FPR	-	-	-	-	0.10	0.07	-	-	0.10	0.08	-	-	0.09	0.09	-	-

5.3.1. Temperature

With the exception of the Brazil and Peru sites, there was high correlation (R

> 0.85) between the daily temperatures measured at stations and those predicted by

the GLDAS model at all sites. While the equivalent correlation estimates were low

for Brazil and, especially, Peru, these metrics did improve when temperature was

aggregated to 7-day averages (as is the case for all sites). Correspondingly, the

Brazil and Peru sites had the lowest level of statistical agreement for temperature

according to the NSE statistics, with the negative values for this metric in Loreto,

Peru suggesting that the station-based average offers a better estimate. Pakistan

was the site with the largest absolute MBE value for temperature and the only one

in which the direction of the bias was negative indicating systematic

underestimation of the station-based temperature measure by the gridded

estimates. The South Africa site, which like the Pakistan site is situated some

distance from the weather-station (respectively 36.9km and 21.9km), also had high

values for MBE and RMSE, but the highest RMSE value for temperature (3.5°C) was

at the Nepal site. At all sites, 7-day temperature averages performed more favorably

than daily estimates according to the R, NSE and RMSE, and made only negligible

differences to the MBE, with the notable exception of NSE in Peru which

deteriorated substantially upon aggregation. Temperature estimates tended to

exhibit higher correlation and agreement than other variables but could be biased in

either direction by up to 3 degrees. The fact that the lowest correlation coefficients

tended to be in Brazil and Peru illustrates the limitation of relying solely on that

metric, since these are the two sites that, being closest to the equator, have the least

dominant seasonal temperature signal. The TPR for days in the upper quintile was very low and did not improve substantially following averaging over 7-days in all sites, but particularly in Dhaka, Bangladesh and with the exception of Naushero Feroze, Pakistan where around 80% of such extreme-temperature days according to the weather station data where characterized as such by GLDAS. The FPR was highest at the Peru site and lowest at the India site.

Correlation in temperature attenuated only slightly or not at all when data from the off-peak times of the year were excluded at all sites except for in Brazil, where the decrease in this metric was more pronounced, and for 7-day averages in Peru, for which it increased. Similarly, the effect on the NSE for temperature of restricting to rotavirus peak season was mostly slight except in Brazil and for 7-day averages in South Africa, where it changed qualitatively from a positive to a negative value. The MBE and RMSE for temperature increased only slightly in most cases, but in Brazil the direction of the bias changed, while the effect on the TPR and FPR was inconsistent across sites, improving most in Pakistan, deteriorating in Brazil and changing very little in Bangladesh.

A 1 degree increase in daily temperature was highly statistically significantly associated with a 7% increase in the odds of rotavirus detection when measured at the weather stations, an equally statistically significant 4% increase predicted by the EO estimates, and a 5% increase when the combined data were used. The results were similar when the same models were fitted using the 7-day average

95

temperatures, though with magnitude of the increase predicted by the station-based

estimates increasing to 8%
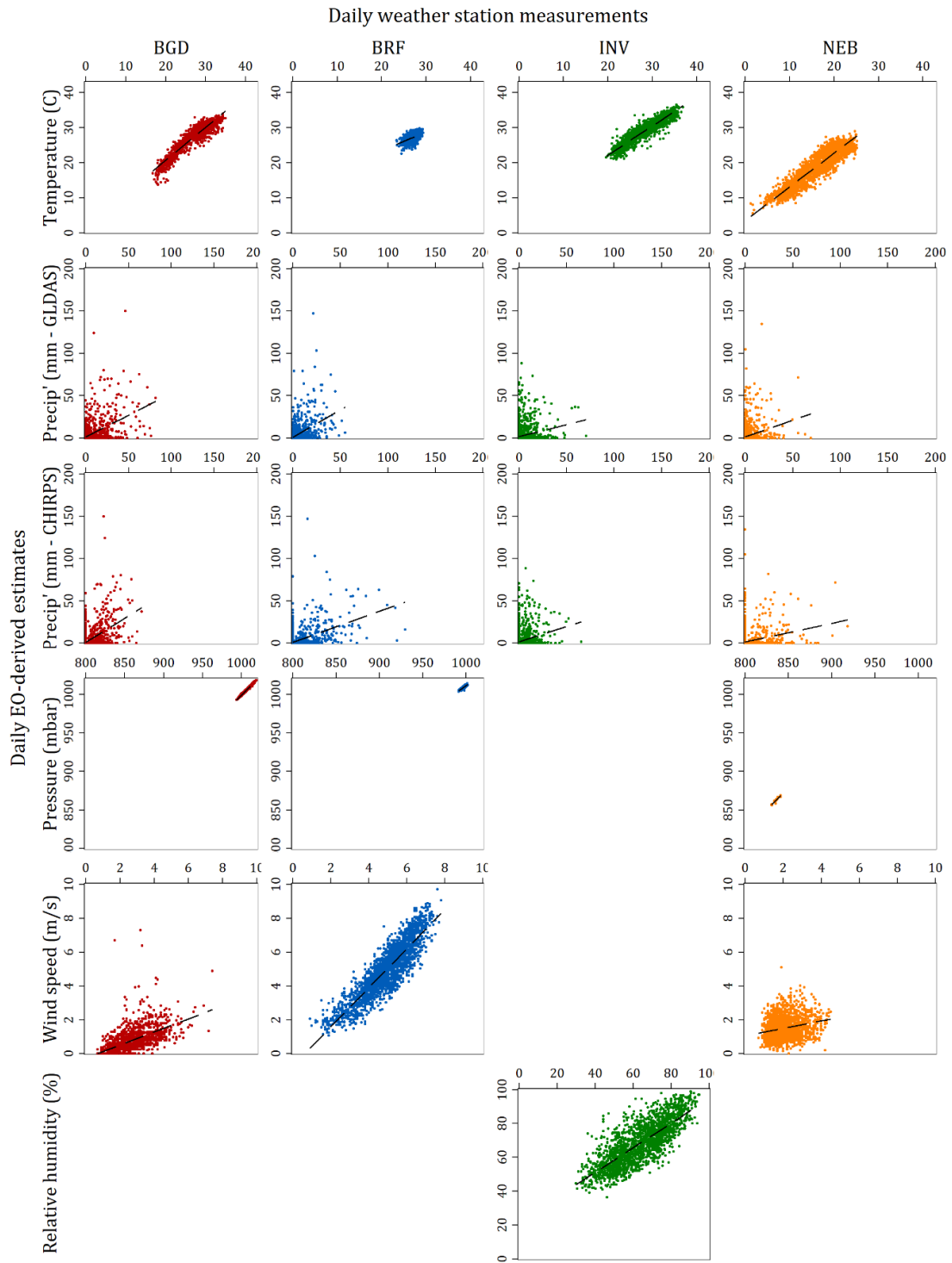
Figure 14a: Scatter plot matrix of EO-derived daily variable estimates against station-based equivalents
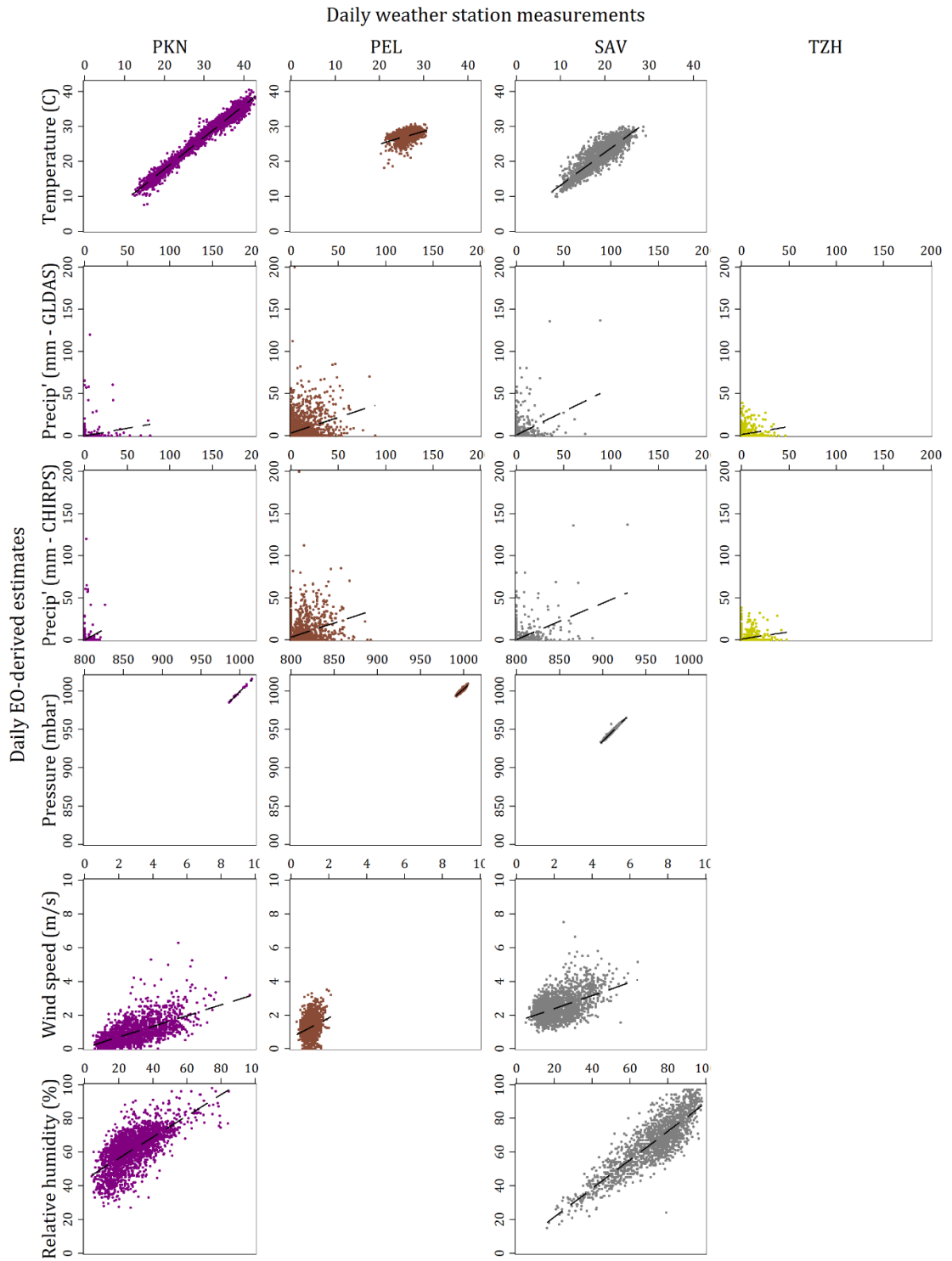
Figure 14b: Scatter plot matrix of EO-derived daily variable estimates against station-based equivalents
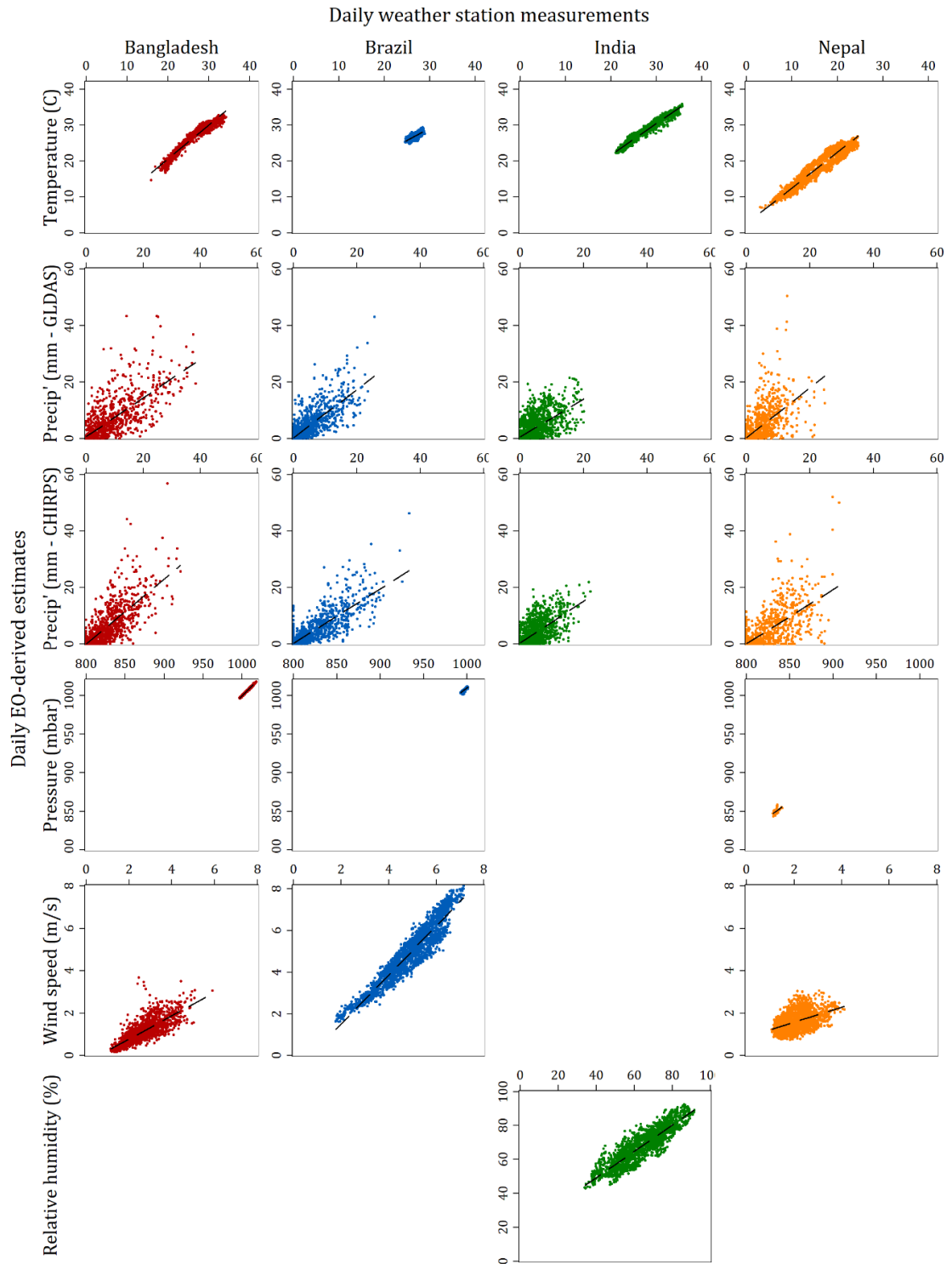
Figure 15a: Scatter plot matrix of EO-derived 7-day average variable estimates against station-based equivalents
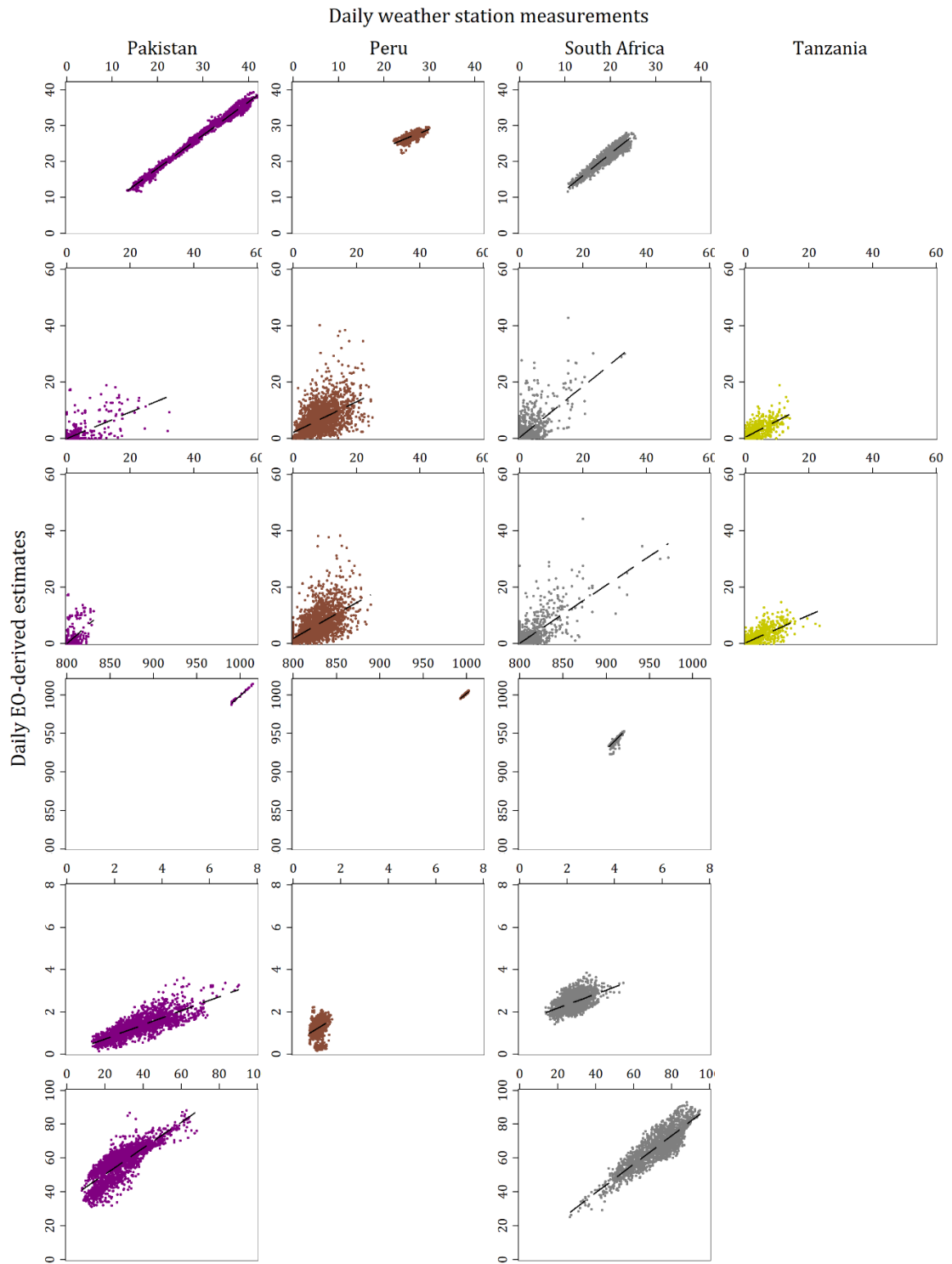
Figure 15b: Scatter plot matrix of EO-derived 7-day average variable estimates against station-based equivalents

5.3.2. Precipitation

The distribution of the precipitation parameter is highly skewed to the right with a high proportion of days with zero rainfall at all sites and for all data sources, but particularly in Pakistan and at the two African sites. Both the GLDAS and the CHIRPS precipitation estimates were poorly correlated (R < 0.55) with the station-based daily volumes at all sites. When the precipitation variables were aggregated into 7-day averages, the correlation improved at all sites with CHIRPS tending to outperform GLDAS, exceeding R = 0.6 in all but Peru and Pakistan. In several sites, the two products were biased in opposite directions, and in no site did one outperform the other across all evaluation statistics. Notably, there also seemed to be no clear relation between a site's distance from its weather-station and the performance of its precipitation estimate. Both GLDAS and CHIRPS had very low sensitivity (TPR) for classifying extreme precipitation days, although for both products, this metric improved greatly in Brazil and Pakistan for 7-day average precipitation. By far the highest FPRs were seen for the Tanzania site for both products.

For the precipitation variables, there was considerable variation in how the evaluation metrics changed when the analysis focused on the peak rotavirus season depending on site, period of aggregation and, to some extent, source (i.e. GLDAS compared to CHIRPS). At the Nepal, Peru and South Africa sites the differences were slight across all metrics for both sources, while in the other sites, but most markedly, in Tanzania, there was an apparent tendency for correlation and NSE to decrease substantially, while MBE and RMSE reduced slightly. The TPR tended to

decline for the peak rotavirus period compared with the full dataset, while the high

FPR for the Tanzania site decreased for the daily averages (most markedly with

CHIRPS), but increased slightly for the 7-day averages.

No statistically significant association was found between rotavirus positivity

and precipitation measured by weather stations either at a daily scale or averaged

over 7 days. By contrast, the GLDAS estimates predicted a highly statistically

significant 1% increase in the odds of a rotavirus positive stool for every millimeter

increase in daily precipitation, and a much larger and similarly significant effect size

when averaged over 7 days. Daily CHIRPS estimates showed no association with the

outcome, but 7-day average CHIRPS values were highly statistically significantly

associated with a 2% increase in the odds of rotavirus per millimeter. Substituting

missing precipitations data from stations with EO-estimates did not improve the

ability of the model to detect a statistically significant effect compared to the

station-only model when a daily resolution was considered, however, this combined

variable showed respectively a highly and a moderately statistically significant

association when averaged over seven days using GLDAS and CHIRPS.

Table 11: Evaluation statistics for hydrometeorological variables during peak rotavirus season in the eight MAL-ED sites[xi]

		BGD		BRF		INV		NEB		PKN		PEL		SAV		TZH	
Variable		Daily	7-day average	Daily	7-day average	Daily	7-day average	Daily	7-day average	Daily	7-day average	Daily	7-day average	Daily	7-day average	Daily	7-day average
Peak season		Nov - Feb		Sep - Nov		Dec - Mar		Dec - Feb		Nov - Feb		Mar - Jun		Feb - Jun		May - Jul	
Median temperature (C)	n	364		545		622		1,756		704		732		877		-	
	R	0.89	0.93	0.40	0.56	0.90	0.95	0.95	0.97	0.95	0.98	0.42	0.64	0.88	0.95	-	-
	NSE	0.77	0.82	-0.41	-0.22	0.56	0.58	0.55	0.58	0.51	0.57	-2.20	-5.83	0.01	-0.11	-	-
	MBE	0.18	0.28	-0.15	-0.15	1.21	1.20	3.30	3.30	-2.32	-2.31	1.48	1.48	3.05	3.05	-	-
	RMSE	1.41	1.12	0.71	0.50	1.63	1.37	3.77	3.60	2.63	2.41	1.97	1.62	3.47	3.19	-	-
	TPR	0.27	0.28	0.35	0.50	0.59	0.63	0.53	0.56	0.86	0.89	0.39	0.56	0.62	0.69	-	-
	FPR	0.09	0.07	0.14	0.12	0.03	0.01	0.10	0.09	0.04	0.02	0.15	0.11	0.08	0.06	-	-
Precipitation - GLDAS	n	364		544		626		1,487		704		665		817		111	
	R	0.50	0.52	0.10	0.16	0.15	0.65	0.27	0.60	-0.01	0.34	0.30	0.34	0.35	0.67	-0.04	-0.06
	NSE	0.24	0.23	-1.12	-1.29	-0.21	0.37	-0.10	0.35	-2.63	-2.74	-0.37	-0.40	-0.07	0.43	-18.12	-9.86
	MBE	0.06	0.06	-0.01	-0.04	0.21	0.16	0.24	0.16	-0.06	-0.06	-3.13	-2.92	0.07	0.09	-0.42	-0.43
	RMSE	2.47	1.55	1.20	0.57	5.91	1.99	10.04	4.49	1.00	0.36	16.27	6.61	5.53	2.18	3.43	1.48
	TPR	0.14	0.27	0.24	0.41	0.31	0.39	0.41	0.44	0.07	0.61	0.30	0.31	0.41	0.47	0.16	0.18
	FPR	0.11	0.10	0.12	0.14	0.15	0.12	0.09	0.07	0.06	0.14	0.16	0.16	0.12	0.09	0.16	0.27
Precipitation - CHIRPS	n	364		544		626		1,487		704		665		817		111	
	R	0.79	0.74	0.06	0.04	0.16	0.57	0.27	0.68	-0.01	0.12	0.29	0.45	0.39	0.72	0.03	-0.03
	NSE	0.55	0.37	-0.12	-0.25	-0.12	0.29	-0.71	0.26	-1.21	-1.33	-0.35	0.04	-0.38	0.44	-15.56	-7.90
	MBE	0.07	0.09	0.17	0.16	0.43	0.38	-1.00	-1.19	-0.01	-0.01	-1.29	-1.05	-0.16	-0.22	-0.42	-0.42
	RMSE	1.90	1.40	0.87	0.42	5.70	2.11	12.49	4.78	0.78	0.29	16.16	5.49	6.28	2.15	3.19	1.34
	TPR	0.05	0.23	0.06	0.21	0.10	0.32	0.38	0.46	0.07	0.39	0.31	0.28	0.25	0.48	0.04	0.18
	FPR	0.03	0.16	0.01	0.05	0.05	0.15	0.11	0.06	0.02	0.07	0.15	0.17	0.05	0.10	0.04	0.29

[xi] n = number of observations; R = correlation coefficient; NSE = Nash-Sutcliffe efficiency coefficient; MBE = mean bias error; RMSE = Root mean square error; TPR/FPR = True/false positive rate (for days exceeding the 80th percentile); BGD = Dhaka, Bangladesh; BRF = Fortaleza, Brazil; INV = Vellore, India; PKN = Naushero Feroze, Pakistan; PEL = Loreto, Peru; SAV = Venda, South Africa; TZH = Haydom, Tanzania.

Table 11: Evaluation statistics for hydrometeorological variables during peak rotavirus season in the eight MAL-ED sites[xi]

Variable	BGD Daily	BGD 7-day average	BRF Daily	BRF 7-day average	INV Daily	INV 7-day average	NEB Daily	NEB 7-day average	PKN Daily	PKN 7-day average	PEL Daily	PEL 7-day average	SAV Daily	SAV 7-day average	TZH Daily	TZH 7-day average
Peak season	Nov - Feb		Sep - Nov		Dec - Mar		Dec - Feb		Nov - Feb		Mar - Jun		Feb - Jun		May - Jul	
Surface pressure (Pa)																
n	362		405		-		24		4		669		869		-	
R	0.98	0.95	0.95	0.96	-	-	0.95	0.85	0.97	0.93	0.92	0.95	0.99	0.99	-	-
NSE	0.82	0.67	-56.01	-69.72	-	-	-65.65	-74.68	0.83	0.00	-1.12	-2.48	-67.35	-121.09	-	-
MBE	-0.90	-0.99	12.47	12.44	-	-	26.74	26.85	-1.03	-2.90	2.35	2.35	35.89	35.88	-	-
RMSE	1.02	1.19	12.48	12.45	-	-	26.76	26.90	1.31	3.15	2.45	2.39	35.90	35.88	-	-
TPR	0.32	0.30	0.45	0.43	-	-	0.20	0.20	0.20	0.20	0.59	0.61	0.81	0.82	-	-
FPR	0.03	0.05	0.03	0.04	-	-	0.00	0.00	0.25	0.25	0.06	0.05	0.03	0.02	-	-
Wind speed (m/s)																
n	363		545		-		1,756		704		715		877		-	
R	0.50	0.64	0.80	0.81	-	-	0.14	0.16	0.57	0.67	0.35	0.40	0.44	0.32	-	-
NSE	-13.41	-31.06	0.53	0.50	-	-	-1.53	-2.26	-9.08	-21.94	-0.35	-1.03	-0.55	-1.37	-	-
MBE	-1.68	-1.70	0.37	0.36	-	-	-0.43	-0.42	-1.28	-1.27	0.34	0.34	0.35	0.33	-	-
RMSE	1.79	1.74	0.84	0.72	-	-	0.89	0.72	1.42	1.31	0.58	0.45	0.86	0.55	-	-
TPR	0.25	0.28	0.61	0.79	-	-	0.25	0.32	0.56	0.66	0.39	0.33	0.43	0.40	-	-
FPR	0.08	0.08	0.09	0.05	-	-	0.16	0.17	0.10	0.08	0.14	0.15	0.11	0.14	-	-
Relative humidity (%)																
n	-		-		625		-		709		-		625		-	
R	-	-	-	-	0.79	0.83	-	-	0.54	0.66	-	-	0.79	0.72	-	-
NSE	-	-	-	-	0.49	0.55	-	-	-25.69	-45.55	-	-	0.18	-0.24	-	-
MBE	-	-	-	-	2.38	2.42	-	-	37.64	37.72	-	-	-8.49	-8.60	-	-
RMSE	-	-	-	-	7.60	6.16	-	-	38.66	38.17	-	-	11.77	11.05	-	-
TPR	-	-	-	-	0.41	0.44	-	-	0.46	0.58	-	-	0.33	0.34	-	-
FPR	-	-	-	-	0.11	0.10	-	-	0.14	0.11	-	-	0.09	0.09	-	-

5.3.3. Surface pressure

GLDAS surface pressure estimates were highly correlated with station-based estimates (R > 0.85), even at those sites with very few such observations, namely Nepal (n = 25) and Pakistan (n = 21). Performance according to the other metrics was much more varied, with the Brazil, South Africa and Nepal sites showing a very clear systematic bias towards over-estimation of the station-based measures by the gridded estimates and poor statistical agreement between the two sources according to the NSE. The high RMSE at these sites were in part due to the fact that this statistic is expressed in the same units as the variable itself, in this case millibars with values at a higher order of magnitude than the other variables. The TPR tended to be low for both daily and 7-day surface pressure estimates, particularly at the sites with very few observations, but with the exception of Peru and, especially, South Africa. The FPR for surface pressure was low at all sites compared with other parameters. In general, the evaluation statistics for surface pressure deteriorated when only the peak season was considered, but not substantially. Only a very small number of pressure observations from the Pakistan site occurred in the peak rotavirus season. For both daily and 7-day average estimates, a one millibar increase in surface pressure above 1,000 mbar was associated with a highly statistically significant 1% increase in the odds of rotavirus detection when weather station records were used, and a moderately statistically significant <0.5% increase when either GLDAS or combined estimates were used.

5.3.4. Wind speed

The wind speed estimated by GLDAS tended to show poor correlation with those recorded by the weather stations and exhibit considerable biases and poor statistical agreement. An exception to this was the Brazil site, where the EO-derived wind speed estimate was notable for showing exceptionally high correlation and agreement and minimal bias. The TPR for wind speed estimated by GLDAS was low and the FPR high relative to other parameters, again with the exception of the Brazil site. A highly statistically significant inverse association between weather station measurements of wind speed and the outcome was observed with a 1 m/s increase in daily wind speed predicting a 22% decline and 7-day average wind speed a 27% decline in the odds of rotavirus infection. When GLDAS daily estimates were used instead, the direction of the association changed but was no longer statistically significant and when combined values were used, no statistically significant effect was discernible.

Table 12: Odds ratios (with 95% confidence intervals) for associations between hydrometeorological variables and rotavirus infection adjusting for age, seasonality and calendar time predicted by logistic model fitted with GLM[xii]

		Daily			7-day average		
		Station	**EO-derived**	**Combined[xiii]**	**Station**	**EO-derived**	**Combined[xiv]**
Temperature (C)		1.06*** (1.03, 1.09)	1.02* (1.00, 1.04)	1.02 (1.00, 1.04)	1.05*** (1.02, 1.09)	1.02 (1.00, 1.03)	1.01 (0.99, 1.03)
Precipitation (mm)	**GLDAS**	1.00 (0.99, 1.01)	1.02*** (1.01, 1.03)	1.00 (1.00, 1.01)	1.02* (1.01, 1.04)	1.04*** (1.03, 1.06)	1.03*** (1.01, 1.04)
	CHIRPS		1.01*** (1.01, 1.02)	1.00 (1.00, 1.01)		1.04*** (1.02, 1.05)	1.02** (1.01, 1.04)
Surface pressure (mbar)		1.01*** (1.01, 1.02)	1.00** (1.00, 1.00)	1.00 (1.00, 1.00)	1.01*** (1.01, 1.02)	1.00** (1.00, 1.00)	1.00 (1.00, 1.00)
Wind speed (m/s)		0.76*** (0.69, 0.83)	0.92* (0.86, 0.99)	0.95 (0.89, 1.01)	0.70*** (0.62, 0.78)	0.87*** (0.81, 0.95)	0.91** (0.85, 0.97)
Relative humidity (%)		0.99 (0.98, 1.00)	1.01** (1.00, 1.01)	1.01* (1.00, 1.01)	0.98* (0.96, 1.00)	1.01** (1.00, 1.01)	1.01** (1.00, 1.02)
Specific humidity (g/kg)[xiv]		-	1.06*** (1.04, 1.09)	-	-	1.07*** (1.04, 1.09)	-
Solar radiation (W/m²)		-	1.00** (1.00, 1.00)	-	-	1.00*** (0.99, 1.00)	-
Soil moisture (%)		-	1.04*** (1.03, 1.05)	-	-	1.04*** (1.03, 1.05)	-
Surface runoff (mm)		-	1.08*** (1.05, 1.12)	-	-	1.30*** (1.18, 1.43)	-

[xii] *** p <0.001, ** p = 0.001 – 0.01, * p = 0.01 - 0.05. Temperature variables were centered at 25C; surface pressure at 1,000 mbar; wind speed at 2 m/s; relative humidity at 40%; specific humidity at 15 g/kg; solar radiation at 200 W/m²; and soil moisture at 25%.

[xiii] In the "combined" data, values that were missing in the station data were replaced by their equivalent EO-derived estimates.

[xiv] Specific humidity was converted to grams per kilogram (multiplied by 1,000) so that the coefficient the change for a plausible one-unit increment.

5.3.5. Relative humidity

Relative humidity showed moderate correlation between the GLDAS and station-based data in South Africa and, at least considering the 7-day average, in India, showing moderate statistical agreement by the NSE, but with notable error and bias in opposite directions in the two sites. In Pakistan, the equivalent estimates were only moderately correlated and showed low agreement and high bias and error. It should be noted in interpreting these statistics, that the station-based estimates for Pakistan and India were calculated as the average of two daily measurements taken at time points representing extremes of the daily cycle of humidity, while the GLDAS indicator was an average of 3-hourly estimates within a day, a fact which may explain some of the bias and error seen at these sites. Correlation between the EO-derived and weather station estimates of relative humidity from all three sites either decreased slightly or did not change when only the peak rotavirus season was considered. At the India site NSE increased slightly while MBE and RMSE decreased while in the other two sites with relative humidity data, the opposite was the case. Weather station records of relative humidity were moderately significantly associated with rotavirus at a daily resolution and only slightly so when aggregated to 7-day means, however the EO-derived estimates were not statistically significant at either levels and the combined data was only slightly statically significantly associated with the outcome when 7-day averages were considered.

5.3.6. Other parameters

Of the four GLDAS variables for which no station-based equivalents were available (specific humidity, solar radiation, soil moisture and surface runoff), all showed a highly statistically significantly direct association with rotavirus infection for both daily and 7-day average estimates with the exception of solar radiation, for which no association was seen using daily estimates, and surface runoff, for which the association with daily estimates was only slightly statistically significant.

5.4. Discussion

The increased availability of historic meteorological data offers great potential to environmental epidemiologists that has yet to be fully explored. While weather stations may record a small number of parameters at particular strategic locations to varying degrees of accuracy, EO-derived products aim to provide meteorological estimates where direct measurements do not exist and therefore merit assessment as potential surrogates. Although such data are starting to be used in studies of human health, livelihood and vulnerability[20,76,108,192], as yet there have been no systematic attempts to evaluate the relative validity and utility of hydrometeorological data from different sources for modeling health outcomes. This study represents an initial attempt to do this and the results indicate there may be certain pitfalls to straightforwardly substituting ground-based observations for their EO-derived equivalents and that researchers should be cautious about the unreflective reliance on these without proper consideration of their limitations.

According to the evaluation statistics the performance of the two gridded EO

data products assessed here was highly dependent on the location, the variable, the

evaluation metric and the distance from the study site (the location at which the

data were extracted from the grid) to the weather station at which the *in situ* data

were recorded. Furthermore, several variables differed considerably in their ability

to statistically significantly predict rotavirus infection depending on whether the

station-based or EO-derived data were used and, when the latter was used to fill

gaps in the former, it often led to a considerable attenuation of the significance level.

Temperature estimates from GLDAS were one of the best performing variables

according to the evaluation statistics, and showed a highly statistically significant

association with the rotavirus outcome, both when GLDAS estimates and, despite

their incompleteness, station measurements were used. Estimates of precipitation

performed most favorably according to the evaluation statistics when they were

extracted from CHIRPS rather than GLDAS and were aggregated to 7-day averages.

In absolute terms, however, precipitation was one of the variables for which EO-

derived data performed the worst. This is perhaps unsurprising, given that it is a

challenging variable both to measure remotely and to model, since variation in

rainfall can be so localized as to confound simple grid-to-station comparisons,

especially at the comparatively low resolution of CHIRPS and GLDAS. In spite of this,

GLDAS-derived precipitation estimates showed strongly statistically significant

associations with rotavirus where gauge-based estimates showed none, though for

CHIRPS, an association was only found using the 7-day averages. This may be

because the model-derived estimates are not in fact reflective of rainfall per se, but

are some aggregate of closely related factors like humidity, cloud and wind which

correlate with precipitation at large scale, but differ from true precipitation in subtle

ways that collectively make them a stronger driver of rotavirus transmission. It is

also conceivable that EO actually provide more meaningful rainfall estimates than

station data in some cases, due to station equipment malfunction or siting bias. This

is difficult to evaluate with available data.

Daily surface pressure from GLDAS was the variable that showed the highest

level of correlation with the station-based measures both for the full annual cycle

and for the peak-season, including in Brazil and Peru which, as with temperature,

had the smallest seasonal variation. The biases observed for this variable are

consistent with differences in altitude between the sites and their respective

weather stations. Confirming the findings of Hervas and colleagues[19], pressure was

statistically significantly associated with rotavirus. This was the case whether EO

data, station data, or the two in combination were used. That wind speed mostly

performed poorly is largely to be expected since most weather stations only report

winds at a 2m height on a very localized scale, while GLDAS produces broader scale

estimates of 10m height winds. This may also explain why station-based wind

measurement were strongly associated with rotavirus, while GLDAS-derived

estimates showed no association. Wind speeds at 2m are more likely to facilitate the

transmission of the virus than at 10m. That the station-based measurements of

surface pressure and wind speed are so highly statistically significant in spite of

their incompleteness is suggestive of a strong and hitherto underexplored

association [19,21]. In an analysis that had used only EO-derived estimates of wind

speed, this association would have gone unidentified, illustrating that a poorly informed choice of meteorological data can be a potential source of type II error.

In line with previously documented evidence [21,59,60], an association was found between relative humidity and rotavirus over the 3-day lag used here, however, it was one of the weaker associations identified and only apparent when station measurements, not GLDAS or combined, estimates were used. It should be noted that the choice to examine the association over a 3-day lag, though guided by biology, was to some extent arbitrary. Further exploration of the exposure-lag-response structure may reveal a stronger association operating over longer time windows but is beyond the scope of this paper.[193] It is notable that the two measures of humidity differed substantially in their association with rotavirus, indicating a highly statistically significant association with specific humidity but none for relative humidity. This demonstrates the importance of considering the physical meaning of related but distinct variables: specific humidity is highly sensitive to air temperature, and thus reflects a combination of temperature and humidity conditions, where relative humidity is standardized to temperature and represents degree of saturation.

With a few notable site-specific exceptions, the EO data performed very poorly in detecting extremes in the weather station data, which here we defined as sensitivity in classifying days in which a given parameter exceeded the 80th percentile of its overall distribution. Researchers wishing to assess the impact of extreme weather events on health outcomes, are encouraged to explore multiple

cutoffs and definitions, as well as lower extremes, which were beyond the scope of this paper, but may be particularly relevant for parameters such as wind speed given its inverse association with rotavirus identified here. In this preliminary analysis we assumed linearity in all associations between hydrometeorological parameters and rotavirus. Future analyses should include methods capable of taking into account non-linear relationships such as polynomial transformations or natural cubic splines.

That the global models and EO data used to generate the variables included here do not perform perfectly should not be surprising. Neither GLDAS nor CHIRPS purport to be entirely representative at local scale or daily resolution, however they do offer the advantage that findings can be generalized to other locations and results mapped continuously across the landscape. While there is no *a priori* reason to suppose that one EO product is better than another, GLDAS and CHIRPS were chosen for this analysis because they are two products for which promising validation work has been published. There are numerous initiatives underway to evaluate these and similar datasets in a more robust way across multiple locations, however, to date, most validation efforts have been piecemeal and the reality remains that in most locations, like those of the MAL-ED study sites, the data remain unevaluated. Spot-check comparisons like the one reported in this study often yield conflicting or inconclusive conclusions, since the station-based data do not always represent a gold standard of comparison for estimates extracted from gridded products at precise coordinates, especially for parameters like precipitation that vary on such local scales. Different weather stations may use different equipment to

measure the same parameter and, because of this and other factors, may vary widely in their accuracy in characterizing conditions at their own locations and in the extent to which such characterizations can be extrapolated to nearby population settlements. When, as is the case with this study, there is disagreement between the two data sources, it is near impossible to attribute this to specific sources of error such as deficiencies with the model used to derive the EO estimate, distance between the study site and its nearest weather station or incompleteness or inaccuracies in the station records. Poor performance of data from one source relative to the other is problematic insofar as it impedes the ability of a study to detect an association and may be tolerated to the extent that it is still possible to detect and quantify their effect on outcomes that there is *a priori* reason to believe are climate-sensitive.

As weather stations become more affordable, accurate, easy to install and offer a wider suite of measurements, environmental epidemiologists working in remote and underserved field sites should consider installing these instruments themselves. Otherwise, secondary station-based observation data should be given preference when they are complete and measured at a location that is close to the study site, and exhaustive attempts should be made to coordinate with local agencies that might be able to provide such data when they are not publicly accessible. When, as is often the case, high quality observational data are not available, EO-derived products may be introduced in a number of ways: to fill gaps, either by direct substitution, or as covariates for multiple imputation of missing data; as surrogates for variables that are not commonly measured at weather

stations (e.g. soil moisture, surface runoff etc.); to generate data ensembles by averaging over multiple EO-derived data sets (the "wisdom of crowds" approach); to set uncertainty bounds when applying data to health risk assessment.

Researchers may feel justified in using gridded products as surrogates to the extent that they are the best hydrometeorological monitoring tools available at global scale and daily resolution. Alternatively, they may use observed data to calibrate and adjust the gridded estimates if they have a level of confidence that the station-based records truly represent the historical conditions at their sites, or attempt custom corrections based on characteristics of the study area, where these are known to high degree of certainty. Where observational data is available but incomplete, studies should report associations with the observed as well as EO-derived data as a sensitivity analysis. The nature of the research question will, in some respects, determine the relative importance of the different evaluation metrics. If the absolute values of the hydrometeorological variable are of interest, minimizing bias will be a priority, whereas if climate anomalies relative to the normal range are the predictor of interest, then more bias may be tolerated. Several of the commonly used evaluation metrics may be sensitive to the averaging period and, as demonstrated here, to the seasonal cycle. The performance of peak-season data would be most important when developing predictive models intended to predict more than just the seasonality of a disease process. However, what is significant from the point of view of the data's epidemiological application is how sensitive the analysis is to moderate inaccuracies in the weather data.

6. Chapter 6: Using hydrometeorological variables to model and predict rotavirus infection episodes

6.1. Background

Diarrheal disease is the fourth leading cause of under-5 mortality, accounting for 8.8% (526,000) of all such deaths globally in 2015 with rotavirus thought to be the pathogen responsible for the largest share of the vaccine-preventable EID burden.[6,171] Climate is a major determinant in the epidemiology of infectious diseases including EIDs since environmental and meteorological factors determine the survival and dispersal of transmissible microorganisms and the distribution of their hosts and reservoirs.[32,41] Indeed, changes in the distribution and burden of EIDs are one of the principal manifestations by which climate change is predicted to impact human health.[106] Under a likely greenhouse gas emissions scenario, it has been estimated that the year 2030 will see an additional 50,000 diarrheal disease deaths attributable to climate change, threatening to undermine recent progress in the reduction of child deaths by this cause.[106] However, predictions like these have wide ranges of uncertainty due to a lack of evidence of how individual environmental factors impact transmission of specific enteric pathogens.[106,107]

Recognizing the complexity of the ecological systems within which weather influences EIDs, there have been calls for a research agenda to characterize the differential associations between multiple meteorological exposures and individual EID agents.[19,105] Numerous time-series analyses and systematic reviews have explored associations between meteorological exposures and either general diarrheal disease or rotavirus-specific outcomes at particular locations, with several

reporting statistically significant relationships with temperature,[20,21,60,131–133,194] relative humidity,[21,59,60,194] and precipitation.[20,21,133] Associations with river levels,[59] NDVI,[20] atmospheric pressure, wind speed and solar radiation[19] have also been documented. Hypothesized explanations for these findings emphasize several possible pathways:

- **Survivability of the virus outside the host:** Laboratory findings indicate that temperature, humidity and soil moisture content impact the duration of survival of rotavirus when suspended in airborne aerosols or attached to surfaces and fomites.[135,141,163] Conversely, UV radiation may have an inactivating effect on rotavirus similar to that demonstrated for feline calicivirus.[60,164]

- **Dispersion of the virus through the environment:** Rainfall runoff and flooding are means by which agents like rotavirus may disseminate, collect and suspend in surface or ground water or be rinsed from the soil[32,105] while wind may promote, and high pressure inhibit, their areal transport in aerosols and dust particles.[19,21]

- **Host factors and behaviors:** Human behavior in response to certain weather conditions may mediate or interact with climatic exposures. For example rainfall may alter contact rates between infected and susceptible individuals as they congregate indoors, migratory patterns may be influenced by harvest seasons, and seasonal changes in diet or water sources may alter exposure to waterborne and foodborne pathogens.[44]

Significant questions remain about the relative importance of these pathways. Most previous studies have had to rely on a small subset of the most easily measured meteorological variables with their values aggregated to weekly or monthly averages. The integrity of the outcomes in these analyses are often limited by short surveillance periods (often of a year or less) or unstandardized diagnostic testing of a few thousand samples of a few hundred subjects at individual sites. There is a need to evaluate the combined impact of multiple meteorological exposures to pathogen-specific EID outcomes at a level of spatiotemporal disaggregation sufficient to characterize potential lag effects, interactions and non-linearity.[19,105] EO climate data products like GLDAS, which are derived from satellites and model-based reanalysis offer an opportunity to address this gap.[195] The aim of this analysis was to characterize the associations between a suite of nine GLDAS hydrometeorological variables and rotavirus infection status ascertained through community-based surveillance and use the resulting estimates to predict future trends in disease burden.

6.2. Statistical methods

This analysis aimed to model the rotavirus isolation rate over time by approximating the probability of a stool sample being positive for the virus on a given day calculated from odds ratios from fitting logistic models to the binary infection status outcome.[58,133] Three separate effects of the hydrometeorological exposures on the rotavirus outcome were approximated:

1. ***The absolute effect:*** This analysis made no adjustment for seasonality and therefore incorporated the effect of medium-term, intra-annual variability due to shifts in both climate and rotavirus incidence over the yearly cycle. As such, it does not account for the potential confounding effects of latent co-seasonal behavioral or host factors, which may be associated with both the main exposures and the outcome. Relative humidity was centered at 40%, soil moisture at 25%, solar radiation at 200 W/m^2, specific humidity at 0.015 kg/kg, temperature at 25°C, and wind speed at 2 m/s so that when combined in a multi-variable model, the intercept for these variables could be interpreted as a value close to its pooled (as opposed to its site-specific) median.

2. ***The standardized effect:*** In this analysis, the hydrometeorological variables were standardized to their local distributions by recalculating each one as the deviation from its site-specific median value. The purpose of this was to adjust for between-site variation in the distribution of the parameters by scaling each one to its variability at a particular site and thereby remove the potential confounding effect of sites that may be outliers with respect to both the exposures and the outcome. Furthermore, it was reasoned that shifts in climate parameters relative to their typical values at a given location, rather than their absolute values, may be more important for transmission. The median was used instead of the mean since several variables were highly right-skewed and the deviations were expressed in the original units of each variable rather than as Z-scores to retain interpretability.

3. ***The seasonality-adjusted standardized effect:*** This analysis also used the site-specific deviations but, in addition, adjusted for co-seasonality of rotavirus incidence (which has been demonstrated to have two annual peaks at most of the MAL-ED sites[58]) and meteorological conditions by including annual and biannual Fourier-series sine and cosine functions as terms in the model. This reduced the influence of intra-annual variation and confounding, while retaining the effect of short-term, interdiurnal variability. Since the timing of seasonal rotavirus patterns vary between the different site locations, only the terms for the interaction between indicator variables for the eight sites and the annual and biannual harmonics were included, with the terms for the main effect omitted.[196]

Several potential confounders were included as covariates in all models. A binary variable was included that grouped the sites according to whether they are located in a country that had introduced the rotavirus vaccine at the time of data collection (Brazil, Peru and South Africa) or not (Bangladesh, India, Nepal, Pakistan and Tanzania – the reference category). This was to adjust for the reduction in background transmission levels brought about by the introduction of the vaccine. The infant's age in continuous months at the time of the stool sample was centered at 12 months of age and represented in the model using both linear and quadratic terms with vaccine category-specific interaction terms to account for differences in the relationship between age and rotavirus transmission in the two groups of sites (which were identified through exploratory analysis). Finally, a second binary variable was included representing whether the stool sample was collected during a monthly assessment or a diarrheal episode. This was to adjust for the differential

probability that a sample obtained by active surveillance (monthly) would be positive compared to one obtained by passive surveillance (diarrheal) and allowed for the separate approximation of the effect on symptomatic relative to asymptomatic rotavirus episodes. Hydrometeorological variables that showed non-linear associations with rotavirus were represented in the models by restricted cubic spline terms with degrees of freedom determined by comparing AIC statistic from models using 3 – 5 knots positioned at the corresponding percentiles of the variable distribution.

In the initial stage of the analysis, it was necessary to determine the optimal lag length at which the association of these highly mutually correlated and autocorrelated exposure variables with rotavirus was strongest for each of the three effects. Very little literature has been published on this issue, however one study into the associations between weather exposures and weekly rotavirus gastroenteritis hospital admissions defined the lower limit of the lag periods considered in terms of the incubation period plus a 48 hour interval from the onset of symptoms to admission and found the strongest association using a one-week lag.[19] Because rotavirus has a very short incubation period (estimated at 1.4 – 2.4 days by one study[191]) and serial interval (the longest documented being 9 days[131]) and because the MAL-ED project used active, community-based surveillance, thus reducing the interval between symptom onset and outcome ascertainment, this analysis considered lag lengths ranging from 2 – 10 days. For each variable, models were fitted separately for each of these lag lengths in turn ("individual lag models") and the results were compared with those of distributed lag models (DLMs), which

incorporated all terms for all lag lengths of 2 to 10 days. For each variable and effect, lag lengths were selected for inclusion in further analyses if the Wald test statistic for the association with rotavirus was statistically significant at the α =0.05 level in both the individual lag models and the DLM. If no lag length met this criterion, then the single length was selected for which the average of the two p-values was lowest.

The terms for the selected lag lengths for each variable in turn were included in separate, single-variable models (along with the non-hydrometeorological covariates) for each effect to determine their independent associations with rotavirus before adjusting for the others. Terms for the interactions between these main exposures and the stool specimen type (diarrheal/monthly) were included if statistically significant at the α =0.05 level. This was to account for the possibility that, depending on the route of transmission through which they operate, different variables may differentially affect the probability of inducing symptomatic compared to asymptomatic infections due to the infectious dose they are capable of transmitting. To visualize the single-variable associations, the probabilities predicted by these models were plotted across the range of values of each variable for each effect and for monthly and diarrheal samples separately. Then backward stepwise selection was used to find the subset of selected variables that retained a likelihood-ratio test statistic that was significant at the α =0.05 level in the presence of the others for each effect. Specific humidity was excluded from these and subsequent multi-variable models due to it being almost entirely a function of two other included variables (temperature and relative humidity), and only the results of its independent association are reported here. To arrive at a final model for each

effect, terms for the interactions of hydrometeorological variables in the subset with each other and with vaccine category and specimen type were tested for inclusion and retained if significant at the α <=0.01 level when added by forward hierarchical selection in order of their separate statistical significance. Interactions between variables that were represented by multiple spline terms were assessed by including multiplicative terms for all pairwise combinations of the spline terms. The purpose of including model terms for interactions for which there was no *a priori* reason to suppose effect measure modification, was both to inform hypothesis generation regarding which variables might lie on common pathways and to improve the predictive ability of and the proportion of variance explained by the final models. Variables that lost significance in the presence of retained interaction terms for combinations of other variables were excluded from the final models.

The final model for the absolute effect was used to predict the average expected probability (with 95% confidence intervals) of symptomatic and asymptomatic rotavirus infection during the period of follow-up. This was calculated for each site and specimen type using a bootstrapping method that, on each of 2,000 iterations, drew a random sample with replacement of subjects in that cohort and, within each sampled subject, randomly sampled one single day of follow-up. The average of the mean and 2.5th and 97.5th percentile values of the probability of rotavirus infection episodes (both symptomatic and asymptomatic) predicted by the final model for the sampled days and subjects are reported as the estimate and confidence limits. To illustrate how this approach can be combined with projected long-term changes in the distribution of the hydrometeorological

variables due to climate change to arrive at projections of future disease burden,

these predictions were repeated for a basic simulated scenario in which the values

for each of the variables in the final model were all increased by 2%. Applying more

rigorous climate projections for the specific locations of the MAL-ED sites was

beyond the scope of this analysis but will be explored in subsequent publications.

All models were fitted using GLM with robust variance estimation to account

for within-subject clustering of the outcome. The combined significance of the terms

for each covariate – the spline and interaction terms for each meteorological

variable, the linear and quadratic terms for age and their interaction with vaccine

category and the site-specific interaction terms for the four Fourier series functions

- were each assessed using the Wald test. To quantify the proportion of the variance

explained by the final models and by the meteorological variables and their

interactions, full and partial coefficients of discrimination (COD - a measure of the

explanatory power of a logistic regression model that has a similar interpretation to

R^2) were calculated from the final model compared with a null model that excluded

all these terms, and included only the main effects of the non-hydrometeorological

covariates.[197]

As sensitivity analyses, the results of the single variable models were

compared to those obtained using measurements from the nearest weather station

to each site (data which have been described elsewhere[195]) in place of the GLDAS

estimates for those variables and dates for which such data were available using

moving 7-day averages of the GLDAS values. Sensitivity of the results to excluding

the data from the Bangladesh and Peru sites was also assessed. Analyses were carried out using Stata 13.1[177] and R 3.4 (http://www.r-project.org/).

6.3. Results

The box plots in figure 16 show the distribution of each exposure variable for the days on which stool samples were collected. Precipitation and surface runoff have highly skewed distributions and a large number of days on which the value was zero. The values for surface pressure are normally, but narrowly distributed within each site with several sites showing little or no overlap in their distribution with other sites. The highest levels of precipitation and surface runoff and the highest mean values for soil moisture and specific humidity are seen at the Bangladesh and Peru sites, reflecting the conditions at their locations in, respectively, an alluvial, deltaic plain subject to annual monsoon and a flood-prone confluence of several Amazon tributaries.[113,120]
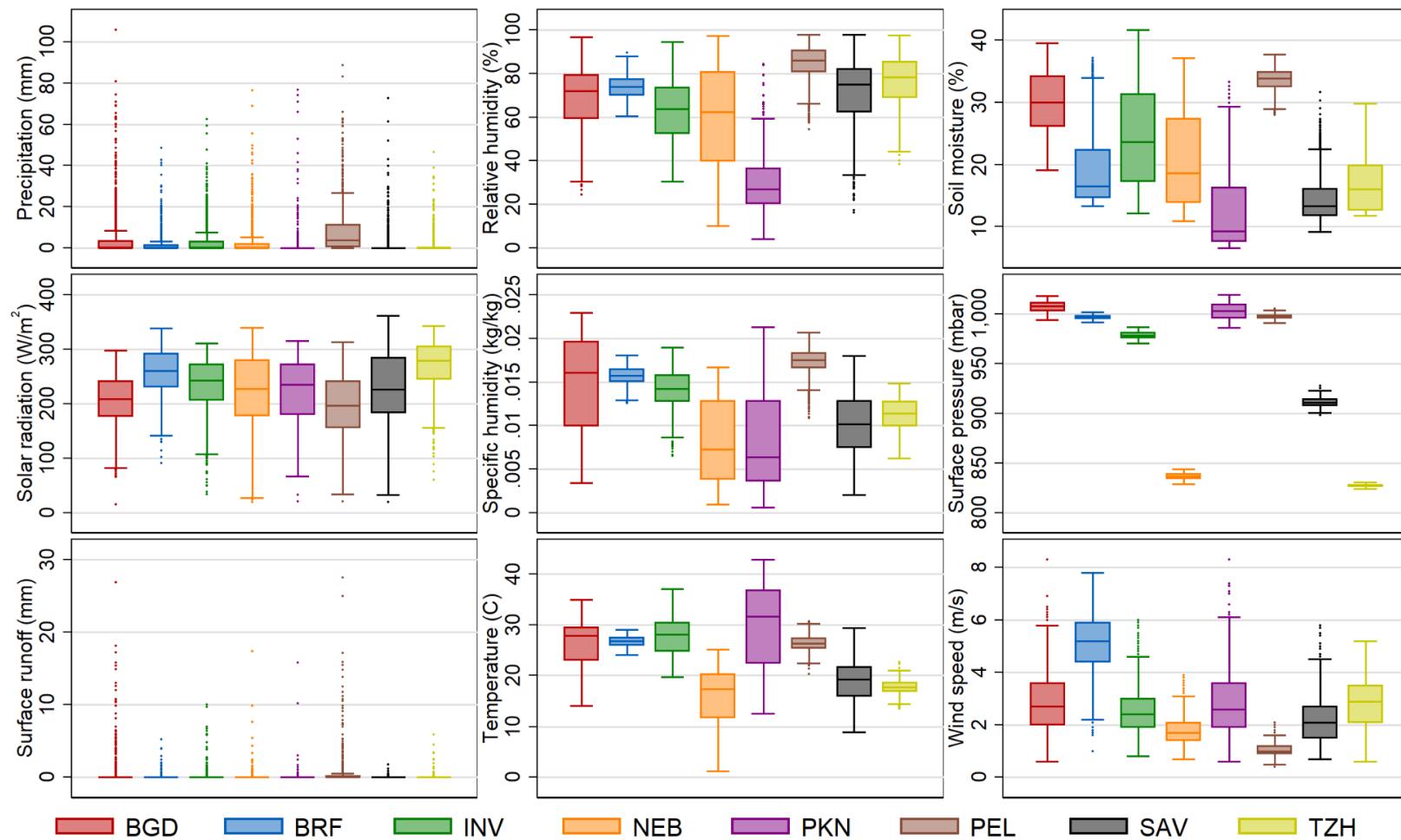
Figure 16: Box-plots of the distributions of the nine hydro-meteorological variables demonstrating significant variability between different hydrometeorological variables at the eight MAL-ED sites. BGD = Dhaka, Bangladesh; BRF = Fortaleza, Brazil; INV = Vellore, India; PKN = Naushero Feroze, Pakistan; PEL = Loreto, Peru; SAV = Venda, South Africa; TZH = Haydom, Tanzania.

Figure 17 shows the *p*-values by significance level from the individual and distributed-lag models and selected lag lengths for the three effects of the nine hydrometeorological variables on rotavirus infection status. Figures 18a-c are visualizations of the probabilities of rotavirus infection predicted by, respectively, the models of the medium-term, absolute, medium-term, within-site and short-term, within-site effects of each of the nine hydrometeorological variables. For variables for which multiple lags met the criteria for inclusion in subsequent models, only the one that showed the highest level of statistical significance is shown in the figures, but the magnitude and shape of the association was not found to change substantial for the other lag lengths (not shown). Tables 13, 14 and 15 summarize the Wald test chi squared statistics for the combined significance of the terms for the main effect of each of the hydrometeorological variables (i.e. the linear or cubic spline terms) from the individual variable models and, for those included in it, the final model, along with those for included interactions with vaccine category, specimen type and between the main exposures themselves

6.3.1. Precipitation

The absolute effect of precipitation on rotavirus only showed a slightly statistically significant association over an 8-day lag, but when site-specific deviations were used instead of absolute values, the effect was slightly statistically significant over several lag lengths, moderately so over a 6- and 8-day lag, but with no lag retaining significance in the DLM. After adjusting the standardized effect for seasonality, the association became statistically significant at almost every lag length and highly so for several, with 8- and 9-day lags retaining significance in the DLM.

The absolute effect of precipitation on rotavirus probability was small in magnitude, linear and direct while the standardized effect was similar, but showed a slightly elevated probability below the site-specific median and neither of these first two effects showed a statistically significant interaction with specimen type. After adjusting for seasonality, the standardized effect predicted a U-shaped association for both episode types in the low extreme of the distribution, with a minimum predicted probability occurring at approximately the site-specific median and a statistically significant interaction with specimen type.

Although only slightly significant in the single-variable model, precipitation over an 8-day lag was retained in the final model for the absolute effect, remaining slight statistically significant in the presence of the other hydrometeorological variables and their interactions. Similarly, for the standardized effect, precipitation over an 8-day lag retained statistical significance that was high in the main effect

and moderate in the final model. In the seasonality adjusted multi-variable models, precipitation over both an 8- and 9-day lag remained statistically significantly associated with the rotavirus outcome in both the main effect and the interaction models. No terms for the interactions between precipitation and any other hydrometeorological variable were retained in any of the three final models, although in the final seasonality-adjusted model, an interaction between precipitation over a 9-day lag and specimen type was included.

		Lag lengths								
		2	3	4	5	6	7	8	9	10
Precipitation (mm)	Absolute effect							▲		
	Standardized effect							▲		
	Seasonality-adjusted effect							▲	▲	
Relative humidity (%)	Absolute effect					▲				
	Standardized effect					▲				▲
	Seasonality-adjusted effect									▲
Soil moisture (%)	Absolute effect								△	▲
	Standardized effect					▲	△		△	▲
	Seasonality-adjusted effect					△	▲			
Solar radiation (W/m2)	Absolute effect	△								
	Standardized effect	▲								
	Seasonality-adjusted effect	▲								
Specific humidity (kg/kg)[xv]	Absolute effect									
	Standardized effect									
	Seasonality-adjusted effect									
Surface pressure (mbar)	Absolute effect	▲								
	Standardized effect		▲	▲						
	Seasonality-adjusted effect		▲							
Surface runoff (mm)	Absolute effect							△		
	Standardized effect							△		
	Seasonality-adjusted effect	△	△					△	△	
Temperature (C)	Absolute effect	△		▲						
	Standardized effect	▲								
	Seasonality-adjusted effect	△							▲	
Wind speed (m/s)	Absolute effect		△					▲	▲	▲
	Standardized effect							△		
	Seasonality-adjusted effect							△		

Legend:
- $p < 0.001$
- $p = 0.001 – 0.01$
- $p = 0.01 - 0.05$
- $p < 0.05$ in DLM

△ Included in stepwise selection   ▲ Selected by stepwise selection

Figure 17: Significance levels of associations found by individual and distributed lag models (DLM) and selected lag lengths for three effects of nine hydrometeorological variables on rotavirus infection status.

[xv] Specific humidity was excluded from the multi-variable models due to it being almost entirely a function of temperature and relative humidity

6.3.2. Relative humidity

Relative humidity was highly statistically significantly associated with rotavirus infection in all three effect models over all considered lag lengths. For the two effects that did not adjust for seasonality, the association retained significance in the DLM at a 6-day lag and for the two standardized effects at a 10-day lag, all of which remained significant in the stepwise selection models. In all three effect models, there was strong statistical evidence for an interaction of relative humidity with specimen type, such that the shape of the effect, as well as the magnitude, is different for symptomatic and asymptomatic episodes. In the model that used absolute values of relative humidity, the effect was small in magnitude - particularly for asymptomatic episodes – direct and fairly linear up to a maximum predicted probability occurring at around 70% relative humidity, at which point the association reached a plateau. When relative humidity was expressed as deviations from its site-specific median value, the overall direction of the association with diarrheal rotavirus reversed, so that the highest probability was seen at the low extreme of the distribution, with a second peak occurring just above the median and the lowest probability above 50 percentage points' deviation. Adjusting for seasonality preserved the shape but reduced the magnitude of this association for symptomatic episodes, while showing a very low-magnitude inverse U-shaped association for asymptomatic infections. Terms for the interaction of relative humidity with temperature and with soil moisture were statistically significant in the final model for all three effects, while its interaction with solar radiation and vaccine category was significant in the standardized effect model.
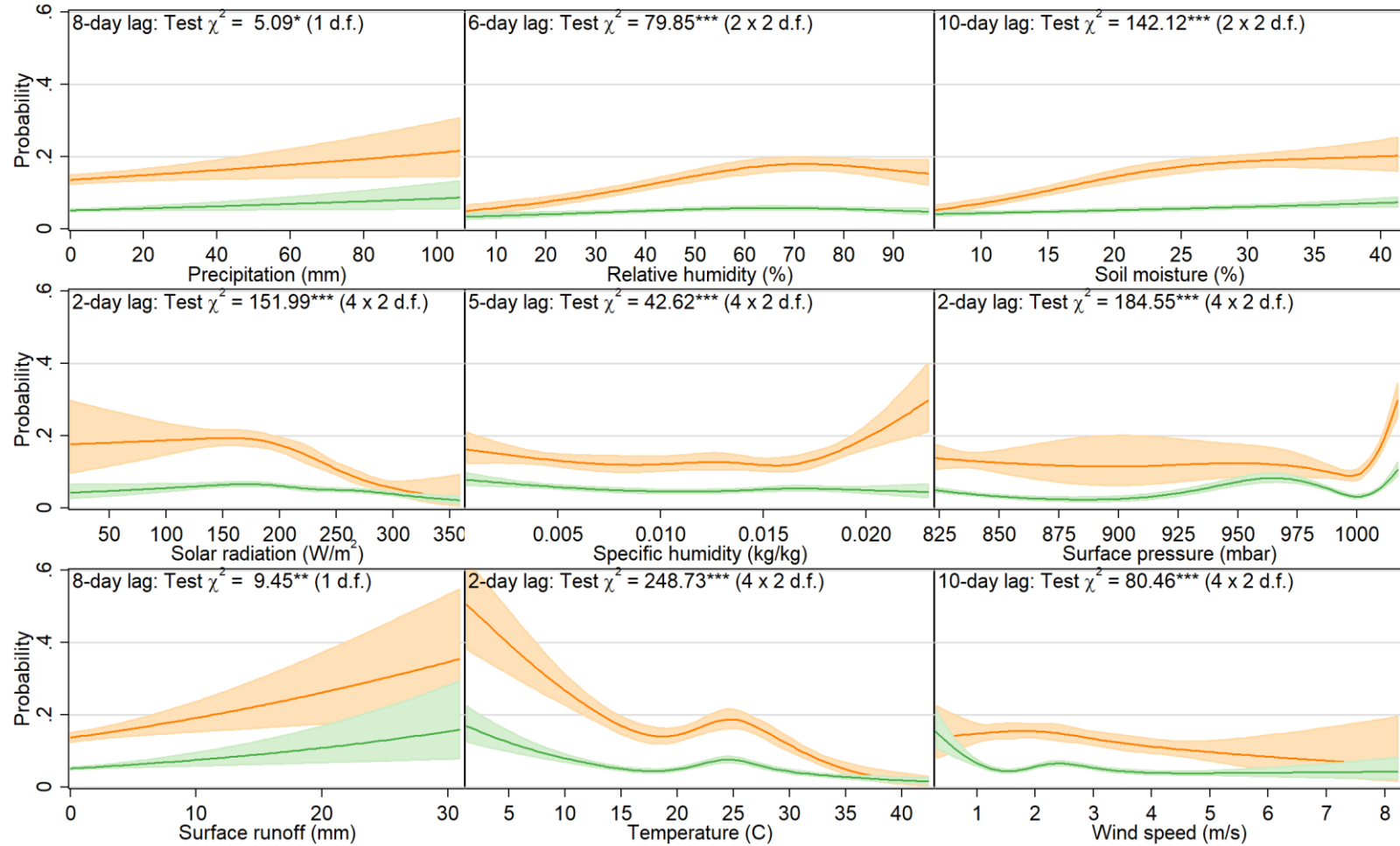
Figure 18a: Probabilities of rotavirus infection predicted by single-variable absolute effect models of for nine hydrometeorological variables in the MAL-ED sites, for both diarrheal (orange) and monthly (green) stool samples[xvi]

[xvi] "d.f." = degrees of freedom. Variables with "x 2 d.f." are those for which terms for an interaction with specimen type were included. *** p<0.001, ** p=0.001 – 0.01, * p=0.01 – 0.05
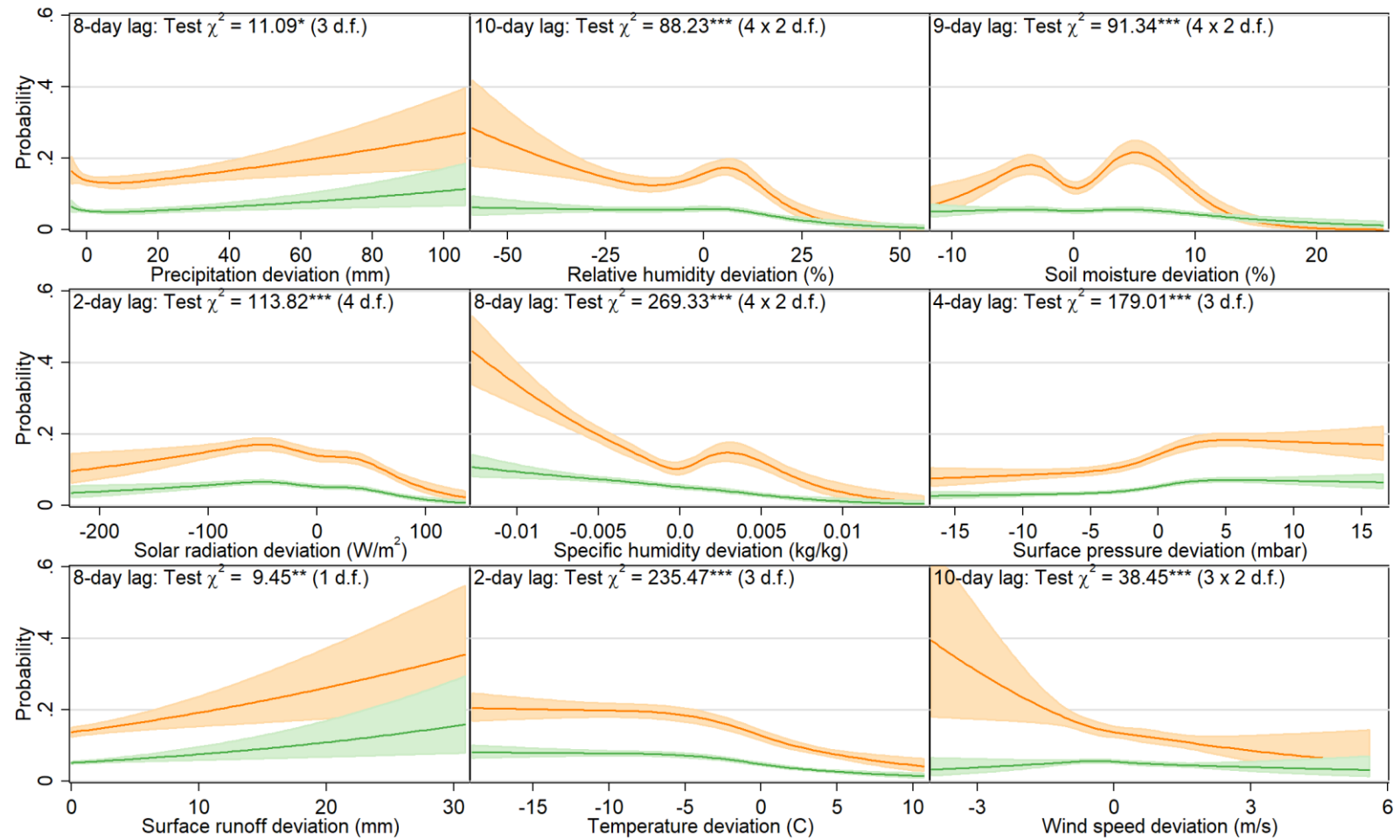
Figure 18b: Probabilities of rotavirus infection predicted by single-variable standardized effect models of for nine hydrometeorological variables in the MAL-ED sites, for both diarrheal (orange) and monthly (green) stool samples[xvii]

[xvii]"d.f." = degrees of freedom. Variables with "x 2 d.f." are those for which terms for an interaction with specimen type were included. *** p<0.001, ** p=0.001 – 0.01, * p=0.01 – 0.05
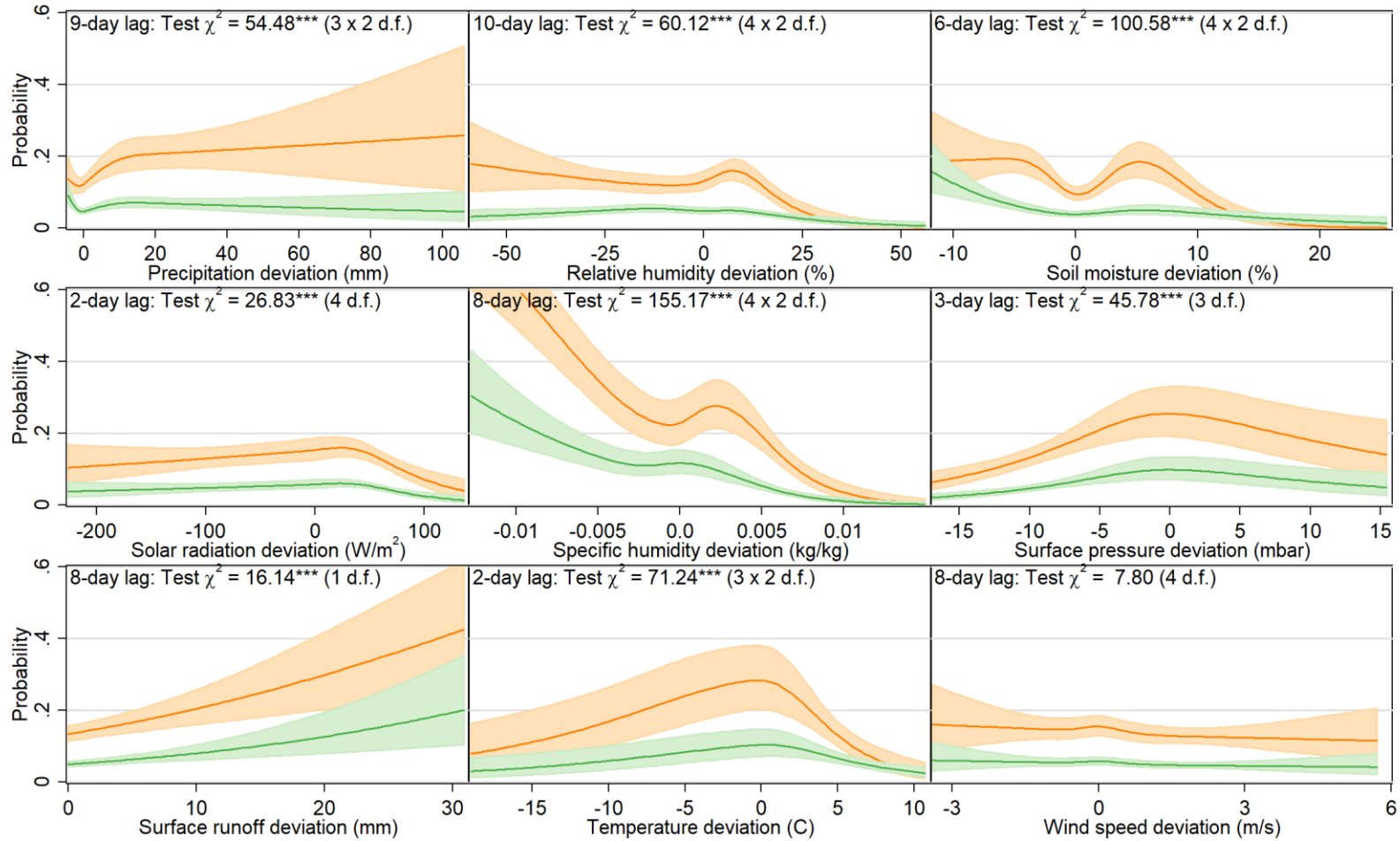
Figure 18c: Probabilities of rotavirus infection predicted by single-variable seasonality-adjusted, standardized effect models of for nine hydrometeorological variables in the MAL-ED sites, for both diarrheal (orange) and monthly (green) stool samples[xviii]

[xviii]"d.f." = degrees of freedom. Variables with "x 2 d.f." are those for which terms for an interaction with specimen type were included. *** p<0.001, ** p=0.001 – 0.01, * p=0.01 – 0.05

6.3.3. Soil moisture

Soil moisture was also highly statistically significantly associated with the outcome at all lag lengths for all three effects, with lags of 9 and 10 days retaining significance in the DLM for the two unadjusted effects and 6- and 7-day lags retained in the DLMs for the two standardized effects. The shape and magnitude of the association between soil moisture and rotavirus closely resembled that of relative humidity in the absolute effect model, but attained a much higher level of statistical significance. When site-specific deviations were used instead of absolute values, the association with diarrheal rotavirus episodes took on a complex, M shape with peaks immediately above and below the site-specific median and the lowest values seen at the extremes of the distribution. After adjusting the site-specific effect for seasonality, the effect at the lower end of the distribution attenuated to a plateau for diarrheal rotavirus, while an indirect association became evident for asymptomatic disease. Terms for the interaction of soil moisture with specimen type, with relative humidity and with temperature were retained in the final models for all three effects, and with vaccine category and in the absolute and standardized (seasonality-unadjusted) effect models.

6.3.4. Solar radiation

The absolute and standardized effects of solar radiation on rotavirus infection status were highly statistically significant across all lag lengths from 2 to 10 days in the individual lag models, while the seasonality-adjusted effect was slightly significant at 8-, 9- and10-day lags, moderately so at 4 and 7 days and only

highly significant at a 2-day lag, the length at which the association retained

statistical significance in the DLM for all three effects. The absolute effect of solar

radiation on rotavirus showed an inverse association above the value of around 170

W/m^2 though the magnitude of this effect was far greater for diarrheal rotavirus

than for asymptomatic episodes. When expressed as local deviations, solar

radiation's association took on an inverse U-shaped with a peak predicted

probability occurring at around 50 W/m^2 below the site-specific median, but when

Fourier terms for seasonality were introduced in the seasonality-adjusted effect

model, the peak shifted to a value above the median, with the most marked effect

being an inverse association in the upper quartile of the distribution. Solar radiation

was excluded from the final model for the absolute effect by stepwise selection, but

was retained in the final models for the two standardized effects, showing evidence

for interaction of this variable with relative humidity before adjusting for

seasonality.

6.3.5. Specific humidity

As with soil moisture, specific humidity showed a highly statistically

significantly association with rotavirus infection status across all lag lengths for all

three effects (with the exception of a 6-day lag for the absolute effect), with a 5-day

lag retaining significance in the all three DLMs, and an 8-day lag in the DLMs for the

two standardized effects. The predicted probability of rotavirus infection decreased

with increasing absolute values of specific humidity for both specimen types, but,

for diarrheal episodes, the direction of the association reversed above a value of

around 0.016 kg/kg and increased markedly to a peak value at the upper end of the

range. The standardized effect of specific humidity had a similar shape and magnitude as relative humidity, but attained a far higher level of statistical significance. Upon adjusting for seasonality, the site-specific effect increased in magnitude considerably but decreased in statistical significance.

6.3.6. Surface pressure

Surface pressure too was statistically significant at all lag lengths for all three modelled effects but, while no lag length retained significance in the DLM for the absolute or seasonality-adjusted effects, 3- and 4-day lags did for the standardized (seasonality-unadjusted) effect. The association between absolute values of surface pressure and predicted rotavirus probability took on a "hockey stick curve" shape, with the most marked effect occurring above a value of around 1,000 mbar. This shape changed considerably when surface pressure deviations were substituted for absolute values such that, in the standardized effect model, the most marked effect was seen within the approximately 10 millibars surrounding the site-specific median. After adjusting for seasonality, the association took on an inverse U shape with a peak predicted probability just below the site-specific median. Surface pressure was retained by stepwise selection in the final models for all three effects, and the terms for its interaction with temperature and soil moisture were retained in the two final interaction models for the standardized effects. The final interaction model for the absolute effect also included terms for surface pressure's interaction with specimen type and vaccine category.

6.3.7. Surface runoff

Because surface runoff has such a highly right-skewed distribution, it had a median value of zero at all sites, meaning that there was no difference between the absolute values and the site-specific deviations. In the two models that did not adjust for seasonality, surface runoff was moderately statistically significant at an 8-day lag and slightly so at 3- and 9-day lags, but in the seasonality-adjusted effect model, it was moderately to highly statistically significant at 2-, 3-, 8- and 9-day lags all of which retained significance in the DLM. In all three effect models, surface runoff had a direct, linear association with rotavirus for both specimen types. Surface runoff was not retained by stepwise selection in any of the final effect models.

6.3.8. Temperature

For all three effects, temperature was highly statistically significantly associated with rotavirus across all lag lengths. In the absolute effect model temperature lagged by 2 and 4 days was included in the stepwise selection model and the 4-day lag was selected by this method, while for the standardized effect, a 2-day lag and for the seasonality-adjusted effect, a 9-day lag were retained in the final models. The absolute effect model predicted a strong and highly statistically significant inverse association with temperature, with the predicted probability of rotavirus declining from a peak at the cold extreme and with a secondary peak at around 25C. The medium term, standardized effect results showed a much lower magnitude but similarly statistically significant inverse association, while after adjusting for seasonality, the seasonality-adjusted effect took on an inverse U-

shaped association with the peak predicted probability occurring at around the site-specific median temperature. In addition to the interactions previously noted, the terms for the interaction of temperature with vaccine category were highly statistically significant in the final seasonality-adjusted effect model.

6.3.9. Wind speed

In the absolute and standardized effect models, wind speed was highly statistically significantly associated with rotavirus at nearly all lag lengths, with lags of 3 and 8 – 10 days all significant in the absolute effect DLM and 8 days in the standardized effect DLM. Furthermore, in the absolute effect model three separate, consecutive lag lengths were retained by stepwise selection, though only one of these (10-days) was kept in the final interaction model since the other two lost all significance in the presence of interactions between other variables. In the individual lag models for the seasonality-adjusted effect, only a 3- and 4-day lag were slightly statistically significant and none were significant in the DLM. Wind speed was not retained in the final models for either of the standardized effects. The absolute effect of wind speed predicted very different patterns for symptomatic – gently increasing probability up to 2 m/s, steadily declining thereafter – compared to asymptomatic – decreasing probability from a peak at 0 to a low at 1.5 m/s with a smaller second peak at around 2.5 m/s. The standardized effect predicted a more linear, inverse association for diarrheal samples, while no association with wind speed was evident from the results of the seasonality-adjusted effect model. Terms for the interaction of wind speed with specimen type in the final absolute effect model were highly statistically significant.

Table 13: Absolute effect - Wald test chi squared statistics (with degrees of freedom) for associations between hydrometeorological variables, covariates and their interactions from logistic models fitted with a GLM using absolute values for the exposures before adjusting for seasonality[xix]

	Lag length (days)	Main effect model	Interaction model						
			Main effect	Precipitation (mm)	Relative humidity (%)	Soil moisture (%)	Surface pressure (mbar)	Temperature (C)	Wind speed (m/s)
Precipitation (mm)	8	7.63**(1)	4.20* (1)						
Relative humidity (%)	6	26.78*** (2)	12.61**(2)	-					
Soil moisture (%)	10	17.75*** (2)	3.64 (2)	-	15.62** (4)				
Surface pressure (mbar)	2	162.03*** (4)	38.63*** (4)	-	-	55.76*** (8)			
Temperature (C)	4	304.46*** (4)	19.01*** (4)	-	43.11*** (8)	42.78*** (8)	82.11*** (16)		
Wind speed (m/s)	8	7.38* (2)	-	-	-	-	-	-	
	9	14.18**(4)	-	-	-	-	-	-	
	10	11.18* (4)	24.71*** (4)	-	-	-	-	-	
Specimen	0	421.87*** (1)	5.97* (1)	-	-	31.75*** (2)	22.02*** (4)	-	20.87*** (2)
Vaccine category	0	55.28*** (1)	0.61 (1)	-	-	7.10* (2)	23.70*** (4)	-	-
Age	0	59.34*** (2)	62.91*** (2)	-	-	-	-	-	-
Age/vaccine interaction	0	6.20* (2)	8.00* (2)	-	-	-	-	-	-
COD for model		4.2%				5.5%			
Partial COD for EO variables		2.0%				3.4%			

[xix] *** p <0.001, ** p = 0.001 – 0.01, * p = 0.01 - 0.05. Numbers in parentheses indicate degrees of freedom – the number of model terms. COD = coefficient of discrimination, EO = earth observation.

Table 14: Standardized effect - Wald test chi squared statistics (with degrees of freedom) for associations between hydrometeorological variables, covariates and their interactions from logistic models fitted with a GLM using deviations of the exposures from the site-specific medians before adjusting for seasonality[xx]

| | Lag length (days) | Main effect model | Interaction model | | | | | | | | | |
| | | | Main effect | Precipitation (mm) | Relative humidity (%) | | Soil moisture (%) | | Solar radiation (W/m²) | Surface pressure (mbar) | | Temperature |
					6-day lag	10-day lag	6-day lag	9-day lag		3-day lag	4-day lag	
Precipitation (mm)	8	46.02*** (3)	12.06** (3)									
Relative humidity (%)	6	25.46*** (4)	0.75 (4)	-								
	10	26.50*** (4)	12.67* (4)	-								
Soil moisture (%)	6	13.00* (4)	7.92 (4)	-	26.28** (12)	28.17* (16)						
	10	13.32** (4)	18.83*** (4)	-	-	-						
Solar radiation (W/m²)	2	33.25*** (4)	15.81** (4)	-	36.34*** (12)	-	-	-				
Surface pressure (mbar)	3	9.98* (3)	25.04*** (3)	-	-	-	-	53.28*** (12)	-			
	4	16.00** (3)	16.87*** (3)	-	-	-	-	-	-	-		
Temperature (C)	2	76.88*** (3)	9.82* (3)	-	32.12** (12)	29.83** (12)	30.67** (12)	39.30*** (12)	-	45.13*** (9)	-	
Specimen	0	391.10*** (1)	33.54*** (1)	-	-	-	-	14.41** (4)	-	-	-	-
Vaccine category	0	112.61*** (1)	21.47*** (1)	-	25.95*** (4)	-	-	39.05*** (4)	-	-	-	-
Age	0	66.60*** (2)	66.67*** (2)	-	-	-	-	-	-	-	-	-
Age/vaccine interaction	0	6.62* (2)	5.68 (2)	-	-	-	-	-	-	-	-	-
COD for model		3.8%						5.4%				
Partial COD for EO variables		1.6%						3.2%				

[xx] *** p <0.001, ** p = 0.001 – 0.01, * p = 0.01 - 0.05. Numbers in parentheses indicate degrees of freedom – the number of model terms. COD = coefficient of discrimination, EO = earth observation.

Table 15: Seasonality-adjusted standardized effect - Wald test chi squared statistics (with degrees of freedom) for associations between hydrometeorological variables, covariates and their interactions from logistic models fitted with a GLM using deviations of the exposures from the site-specific medians after adjusting for seasonality[xxi]

| | Lag length (days) | Main effect model | Interaction model | | | | | | | |
| | | | Main effect | Precipitation (mm) | | Relative humidity (%) | Soil moisture (%) | Solar radiation (W/m²) | Surface pressure (mbar) | Temperature (C) |
				8-day lag	9-day lag					
Precipitation (mm)	8	29.59*** (3)	9.00* (3)							
	9	15.42** (3)	18.60*** (3)	-						
Relative humidity (%)	10	22.86*** (4)	12.13** (3)	-	-					
Soil moisture (%)	10	49.91*** (4)	19.46*** (4)	-	-	26.01* (12)				
Solar radiation (W/m²)	2	15.49** (4)	16.24** (4)	-	-	-	-			
Surface pressure (mbar)	2	10.63* (3)	-	-	-	-	-	-		
Temperature (C)	4	11.56** (3)	4.35 (3)	-	-	46.06*** (9)	28.24**(12)	-	-	
Specimen	0	434.30*** (1)	51.64*** (1)	-	13.02** (3)	-	17.75** (4)	-	-	-
Vaccine category	0	122.91*** (1)	8.70**(1)	-	-	-	-	-	-	13.39** (3)
Age	0	71.03*** (2)	68.01*** (2)	-	-	-	-	-	-	-
Age/vaccine interaction	0	8.72*(2)	8.53* (2)	-	-	-	-	-	-	-
Seasonality	0	326.32*** (32)	296.88*** (32)	-	-	-	-	-	-	-
COD for model		4.8%	5.4%							
Partial COD for EO variables		2.6%	3.2%							

[xxi] *** p <0.001, ** p = 0.001 – 0.01, * p = 0.01 - 0.05. Numbers in parentheses indicate degrees of freedom – the number of model terms. COD = coefficient of discrimination, EO = earth observation.

6.3.10.   Model Predictions

For all three effects, the interaction model explained a larger proportion of the variability on the rotavirus outcome than the main effects model. The final model for the seasonality-adjusted, standardized effect was the model that had the highest explanatory power according to the coefficients of discrimination. In absolute terms, however, this proportion was small, with the hydrometeorological variables and their interactions and seasonality explaining just 3.2% of the variability in rotavirus positivity. Figure 19 shows the bootstrapped probabilities of rotavirus infection in diarrheal (symptomatic) and monthly (asymptomatic) stools predicted by the final absolute effect model for each MAL-ED site for the period of follow-up and for the simulated future scenario. At all the sites and for both types of episode, the model predicted a decrease in probability following an increase in the value of the EO variables, with the exception of Dhaka, Bangladesh, Naushero Feroze Pakistan and, far less markedly, symptomatic episodes in Vellore, India, all of which were predicted to see these probabilities increase. The ranges of uncertainty for these predictions were wide, however, particularly for the Pakistan and Bangladesh sites and only in the Brazil and Peru sites, where a reduction in probability to almost zero was predicted under the future scenario, did the confidence intervals for the two predictions not overlap.
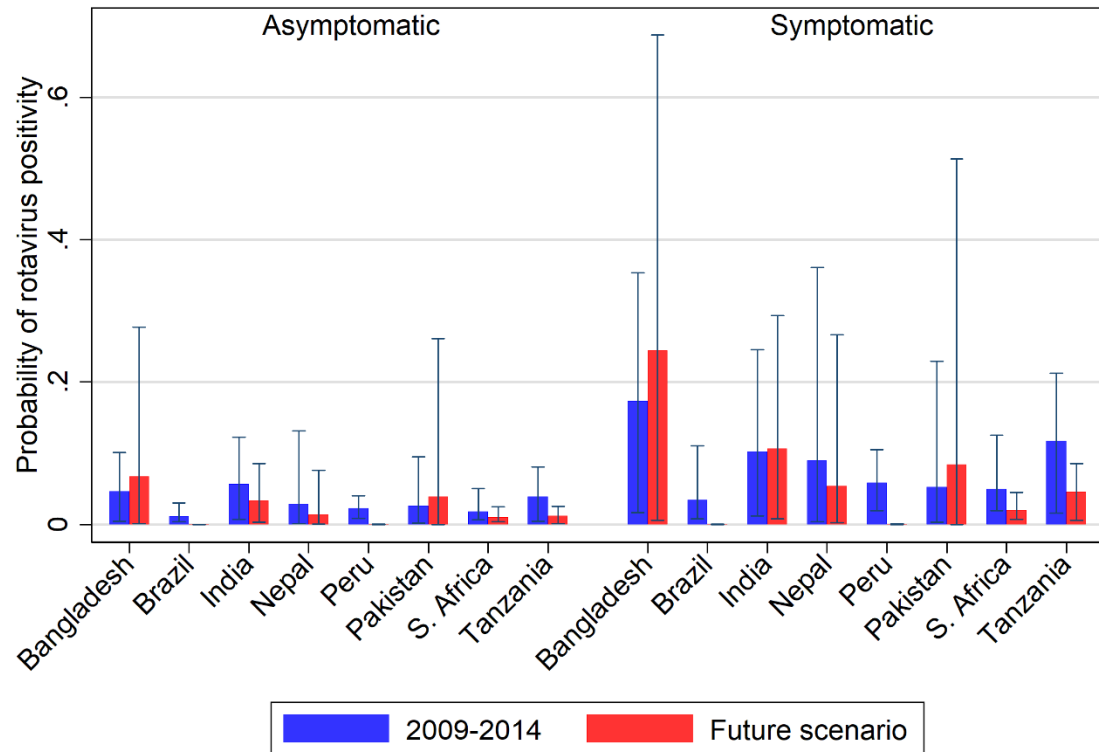
Figure 19: Bootstrapped probabilities (with 95% confidence intervals) of symptomatic and asymptomatic rotavirus episodes predicted by the final absolute effect model for 2009 – 2014 (using the true GLDAS estimates as predictors) and for a simulated future scenario in which the GLDAS values were all increased by 2%[xxii]

[xxii] BGD = Dhaka, Bangladesh; BRF = Fortaleza, Brazil; INV = Vellore, India; NEB = Bhaktapur, Nepal; PKN = Naushero Feroze Pakistan; PEL = Loreto, Peru; SAV = Venda, South Africa; TZH = Haydom, Tanzania

6.4. Discussion

Among the EIDs, rotavirus is perhaps the pathogen for which associations with climate factors have been most thoroughly explored. While recent studies have incorporated multiple climate variables[131], single-day exposure estimates[59,195], differential lag relationships[59,131], interactions[19], non-linearity[59,96], multiple climate zones[20] and adjustment for seasonality[60,198] this analysis is the first to address all of these factors while also assessing differences in effect between symptomatic and asymptomatic infections. An important general finding is that numerous hydrometeorological parameters – including several that are not commonly measured by weather stations – exhibit complex, non-linear associations with rotavirus infection that differ by episode type and may be independently, and highly statistically significant over multiple consecutive or non-consecutive lags, including as short a period as two days. More specifically, the results presented here show evidence for the hypothesis that the effect of climate on rotavirus transmission is mediated by four independently operating mechanisms:

6.4.1. Waterborne dispersion

It is thought that precipitation drives enteric pathogen transmission via different mechanisms at the two extremes, with heavy rainfall and runoff flushing microorganism from soils and surfaces into surface water (and possibly groundwater[199]) sources and drought conditions concentrating them in these environments.[105] The U-shaped association between precipitation and rotavirus infection predicted by the seasonality-adjusted effect model, with low

145

probability occurring at the site-specific median, conforms to this theory. That this effect is only seen in the standardized models is revealing. Small-scale water and drainage systems are generally set up to cope with precipitation and runoff levels within the range that is typical for a particular location and may be overwhelmed levels of precipitation that would be within the normal range at another location. Furthermore, the finding that precipitation was much less statistically significant in the absolute and standardized effect models may be due to these two competing pathways (the "runoff effect" versus the "concentration effect"[105]) influencing rotavirus transmission in opposite directions in the rainy season compared to drier times of the year, which may cancel each other out unless seasonality is adjusted for. The lack of an interaction identified between precipitation and any other hydrometeorological variable is consistent with this variable operating via a separate pathway. That precipitation and the closely related variable of surface runoff both tended to be statistically significant within two lag ranges – 2-4 days and 7-9 days – is also suggestive of a common, independent pathway. It is possible that the shorter lag lengths constitute evidence of primary and the longer, secondary transmission of the pathogen. The seasonality-adjusted, negative effect of solar radiation on rotavirus was also only significant within these two lag ranges, which may represent evidence of confounding by precipitation, since levels of solar radiation reaching the earth will be lower on rainy, overcast days when cloud cover obscures sunlight.

6.4.2. Airborne dispersion

An airborne route of transmission for rotavirus has long been suspected but not conclusively proven.[200] Hypothesized vehicles for this type of transmission include either dried dust[21] or liquid particles[19] which may either be ingested directly[132] or inhaled first, before migrating to the gastrointestinal tract through the swallowing of respiratory secretions.[201] Aerosol transmission of this kind is to be distinguished from droplet transmission by the smaller size of the contaminated particles involved (<5μm diameter according to the WHO's definition), which permits them to suspend in the air for days at a time, settling at a rate that is a function both of their size and the movement of the air and capable of infecting susceptible individuals at a greater distance from their source.[202–204] While aerosol transmission is usually associated with particles secreted from the respiratory tract, formation of aerosols can also occur from disposal of excreta in diapers and through toilet flushing.[19] Once suspended, environmental conditions affect the pathogens' ability to survive and remain infectious in aerosols by determining the size of the particles and the rate at which they desiccate.[202,203,205]

Factors on this pathway that are assessed in this analysis include temperature, humidity pressure and wind speed, a cluster of closely related variables that were all retained in all three final effect models (with the exception of wind speed) and showed considerable mutual interaction. We do not consider it biologically plausible, nor do we find evidence in the literature for, a direct effect of surface pressure on rotavirus infectivity. Instead, we

speculate that the apparent association between surface pressure and rotavirus infection, which is consistently statistically significant across the three effects at all lag lengths is due to GLDAS estimates of pressure being a better predictor of true advective air movement at ground level than GLDAS estimates of wind speed at 10m height. In the seasonality-adjusted model, predicted probability of rotavirus was highest at the site-specific median, at which stiller conditions would be expected. This is consistent with the negative relationship with wind speed seen in the absolute and standardized effect models and suggests that transmission is promoted when aerosols are able to linger in slow moving air, and that stronger wind may inhibit transmission by transporting these particles away from susceptible individuals.

The direct association between absolute values of relative humidity and rotavirus appeared to reverse to a broadly negative relationship when these were substituted for site-specific deviations. This may be due to rotavirus transmission being higher in MAL-ED sites where conditions tend to be more humid. The standardized effect of relative humidity is consistent with dry dust particles being a vehicle for aerosol transmission, since a peak in the probability of infection is seen at lowest humidity. However, the secondary peak (roughly equal in magnitude in the seasonality-adjusted model) at just above the site-specific median relative humidity, may indicate that liquid aerosols also play a role. The low probability of infection at high relative humidity, may occur because liquid aerosols are unable to suspend for long periods in moisture-

dense air. Comparable changes in the shape of the standardized compared to the absolute effect were seen for specific humidity and soil moisture.

6.4.3. Survival on soil and surfaces

Soil moisture was consistently highly statistically significant across effects and lag-lengths and the fact that it remained so in the presence of relative humidity in the final models suggests that it was not merely acting as an indicator of the general moistness of conditions, but in fact had its own independent effect. Non-linear effects on the ability of enteric viruses to remain viable in soil at varying moisture levels have been documented[135], as has increased persistence of rotavirus on porous surfaces at higher humidity.[142] It has been suggested that, once rotavirus-contaminated aerosols settle, the virus may remain viable for longer if it is then able to adhere to a surface that permits retention of its surrounding moisture[19], an interaction with the airborne dispersion route for which there is evidence in the final models.

While the absolute effect indicated a linear, direct association with higher soil moisture causing higher risk, the within-site effect suggests a more complex, M-shaped relationship (at least for diarrheal rotavirus), where soil moisture values of around 5 percentage points above and 5 below the site-specific median are associated with the highest risk. A possible interpretation of this is that rotavirus survival is optimized on soil that is dry enough to form dust particles but not so dry as to desiccate entirely, or moist enough for adhered viruses to remain viable, but not so moist that they are flushed from the soil.

6.4.4. Host factors

The two effect models for temperature that did not adjust for seasonality predicted a peak in predicted rotavirus probability occurring at the coldest extreme of the distribution yet, in the seasonality-adjusted model, the effect takes on an inverse U-shape. A possible explanation for this is that, at colder times of the year, increased contact rates between susceptible and infected individuals as they congregate indoors promotes transmission, a confounding effect that is removed by adjusting for seasonality. A role for UV radiation has been suggested in the downregulation of the human immune system, a host-mediated factor which could exacerbate the course and severity of viral infections.[19,134] Although some superficial evidence for this is suggested by the fact that solar radiation remains significantly negatively associated with rotavirus infection after adjustment for seasonality, we consider it more likely that this represents confounding due to rainclouds obscuring the sun on rainy days.

Although many of the associations identified in this analysis were highly statistically significant, a large proportion of the variability in the outcome went unexplained by the final models, a fact reflected in the small values for the CODs and the wide confidence intervals for the predictions. This may mean that hydrometeorological predictors are of limited explanatory power and clinical significance relative to household and behavioral factors, which determine exposure and genetic and immunological factors that influence the course of an infection once established. Future analyses may attempt to refine the models presented here to

adjust for easily measured, non-environmental exposures (such as breastfeeding status and sanitation) in order to improve their explanatory power. However, individual-level factors that may be the most determinative of disease risk (such as fucosyltransferase-2 secretor status[206]), may be unfeasible to ascertain at this scale. Modest increases in explanatory power can be gained by specifying terms for interactions between hydrometeorological variables for which effect measure modification has not yet been hypothesized, as demonstrated by the increase in COD for the interaction models relative to the main effects.

The observed proportions of rotavirus positive stool samples were all close to and within the confidence limits of those predicted by the model for the period of follow up for all sites and both episode types. Although the simulated values used for the future predictions were an illustrative example and not intended to be realistic, it is noteworthy that applying the same adjustment to all the EO variable values produced changes in the predicted probabilities in different directions for different sites. The two sites in which the predicted probability under the future scenario increased the most relative to the follow up period were Dhaka, Bangladesh and Naushero Feroze Pakistan, were also those in which the confidence intervals for the predictions were by far the widest.

7. Chapter 7: Implications for modeling and predicting pathogen-specific enteric disease burden

Despite widespread recognition of the seasonal nature of the epidemiology of EID, the mechanisms underlying this phenomenon are poorly understood compared with other disease groups.[42,207] With some notable exceptions, most studies of the seasonality of these diseases have been hindered by data being aggregated up to weekly, monthly or quarterly cumulative incidence, rather than daily estimates.[31,53] There is also a paucity of studies on seasonality of EID coming from developing countries – particularly Africa and South America - and those in the tropics and the southern hemisphere.[48]

Similarly, while climatic drivers of some infections, principally vector-borne diseases, have been relatively well-explored, currently little is understood of their role in EID transmission and there is no unified theoretical framework through which to conceptualize the relative influence of individual drivers.[16,20,93] What exploration there has been has tended to focus on non-specific, morbidity-dependent outcomes such as hospitalizations for acute diarrheal disease and outbreaks of gastroenteritis and not on underlying, background endemicity of specific pathogens.[94] On the exposure side, there has been a disproportionate emphasis on extreme weather events and anomalous climate phenomena such as ENSO and the NAO as predictors.[76,95–97] Less has been explored about small-scale meteorological variability within the normal range.

Because of its unique design, the MAL-ED study has the potential to address many of these knowledge gaps and elucidate the drivers of seasonality by comparing pathogen-specific patterns across sites, and multiple pathogens within the same location. The research presented in this thesis represents an initial attempt to do this with a test-case pathogen of known seasonality and climate sensitivity. Having demonstrated that it is possible to characterize these associations for rotavirus, we may proceed with greater confidence that any similar relationships that are observed for other, less well-characterized EIDs using the same methods, represent valid and generalizable inferences. Furthermore, having demonstrated that the effect estimates from the final models presented here can be used to make predictions under illustrative future climate scenarios, researchers may proceed by combining ever more sophisticated and rigorous climate projections to obtain realistic estimates of future disease burden.

Many previously published analyses of the influence of weather on rotavirus or other health outcomes have tended to aggregate the meteorological exposures over large areas or longer time windows (e.g. weeks or months). This study demonstrates for the first time that associations can still be detected using daily estimates, which in many cases were more highly statistically significantly associated with the outcome than 7-day averages. Where outcome data is available with the precise date of ascertainment, an equally high resolution for the exposure data may be preferred in order to retain the variability in the data and for the most precise characterization of lag effects and the temporal order over which multivariate associations operate.

A major limitation of this study is the lack of representativeness of the site locations. Although there were a large number of observations within each site, the fact that they represent only 8 locations restricts the extent to which the results can be extrapolated to other climate zones. This is an inevitable limitation for community-based studies that implement labor- and resource-intensive active surveillance. However, since the EO datasets from which the predictors were extracted are available at global scale and sub-daily resolution and updated continually, as new studies using similar methods are carried out at different locations, they can be added to the MAL-ED data to derive more precise predictions for more diverse conditions. Furthermore, emerging tools for objective climate regionalization can be combined with the results of these models to divide extensive geographic zones into smaller regions that are homogenous with respect to important climate characteristics.[208]

Appendix 1: Descriptive tables

Table 16: Summary of GLDAS, CHIRPS and weather-station-based hydrometeorological variables for the eight MAL-ED study sites, 2009-2014[23]

	Variable	Median temperature (C)		Precipitation (mm)			Surface pressure (mbar)		Wind speed (m/s)	
		GLDAS	Station	GLDAS	CHIRPS	Station	GLDAS	Station	GLDAS	Station
BGD	Median	27.9	28.8	0.0	0.0	0.0	1,007.3	1,005.7	2.6	0.6
	IQR	6.3	5.5	4.3	6.1	1.8	8.9	8.6	1.4	0.8
	Maximum	35.0	33.9	107.1	103.5	150.1	1,019.6	1,018.3	8.3	7.3
	Minimum	14.1	13.8	0.0	0.0	0.0	993.5	992.8	0.6	0.0
	Completeness	100.0%	59.7%	100.0%	100.0%	59.6%	100.0%	59.1%	100.0%	59.3%
BRF	Median	27.1	27.2	0.0	0.0	0.0	996.9	1,008.7	5.0	4.9
	IQR	1.4	1.3	1.7	0.0	1.0	2.7	2.3	1.7	2.3
	Maximum	29.4	29.8	55.0	118.0	147.1	1,002.2	1,014.0	7.8	9.7
	Minimum	23.3	22.5	0.0	0.0	0.0	991.5	1,003.8	0.9	1.1
	Completeness	100.0%	99.8%	100.0%	100.0%	99.0%	100.0%	70.0%	100.0%	99.8%
INV	Median	28.0	28.9	0.0	0.0	0.0	978.0	-	2.4	-
	IQR	5.8	5.1	2.9	4.0	0.2	5.6	-	1.0	-
	Maximum	37.0	36.5	70.7	72.5	88.4	986.8	-	6.2	-
	Minimum	19.6	20.9	0.0	0.0	0.0	968.4	-	0.8	-
	Completeness	100.0%	84.5%	100.0%	100.0%	84.8%	100.0%	-	100.0%	-
NEB	Median	17.8	20.5	0.0	0.0	0.0	837.6	862.1	1.8	1.4
	IQR	8.5	9.5	2.1	0.0	0.0	4.7	2.5	1.0	0.8
	Maximum	25.2	29.0	76.5	142.0	134.6	847.8	868.7	4.5	5.1
	Minimum	1.3	5.5	0.0	0.0	0.0	828.7	856.1	0.7	0.0
	Completeness	100.0%	97.6%	100.0%	100.0%	83.5%	100.0%	1.1%	100.0%	97.6%
PKN	Median	31.8	29.2	0.0	0.0	0.0	1,003.1	994.8	2.5	0.9
	IQR	14.4	12.8	0.0	0.0	0.0	14.2	12.7	1.8	0.8
	Maximum	42.8	40.5	76.9	24.0	119.9	1,019.0	1,015.6	9.7	6.3
	Minimum	12.1	7.5	0.0	0.0	0.0	986.2	985.2	0.5	0.0
	Completeness	100.0%	99.0%	100.0%	100.0%	98.9%	100.0%	1.0%	100.0%	99.0%
PEL	Median	26.2	27.4	4.4	0.0	0.8	997.2	999.8	1.0	1.2
	IQR	1.9	1.4	11.5	11.2	7.4	3.1	3.0	0.3	0.7
	Maximum	30.7	30.6	88.9	84.1	199.9	1,005.4	1,009.4	2.1	3.5
	Minimum	20.4	18.1	0.0	0.0	0.0	991.1	993.1	0.3	0.0
	Completeness	100.0%	100.0%	100.0%	100.0%	90.9%	100.0%	94.0%	100.0%	94.6%
SAV	Median	19.3	21.9	0.0	0.0	0.0	910.6	946.5	2.1	2.4
	IQR	5.4	5.8	0.0	0.0	0.3	5.9	6.8	1.2	0.8
	Maximum	29.3	29.9	88.7	150.2	136.9	927.5	965.0	6.4	7.5
	Minimum	8.1	9.9	0.0	0.0	0.0	898.0	932.7	0.5	1.0
	Completeness	100.0%	96.1%	100.0%	100.0%	88.2%	100.0%	94.9%	100.0%	96.1%
TZH	Median	18.0	-	0.0	0.0	0.0	827.3	-	2.9	-
	IQR	2.0	-	0.4	0.0	0.0	2.0	-	1.3	-
	Maximum	22.6	-	46.5	47.7	38.5	831.4	-	5.8	-
	Minimum	13.4	-	0.0	0.0	0.0	823.4	-	0.6	-
	Completeness	100.0%	-	100.0%	100.0%	36.4%	100.0%	-	100.0%	-

---

[23] BGD = Dhaka, Bangladesh; BRF = Fortaleza, Brazil; INV = Vellore, India; NEB = Bhaktapur, Nepal; PKN = Naushero Feroze Pakistan; PEL = Loreto, Peru; SAV = Venda, South Africa; TZH = Haydom, Tanzania; IQR = Inter-quartile range

Table 16 (cont'd): Summary of GLDAS, CHIRPS and weather-station-based hydrometeorological variables for the eight MAL-ED study sites, 2009-2014

	Variable	Relative humidity (%)		Specific humidity (kg/kg)		Solar radiation (W/m²)		Soil moisture (%)		Surface runoff (mm)	
		GLDAS	Station	GLDAS	Station	GLDAS	Station	GLDAS	Station	GLDAS	Station
BGD	Median	71.7	-	0.016	-	202.8	-	30.2	-	0.0	-
	IQR	23.2	-	0.010	-	64.9	-	8.9	-	0.0	-
	Maximum	97.2	-	0.024	-	298.4	-	39.8	-	36.1	-
	Minimum	23.1	-	0.003	-	15.3	-	19.1	-	0.0	-
	Completeness	100.0%	-	100.0%	-	100.0%	-	100.0%	-	100.0%	-
BRF	Median	72.7	-	0.016	-	259.6	-	17.0	-	0.0	-
	IQR	8.0	-	0.002	-	61.9	-	8.5	-	0.0	-
	Maximum	92.1	-	0.019	-	338.9	-	38.4	-	7.7	-
	Minimum	58.6	-	0.012	-	45.8	-	13.2	-	0.0	-
	Completeness	100.0%	-	100.0%	-	100.0%	-	100.0%	-	100.0%	-
INV	Median	64.3	69.0	0.014	-	238.7	-	23.8	-	0.0	-
	IQR	21.8	19.5	0.003	-	70.4	-	13.1	-	0.0	-
	Maximum	95.8	99.0	0.020	-	311.1	-	41.6	-	31.2	-
	Minimum	29.1	36.5	0.006	-	15.0	-	12.1	-	0.0	-
	Completeness	100.0%	84.6%	100.0%	-	100.0%	-	100.0%	-	100.0%	-
NEB	Median	62.6	-	0.007	-	222.9	-	18.7	-	0.0	-
	IQR	42.5	-	0.009	-	96.1	-	12.9	-	0.0	-
	Maximum	97.5	-	0.017	-	350.7	-	37.1	-	17.4	-
	Minimum	4.6	-	0.001	-	12.8	-	9.9	-	0.0	-
	Completeness	100.0%	-	100.0%	-	100.0%	-	100.0%	-	100.0%	-
PKN	Median	27.3	63.5	0.006	-	235.5	-	7.8	-	0.0	-
	IQR	16.6	13.5	0.009	-	98.5	-	3.3	-	0.0	-
	Maximum	84.5	98.0	0.021	-	322.5	-	33.6	-	15.8	-
	Minimum	4.0	27.0	0.001	-	6.9	-	6.6	-	0.0	-
	Completeness	100.0%	100.0%	100.0%	-	100.0%	-	100.0%	-	100.0%	-
PEL	Median	87.3	-	0.018	-	191.6	-	34.1	-	0.0	-
	IQR	9.8	-	0.002	-	86.3	-	2.1	-	0.2	-
	Maximum	98.1	-	0.021	-	313.4	-	38.2	-	27.5	-
	Minimum	54.5	-	0.011	-	14.8	-	28.0	-	0.0	-
	Completeness	100.0%	-	100.0%	-	100.0%	-	100.0%	-	100.0%	-
SAV	Median	75.2	67.0	0.010	-	222.0	-	14.6	-	0.0	-
	IQR	21.0	19.0	0.005	-	98.7	-	4.8	-	0.0	-
	Maximum	98.0	97.0	0.018	-	362.2	-	31.6	-	3.4	-
	Minimum	16.2	15.0	0.002	-	9.2	-	9.2	-	0.0	-
	Completeness	100.0%	74.4%	100.0%	-	100.0%	-	100.0%	-	100.0%	-
TZH	Median	77.0	-	0.011	-	275.8	-	15.5	-	0.0	-
	IQR	17.1	-	0.003	-	57.6	-	6.7	-	0.0	-
	Maximum	97.7	-	0.015	-	348.2	-	29.8	-	5.9	-
	Minimum	38.6	-	0.006	-	35.1	-	11.7	-	0.0	-
	Completeness	100.0%	-	100.0%	-	100.0%	-	100.0%	-	100.0%	-

Bibliography

1   Liu L, Oza S, Hogan D, *et al.* Global, regional, and national causes of under-5 mortality

in 2000–15: an updated systematic analysis with implications for the Sustainable

Development Goals. *Lancet* 2016; **388**: 3027–35.

2   World Health Organization. Preventive chemotherapy in human helminthiasis.

Geneva, 2006.

3   Korpe PS, Petri WA. Environmental enteropathy: critical implications of a poorly

understood condition. *Trends Mol Med* 2012; **18**: 328–36.

4   Ali A, Iqbal NT, Sadiq K. Environmental enteropathy. *Curr Opin Gastroenterol* 2016;

32: 12–7.

5   Gilmartin AA, Petri WA. Exploring the role of environmental enteropathy in

malnutrition, infant development and oral vaccine response. *Philos Trans R Soc Lond

B Biol Sci* 2015; **370**. DOI:10.1098/rstb.2014.0143.

6   Lanata CF, Fischer-Walker CL, Olascoaga AC, *et al.* Global causes of diarrheal disease

mortality in children. *PLoS One* 2013; **8**: e72788.

7   Platts-Mills JA, Babji S, Bodhidatta L, *et al.* Pathogen-specific burdens of community

diarrhoea in developing countries: a multisite birth cohort study (MAL-ED). *Lancet

Glob Heal* 2015; **3**: e564-75.

8   Hammitt L. Rotavirus Vaccines. 2015.

9   Hall AJ. Noroviruses: the perfect human pathogens? *J Infect Dis* 2012; **205**: 1622–4.

10   Harrison SC. Virology. Looking inside adenovirus. *Science* 2010; **329**: 1026–7.

11   Public Health Agency of Canada. Adenovirus (Serotypes 40 & 41). Pathog. Saf.

Data Sheets Risk Assess. 2010. http://www.phac-aspc.gc.ca/lab-bio/res/psds-

ftss/adenovirus-eng.php (accessed Oct 20, 2016).

12   Centers for Disease Control and Prevention. Adenovirus. CDC A-Z Index. 2015.

http://www.cdc.gov/adenovirus/ (accessed Dec 10, 2015).

13   Bosch A, Pintó RM, Guix S. Human astroviruses. *Clin Microbiol Rev* 2014; **27**: 1048–74.

14   Walter JE, Mitchell DK. Astrovirus infection in children. *Curr Opin Infect Dis* 2003; **16**: 247–53.

15   Guadagnucci Morillo S, Sampaio TavaresTimenetsky M do C. Norovírus: An Overview. *Rev Assoc Med Bras* 2011; **57**: 462–7.

16   Fisman D. Seasonality of viral infections: mechanisms and unknowns. *Clin Microbiol Infect* 2012; **18**: 946–54.

17   Hamborsky J, Kroger A, Wolfe S. Rotavirus. In: Epidemiology and Prevention of Vaccine-Preventable Diseases, 19th edn. Washington, DC: Public Health Foundation, 2015.

18   Payne DC, M. W, Parashar UD. Rotavirus. In: Manual for the surveillance of vaccine-preventable diseases. Atlanta, GA: Centers for Disease Control and Prevention, 2011.

19   Hervás D, Hervás-Masip J, Rosell A, Mena A, Pérez JL, Hervás JA. Are hospitalizations for rotavirus gastroenteritis associated with meteorologic factors? *Eur J Clin Microbiol Infect Dis* 2014; **33**: 1547–53.

20   Jagai JS, Sarkar R, Castronovo D, *et al.* Seasonality of rotavirus in South Asia: a meta-analysis approach assessing associations with temperature, precipitation, and vegetation index. *PLoS One* 2012; **7**: e38168.

21   Levy K, Hubbard AE, Eisenberg JNS. Seasonality of rotavirus disease in the tropics: a systematic review and meta-analysis. *Int J Epidemiol* 2009; **38**: 1487–96.

22   Rouhani S, Peñataro Yori P, Paredes Olortegui M, *et al.* Norovirus Infection and Acquired Immunity in 8 Countries: Results From the MAL-ED Study. *Clin Infect Dis* 2016; **62**: 1210–7.

23   Bernstein DI. Rotavirus overview. *Pediatr Infect Dis J* 2009; **28**: S50-3.

24   Richardson V, Hernandez-Pichardo J, Quintanar-Solares M, *et al.* Effect of rotavirus vaccination on death from childhood diarrhea in Mexico. *N Engl J Med* 2010; **362**: 299–305.

25   de Palma O, Cruz L, Ramos H, *et al.* Effectiveness of rotavirus vaccination against childhood diarrhoea in El Salvador: case-control study. *BMJ* 2010; **340**: c2825.

26   International Vaccine Action Center (IVAC). Vaccine introduction. ROTA Counc. 2017. http://rotacouncil.org/vaccine-introduction/ (accessed Dec 5, 2017).

27   GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet (London, England)* 2015; **385**: 117–71.

28   Murray PR, Rosenthal KS, Pfaller MA. Medical microbiology. Elsevier/Saunders, 2013.

29   Huang DB, Dupont HL. Enteroaggregative Escherichia coli: an emerging pathogen in children. *Semin Pediatr Infect Dis* 2004; **15**: 266–71.

30   Government of Canada PHA of C. Pathogen Safety Data Sheets and Risk Assessment - Public Health Agency of Canada. Lab. Biosaf. Biosecurity. 2016. http://www.phac-aspc.gc.ca/lab-bio/res/psds-ftss/index-eng.php (accessed Dec 2, 2016).

31   Naumova EN, Jagai JS, Matyas B, DeMaria A, MacNeill IB, Griffiths JK. Seasonality in six enterically transmitted diseases and ambient temperature. *Epidemiol Infect* 2007; **135**: 281–92.

32   Hellberg RS, Chu E. Effects of climate change on the persistence and dispersal of foodborne bacterial pathogens in the outdoor environment: A review. *Crit Rev Microbiol* 2015; **0**: 1–25.

33   Mookerjee S, Jaiswal A, Batabyal P, *et al.* Seasonal dynamics of Vibrio cholerae and its

phages in riverine ecosystem of Gangetic West Bengal: cholera paradigm. *Environ Monit Assess* 2014; **186**: 6241–50.

34   Janda JM, Abbott SL, McIver CJ. Plesiomonas shigelloides Revisited. *Clin Microbiol Rev* 2016; **29**: 349–74.

35   Centers for Disease Control and Prevention. DPDx - Laboratory Identification of Parasitic Diseases of Public Health Concern. 2016. https://www.cdc.gov/dpdx/intestinalAmebae/ (accessed Jan 1, 2017).

36   World Health Organization. Soil-transmitted helminth infections. Fact sheet. 2017. http://www.who.int/mediacentre/factsheets/fs366/en/ (accessed Feb 19, 2017).

37   Hotez PJ, Bundy DAP, Beegle K, *et al.* Helminth Infections: Soil-transmitted Helminth Infections and Schistosomiasis. In: Jamison DT, Breman JG, Measham AR, *et al.*, eds. 2nd edn. Washington (DC): World Bank, 2006.

38   Montresor A, Crompton DWT, Hall A, Bundy DAP, Savioli L. Guidelines for the Evaluation of Soil-Transmitted Helminthiasis and Schistosomiasis at Community Level. Geneva, 1998.

39   Christiansen CFF, Pedersen L, Sørensen HTT, Rothman KJJ. Methods to assess seasonal effects in epidemiological studies of infectious diseases--exemplified by application to the occurrence of meningococcal disease. *Clin Microbiol Infect* 2012; **18**: 963–9.

40   Naumova EN. Mystery of seasonality: getting the rhythm of nature. *J Public Health Policy* 2006; **27**: 2–12.

41   Wu X, Lu Y, Zhou S, Chen L, Xu B. Impact of climate change on human infectious diseases: Empirical evidence and human adaptation. *Environ Int* 2016; **86**: 14–23.

42   Grassly NC, Fraser C. Seasonal infectious disease epidemiology. *Proc Biol Sci* 2006; **273**: 2541–50.

43   Islam MS, Sharker MAY, Rheman S, *et al.* Effects of local climate variability on transmission dynamics of cholera in Matlab, Bangladesh. *Trans R Soc Trop Med Hyg* 2009; **103**: 1165–70.

44   Chan T-C, Fu Y-C, Hwang J-S. Changing social contact patterns under tropical weather conditions relevant for the spread of infectious diseases. *Epidemiol Infect* 2015; **143**: 440–51.

45   Yang Y, Guo C, Liu L, Zhang T, Liu W. Seasonality Impact on the Transmission Dynamics of Tuberculosis. *Comput Math Methods Med* 2016; **2016**: 8713924.

46   Lal A, Hales S, French N, Baker MG. Seasonality in human zoonotic enteric diseases: a systematic review. *PLoS One* 2012; **7**: e31883.

47   Varga C, Pearl DL, McEwen SA, Sargeant JM, Pollari F, Guerin MT. Incidence, distribution, seasonality, and demographic risk factors of Salmonella Enteritidis human infections in Ontario, Canada, 2007-2009. *BMC Infect Dis* 2013; **13**: 212.

48   Ahmed SM, Lopman BA, Levy K. A Systematic Review and Meta-Analysis of the Global Seasonality of Norovirus. *PLoS One* 2013; **8**: e75922.

49   Kovats RS, Edwards SJ, Charron D, *et al.* Climate variability and campylobacter infection: an international study. *Int J Biometeorol* 2005; **49**: 207–14.

50   Dorélien AM, Ballesteros S, Grenfell BTB, *et al.* Impact of Birth Seasonality on Dynamics of Acute Immunizing Infections in Sub-Saharan Africa. *PLoS One* 2013; **8**: e75806.

51   Neuzil KM, Kotloff KL. Community-acquired diarrhoea in a world with rotavirus vaccine: a glimpse into the future. *Lancet Glob Heal* 2015; **3**: e510–1.

52   Chang MR, Velapatiño G, Campos M, *et al.* Rotavirus seasonal distribution and prevalence before and after the introduction of rotavirus vaccine in a peri-urban community of Lima, Peru. *Am J Trop Med Hyg* 2015; **92**: 986–8.

53   Patel MM, Pitzer VE, Alonso WJ, *et al.* Global seasonality of rotavirus disease. *Pediatr Infect Dis J* 2013; **32**: e134-147.

54   Paul A, Gladstone BP, Mukhopadhya I, Kang G. Rotavirus infections in a community based cohort in Vellore, India. *Vaccine* 2014; **32 Suppl 1**: A49-54.

55   Rohayem J. Norovirus seasonality and the potential impact of climate change. *Clin Microbiol Infect* 2009; **15**: 524–7.

56   Sarkar R, Kang G, Naumova EN. Rotavirus seasonality and age effects in a birth cohort study of southern India. *PLoS One* 2013; **8**: e71616.

57   Rahman M, Sultana R, Ahmed G, *et al.* Prevalence of G2P[4] and G12P[6] rotavirus, Bangladesh. *Emerg Infect Dis* 2007; **13**: 18–24.

58   Colston JM, Ahmed AMS, Soofi SB, *et al.* Seasonality and within-subject clustering of rotavirus infections in an eight-site birth cohort study. *Epidemiol Infect* 2018; **146**. DOI:10.1017/S0950268818000304.

59   Hashizume M, Armstrong B, Wagatsuma Y, Faruque ASG, Hayashi T, Sack DA. Rotavirus infections and climate variability in Dhaka, Bangladesh: a time-series analysis. *Epidemiol Infect* 2008; **136**: 1281–9.

60   D'Souza RM, Hall G, Becker NG. Climatic factors associated with hospitalizations for rotavirus diarrhoea in children under 5 years of age. *Epidemiol Infect* 2008; **136**: 56–64.

61   Colas de la Noue A, Estienney M, Aho S, *et al.* Absolute Humidity Influences the Seasonal Persistence and Infectivity of Human Norovirus. *Appl Environ Microbiol* 2014; **80**: 7196–205.

62   Dey SK, Shimizu H, Phan TG, *et al.* Molecular epidemiology of adenovirus infection among infants and children with acute gastroenteritis in Dhaka City, Bangladesh. *Infect Genet Evol* 2009; **9**: 518–22.

63 Dey SK, Hoq I, Okitsu S, Hayakawa S, Ushijima H. Prevalence, seasonality, and peak age of infection of enteric adenoviruses in Japan, 1995-2009. *Epidemiol Infect* 2013; **141**: 958–60.

64 de Jong JC, Bijlsma K, Wermenbol AG, *et al.* Detection, typing, and subtyping of enteric adenoviruses 40 and 41 from fecal samples and observation of changing incidences of infections with these types and subtypes. *J Clin Microbiol* 1993; **31**: 1562–9.

65 Moyo SJ, Hanevik K, Blomberg B, *et al.* Prevalence and molecular characterisation of human adenovirus in diarrhoeic children in Tanzania; a case control study. *BMC Infect Dis* 2014; **14**: 666.

66 Soares CC, Volotão EM, Albuquerque MCM, *et al.* Prevalence of enteric adenoviruses among children with diarrhea in four Brazilian cities. *J Clin Virol* 2002; **23**: 171–7.

67 Mena KD, Gerba CP. Waterborne adenovirus. *Rev Environ Contam Toxicol* 2009; **198**: 133–67.

68 Strachan NJC, Rotariu O, Smith-Palmer A, *et al.* Identifying the seasonal origins of human campylobacteriosis. *Epidemiol Infect* 2013; **141**: 1267–75.

69 Louis VR, Gillespie IA, O'Brien SJ, Russek-Cohen E, Pearson AD, Colwell RR. Temperature-driven Campylobacter seasonality in England and Wales. *Appl Environ Microbiol* 2005; **71**: 85–92.

70 Kovats RS, Edwards SJ, Hajat S, Armstrong BG, Ebi KL, Menne B. The effect of temperature on food poisoning: a time-series analysis of salmonellosis in ten European countries. *Epidemiol Infect* 2004; **132**: 443–53.

71 Ravel A, Smolina E, Sargeant JM, *et al.* Seasonality in human salmonellosis: assessment of human activities and chicken contamination as driving factors. *Foodborne Pathog Dis* 2010; **7**: 785–94.

72 Philipsborn R, Ahmed SM, Brosi BJ, Levy K. Climatic drivers of diarrheagenic

Escherichia coli : A systematic review and meta-analysis. *J Infect Dis* 2016; : jiw081.

73   Freeman JT, Anderson DJ, Sexton DJ. Seasonal peaks in Escherichia coli infections: possible explanations and implications. *Clin Microbiol Infect* 2009; **15**: 951–3.

74   Emch M, Feldacker C, Islam MS, Ali M. Seasonality of cholera from 1974 to 2005: a review of global patterns. *Int J Health Geogr* 2008; **7**: 31.

75   Pascual M, Rodó X, Ellner SP, Colwell R, Bouma MJ. Cholera dynamics and El Niño-Southern Oscillation. *Science* 2000; **289**: 1766–9.

76   Moore SM, Azman AS, Zaitchik BF, *et al.* El Niño and the shifting geography of cholera in Africa. *Proc Natl Acad Sci U S A* 2017; : 201617218.

77   Joh RI, Hoekstra RM, Barzilay EJ, *et al.* Dynamics of shigellosis epidemics: estimating individual-level transmission and reporting rates from national epidemiologic data sets. *Am J Epidemiol* 2013; **178**: 1319–26.

78   Britton E, Hales S, Venugopal K, Baker MG. The impact of climate variability and change on cryptosporidiosis and giardiasis rates in New Zealand. *J Water Health* 2010; **8**: 561–71.

79   Jagai JS, Castronovo DA, Monchak J, Naumova EN. Seasonality of cryptosporidiosis: A meta-analysis approach. *Environ Res* 2009; **109**: 465–78.

80   Naous A, Naja Z, Zaatari N, Kamel R, Rajab M. Intestinal amebiasis: a concerning cause of acute gastroenteritis among hospitalized lebanese children. *N Am J Med Sci* 2013; **5**: 689–98.

81   Brooker S, Bethony J, Hotez PJ. Human hookworm infection in the 21st century. *Adv Parasitol* 2004; **58**: 197–288.

82   Chammartin F, Guimarães LH, Scholte RG, Bavia ME, Utzinger J, Vounatsou P. Spatio-temporal distribution of soil-transmitted helminth infections in Brazil. *Parasit Vectors* 2014; **7**: 440.

83   Centers for Disease Control and Prevention. Cryptosporidiosis. CDC A-Z Index. 2015. http://wwwnc.cdc.gov/travel/yellowbook/2016/infectious-diseases-related-to-travel/cryptosporidiosis (accessed Dec 11, 2015).

84   Centers for Disease Control and Prevention. Campylobacter. Food Saf. 2014. http://www.cdc.gov/foodsafety/diseases/campylobacter/index.html (accessed Aug 18, 2016).

85   Centers for Disease Control and Prevention. Salmonella. CDC A-Z Index. 2015. http://www.cdc.gov/salmonella/general/technical.html (accessed Dec 11, 2015).

86   Centers for Disease Control and Prevention. Giardia. CDC A-Z Index. 2015. http://www.cdc.gov/parasites/giardia/general-info.html (accessed Dec 11, 2015).

87   Centers for Disease Control and Prevention. Norovirus. CDC A-Z Index. 2015. http://wwwnc.cdc.gov/travel/yellowbook/2016/infectious-diseases-related-to-travel/norovirus (accessed Dec 11, 2015).

88   Centers for Disease Control and Prevention. Amebiasis. CDC A-Z Index. 2015. http://wwwnc.cdc.gov/travel/yellowbook/2016/infectious-diseases-related-to-travel/amebiasis (accessed Dec 11, 2015).

89   Centers for Disease Control and Prevention. E.coli (Echerichia coli). CDC A-Z Index. 2015. http://www.cdc.gov/ecoli/general/index.html (accessed Dec 11, 2015).

90   Vila J, Ruiz J, Gallardo F, *et al.* Aeromonas spp. and traveler's diarrhea: clinical features and antimicrobial resistance. *Emerg Infect Dis* 2003; **9**: 552–5.

91   World Health Organization. The global view of campylobacteriosis: report of an expert consultation. Geneva, 2014.

92   Rosner BM, Stark K, Werber D. Epidemiology of reported Yersinia enterocolitica infections in Germany, 2001-2008. *BMC Public Health* 2010; **10**: 337.

93   Pop-Jordanova N, Grigorova E. Influence of Climate Changes on Health (Review).

PRILOZI 2015; **36**: 119–25.

94   Guzman Herrador BR, de Blasio BF, MacDonald E, *et al.* Analytical studies assessing

the association between extreme precipitation or temperature and drinking water-

related waterborne infections: a review. *Environ Heal* 2015; **14**: 29.

95   Morand S, Owers KA, Waret-Szkuta A, McIntyre KM, Baylis M. Climate variability and

outbreaks of infectious diseases in Europe. *Sci Rep* 2013; **3**: 1774.

96   Checkley W, Epstein LD, Gilman RH, *et al.* Effects of EI Niño and ambient temperature

on hospital admissions for diarrhoeal diseases in Peruvian children. *Lancet* 2000;

355: 442–50.

97   McMichael AJ. Extreme weather events and infectious disease outbreaks. *Virulence*

2015; **6**: 543–7.

98   Intergovernmental Panel on Climate Change (IPCC) Working Group. Climate Change

2013: The Physical Science Basis. Stockholm, 2013.

99   Franchini M, Mannucci PM. Impact on human health of climate changes. *Eur J Intern*

Med 2015; **26**: 1–5.

100   Hofstra N. Quantifying the impact of climate change on enteric waterborne pathogen

concentrations in surface water. *Curr Opin Environ Sustain* 2011; **3**: 471–9.

101   Rodó X, Pascual M, Doblas-Reyes FJ, *et al.* Climate change and infectious diseases: Can

we meet the needs for better prediction? *Clim Change* 2013; **118**: 625–40.

102   Bouzid M, Colón-González FJ, Lung T, Lake IR, Hunter PR. Climate change and the

emergence of vector-borne diseases in Europe: case study of dengue fever. *BMC*

Public Health 2014; **14**: 781.

103   Wallace JM, Hobbs PV. Atmospheric Science, Second Edition: An Introductory Survey,

2 edition. Amsterdam; Boston: Academic Press, 2006.

104   Marshall J, R. Alan P. Atmosphere, Ocean and Climate Dynamics: An Introductory

Text, 1 edition. Amsterdam; Burlington, MA: Academic Press, 2007.

105   Levy K, Woster AP, Goldstein RS, Carlton EJ. Untangling the Impacts of Climate

        Change on Waterborne Diseases: a Systematic Review of Relationships between

        Diarrheal Diseases and Temperature, Rainfall, Flooding, and Drought. *Environ Sci*

        Technol 2016; **50**: 4905–22.

106   World Health Organization. Quantitative risk assessment of the effects of climate

        change on selected causes of death, 2030s and 2050s. 2014.

        http://www.who.int/globalchange/publications/quantitative-risk-assessment/en/

        (accessed Jan 1, 2017).

107   Kolstad EW, Johansson KA. Uncertainties Associated with Quantifying Climate

        Change Impacts on Human Health: A Case Study for Diarrhea. *Environ Health*

        Perspect 2011; **119**: 299–305.

108   Grace K, Davenport F, Hanson H, Funk C, Shukla S. Linking climate change and health

        outcomes: Examining the relationship between temperature, precipitation and birth

        weight in Africa. *Glob Environ Chang* 2015; **35**: 125–37.

109   MAL-ED Network Investigators, The MAL-ED Network Investigators, MAL-ED

        Network Investigators. The MAL-ED study: a multinational and multidisciplinary

        approach to understand the relationship between enteric pathogens, malnutrition,

        gut physiology, physical growth, cognitive development, and immune responses in

        infants and children up to 2 years of. *Clin Infect Dis* 2014; **59 Suppl 4**: S193-206.

110   Houpt E, Gratz J, Kosek M, *et al.* Microbiologic Methods Utilized in the MAL-ED Cohort

        Study. *Clin Infect Dis* 2014; **59**: S225–32.

111   Institute for Vetinary Public Health. Observed and Projected Climate Shifts 1901-

        2100 Depicted by World Maps of the Köppen-Geiger Climate Classification. World

        Maps Köppen-Geiger Clim. Cassification. 2011. http://koeppen-geiger.vu-

wien.ac.at/shifts.htm (accessed Aug 1, 2016).

112   MAL-ED. The Interactions of Malnutrition and Enteric Infections: Consequences for

      Child Health and Development. 2015. http://mal-ed.fnih.org/ (accessed Oct 25,

      2015).

113   Ahmed T, Mahfuz M, Islam MM, *et al.* The MAL-ED cohort study in Mirpur,

      Bangladesh. *Clin Infect Dis* 2014; **59 Suppl 4**: S280-6.

114   Bessong PO, Nyathi E, Mahopo TC, Netshandama V, MAL-ED South Africa.

      Development of the Dzimauli community in Vhembe District, Limpopo province of

      South Africa, for the MAL-ED cohort study. *Clin Infect Dis* 2014; **59 Suppl 4**: S317-24.

115   John SM, Thomas RJ, Kaki S, *et al.* Establishment of the MAL-ED birth cohort study

      site in Vellore, Southern India. *Clin Infect Dis* 2014; **59 Suppl 4**: S295-9.

116   Lima AAM, Oriá RB, Soares AM, *et al.* Geography, population, demography,

      socioeconomic, anthropometry, and environmental status in the MAL-ED cohort and

      case-control study Sites in Fortaleza, Ceará, Brazil. *Clin Infect Dis* 2014; **59 Suppl 4**:

      S287-94.

117   Mduma ER, Gratz J, Patil C, *et al.* The etiology, risk factors, and interactions of enteric

      infections and malnutrition and the consequences for child health and development

      study (MAL-ED): description of the Tanzanian site. *Clin Infect Dis* 2014; **59 Suppl 4**:

      S325-30.

118   Shrestha PS, Shrestha SK, Bodhidatta L, *et al.* Bhaktapur, Nepal: The MAL-ED Birth

      Cohort Study in Nepal. *Clin Infect Dis* 2014; **59**: S300–3.

119   Turab A, Soofi SB, Ahmed I, Bhatti Z, Zaidi AKM, Bhutta ZA. Demographic,

      socioeconomic, and health characteristics of the MAL-ED network study site in rural

      Pakistan. *Clin Infect Dis* 2014; **59 Suppl 4**: S304-9.

120   Yori PP, Lee G, Olortegui MP, *et al.* Santa Clara de Nanay: The MAL-ED Cohort in Peru.

Clin Infect Dis 2014; **59**: S310–6.

121   Kosek M, Guerrant RL, Kang G, *et al.* Assessment of environmental enteropathy in the

       MAL-ED cohort study: theoretical and analytic framework. *Clin Infect Dis* 2014; **59**

       Suppl 4: S239-47.

122   Richard S a, Barrett LJ, Guerrant RL, Checkley W, Miller M a. Disease Surveillance

       Methods Used in the 8-Site MAL-ED Cohort Study. *Clin Infect Dis* 2014; **59**: S220-224.

123   Liu J, Gratz J, Amour C, *et al.* Optimization of Quantitative PCR Methods for

       Enteropathogen Detection. *PLoS One* 2016; **11**: e0158199.

124   Barnett AG, Dobson AJ. Analysing Seasonal Health Data. Berlin, Heidelberg: Springer

       Berlin Heidelberg, 2010.

125   Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ* 2006;

       332: 1080–1080.

126   Simpson G. Modelling seasonal data with GAMs. From Bottom Heap. 2014.

       http://www.fromthebottomoftheheap.net/2014/05/09/modelling-seasonal-data-

       with-gam/ (accessed Oct 14, 2016).

127   Barnett AG, Baker P, Dobson AJ. Analysing Seasonal Data. *R J* 2012; **4**: 5–10.

128   Serfling RE. Methods for current statistical analysis of excess pneumonia-influenza

       deaths. *Public Health Rep* 1963; **78**: 494–506.

129   Naumova EN, MacNeill IB. Seasonality assessment for biosurveillance systems. In:

       Auget J-L, Balakrishnan N, Mesbah M, Molenberghs G, eds. Advances in Statistical

       Methods for the Health Sciences: aplications to cancer and AIDS studies, genome

       sequence analysis and survival analysis. Boston: Birkhauser, 2006.

130   Imai C, Armstrong B, Chalabi Z, Mangtani P, Hashizume M. Time series regression

       model for infectious disease and weather. *Environ Res* 2015; **142**: 319–27.

131   Atchison CJ, Tam CC, Hajat S, van Pelt W, Cowden JM, Lopman BA. Temperature-

dependent transmission of rotavirus in Great Britain and The Netherlands. *Proc Biol Sci* 2010; **277**: 933–42.

132   Barril PA, Fumian TM, Prez VE, *et al.* Rotavirus seasonality in urban sewage from Argentina: effect of meteorological variables on the viral load and the genetic diversity. *Environ Res* 2015; **138**: 409–15.

133   Sumi A, Rajendran K, Ramamurthy T, *et al.* Effect of temperature, relative humidity and rainfall on rotavirus infections in Kolkata, India. *Epidemiol Infect* 2013; **141**: 1652–61.

134   Norval M. The Effect of Ultraviolet Radiation on Human Viral Infections. *Photochem Photobiol* 2006; **82**: 1495.

135   Hurst CJ, Gerba CP, Cech I. Effects of environmental variables and soil characteristics on virus survival in soil. *Appl Environ Microbiol* 1980; **40**: 1067–79.

136   Rodell M, Houser PR, Jambor U, *et al.* The Global Land Data Assimilation System. *http://dx.doi.org/101175/BAMS-85-3-381* 2004.

137   Chen F, Mitchell K, Schaake J, *et al.* Modeling of land surface evaporation by four schemes and comparison with FIFE observations. *J Geophys Res Atmos* 1996; **101**: 7251–68.

138   Goddard Earth Sciences Data and Information Services Center (GES DISC), National Aeronautics and Space Administration (NASA). Documentation. Glob. L. Data Assim. Syst. 2016. http://disc.sci.gsfc.nasa.gov/hydrology/documentation/hydro_doc.shtml#NOAH (accessed Jan 7, 2016).

139   Kato H, Rodell M, Beyrich F, *et al.* Sensitivity of land surface simulations to model physics, land characteristics, and forcings, at four CEOP sites. 2007; **85 A**: 187–204.

140   Funk C, Peterson P, Landsfeld M, *et al.* The climate hazards infrared precipitation

with stations—a new environmental record for monitoring extremes. *Sci Data* 2015;
2: 150066.

141   Moe K, Shirley JA. The effects of relative humidity and temperature on the survival of
human rotavirus in faeces. *Arch Virol* 1982; **72**: 179–86.

142   Abad FX, Pintó RM, Bosch A. Survival of enteric viruses on environmental fomites.
Appl Environ Microbiol 1994; **60**: 3704–10.

143   Mbithi JN, Springthorpe VS, Sattar SA. Effect of relative humidity and air temperature
on survival of hepatitis A virus on environmental surfaces. *Appl Environ Microbiol*
1991; **57**: 1394–9.

144   Ji L, Senay GB, Verdin JP. Evaluation of the Global Land Data Assimilation System
(GLDAS) air temperature data products. 2015; **16**: 2463–80.

145   Lloyd S, Kovats R, Armstrong B. Global diarrhoea morbidity, weather and climate.
Clim Res 2007; **34**: 119–27.

146   Ceccherini G, Ameztoy I, Hernández CPR, Moreno CC. High-resolution precipitation
datasets in South America and West Africa based on satellite-derived rainfall,
enhanced vegetation index and digital elevation model. 2015; **7**: 6454–88.

147   Funk C, Husak G, Michaelson J, Shukla S, Hoell A. Attribution of 2012 and 2003-12
Rainfall Deficits in Eastern Kenya and Southern Somalia. *Bull Am Meteorol Soc* 2013;
94: S45–8.

148   Shukla S, McNally A, Husak G, Funk C. A seasonal agricultural drought forecast
system for food-insecure regions of East Africa. *Hydrol Earth Syst Sci* 2014; **18**: 3907–
21.

149   Toté C, Patricio D, Boogaard H, van der Wijngaart R, Tarnavsky E, Funk C. Evaluation
of satellite rainfall estimates for drought and flood monitoring in Mozambique. 2015;
7: 1758–76.

150 Katsanos D, Retalis A, Michaelides S. Validation of a high-resolution precipitation database (CHIRPS) over Cyprus for a 30-year period. 2016; **169**: 459–64.

151 Sattar SA, Ijaz MK, Johnson-Lussenburg CM, Springthorpe VS. Effect of relative humidity on the airborne survival of rotavirus SA11. *Appl Environ Microbiol* 1984; **47**: 879–81.

152 Girard M, Ngazoa S, Mattison K, Jean J. Attachment of noroviruses to stainless steel and their inactivation, using household disinfectants. *J Food Prot* 2010; **73**: 400–4.

153 HEMMES JH, WINKLER KC, KOOL SM. Virus survival as a seasonal factor in influenza and poliomylitis. *Antonie Van Leeuwenhoek* 1962; **28**: 221–33.

154 Dixon GJ, Sidwell RW, McNeil E. Quantitative studies on fabrics as disseminators of viruses. II. Persistence of poliomyelitis virus on cotton and wool fabrics. *Appl Microbiol* 1966; **14**: 183–8.

155 Mahl MC, Sadler C. Virus survival on inanimate surfaces. *Can J Microbiol* 1975; **21**: 819–23.

156 Fujita Y. [Rotavirus infection--clinical symptoms and influence of climate]. *Kansenshōgaku zasshi J Japanese Assoc Infect Dis* 1990; **64**: 1255–63.

157 Haffejee IE, Moosa A. Rotavirus studies in Indian (Asian) South African infants with acute gastro-enteritis: I. Microbiological and epidemiological aspects. *Ann Trop Paediatr* 1990; **10**: 165–72.

158 Mitui MT, Chan PKS, Nelson EAS, Leung TF, Nishizono A, Ahmed K. Co-dominance of G1 and emerging G3 rotaviruses in Hong Kong: A three-year surveillance in three major hospitals. *J Clin Virol* 2011; **50**: 325–33.

159 Gomwalk NE, Gosham LT, Umoh UJ. Rotavirus gastroenteritis in pediatric diarrhoea in Jos, Nigeria. *J Trop Pediatr* 1990; **36**: 52–5.

160 Lopman B, Armstrong B, Atchison C, Gray JJ. Host, weather and virological factors

drive norovirus epidemiology: time-series analysis of laboratory surveillance data in England and Wales. *PLoS One* 2009; **4**: e6671.

161   Santos RAT, Borges AMT, da Costa PSS, *et al.* Astrovirus infection in children living in the Central West region of Brazil. *Memórias do Inst Oswaldo Cruz* 2007; **102**: 209–13.

162   Qi W, Zhang C, Fu G, Zhou H. Global Land Data Assimilation System data assessment using a distributed biosphere hydrological model. 2015; **528**: 652–67.

163   Ijaz MK, Sattar SA, Johnson-Lussenburg CM, Springthorpe VS, Nair RC. Effect of relative humidity, atmospheric temperature, and suspending medium on the airborne survival of human rotavirus. *Can J Microbiol* 1985; **31**. DOI:10.1139/m85-129.

164   Nuanualsuwan S, Mariam T, Himathongkham S, Cliver DO. Ultraviolet inactivation of feline calicivirus, human enteric viruses and coliphages. *Photochem Photobiol* 2002; **76**: 406–10.

165   Kim D, Lim Y-J, Kang M, Choi M. Land response to atmosphere at different resolutions in the common land model over East Asia. *Adv Atmos Sci* 2016; **33**: 391–408.

166   Chen Y, Yang K, Qin J, Zhao L, Tang W, Han M. Evaluation of AMSR-E retrievals and GLDAS simulations against observations of a soil moisture network on the central Tibetan Plateau. 2013; **118**: 4466–75.

167   Zaitchik BF, Rodell M, Olivera F. Evaluation of the Global Land Data Assimilation System using global river discharge data and a source-to-sink routing scheme. 2010; **46**. DOI:10.1029/2009WR007811.

168   National Oceanic and Atmospheric Administration. NNDC Climate Data Online. NOAA Satallite Inf. Serv. 2016. https://www7.ncdc.noaa.gov/CDO/cdoselect.cmd?datasetabbv=GSOD&countryabbv=&georegionabbv= (accessed Jan 1, 2017).

169   National Climatic Data Center - Climate Services Branch. Federal Climate Complex -
       Global Surface Summary of Day Data (version 7). 2006.
       https://www7.ncdc.noaa.gov/CDO/GSOD_DESC.txt (accessed Jan 1, 2016).

170   World Meteorological Organization. Technical Regulations: Basic Documents No. 2,
       Volume I - General Meteorological Standards and Recommended Practices. 2015.

171   Liu L, Oza S, Hogan D, *et al.* Global, regional, and national causes of under-5 mortality
       in 2000–15: an updated systematic analysis with implications for the Sustainable
       Development Goals. *Lancet* 2016; **388**: 3027–35.

172   Cook SM, Glass RI, LeBaron CW, Ho MS. Global seasonality of rotavirus infections. *Bull
       World Health Organ* 1990; **68**: 171–7.

173   Stolwijk AM, Straatman H, Zielhuis GA. Studying seasonality by using sine and cosine
       functions in regression analysis. *J Epidemiol Community Health* 1999; **53**: 235–8.

174   Royston P, Sauerbrei W. Multivariable modeling with cubic regression splines: A
       principled approach. *Stata J* 2007; **7**: 45–70.

175   Mohan VR, Karthikeyan R, Babji S, *et al.* Rotavirus Infection and Disease in a Multisite
       Birth Cohort: Results From the MAL-ED Study. *J Infect Dis* 2017; **92**: 680–5.

176   Linhares AC, Lanata CF, Hausdorff WP, Gabbay YB, Black RE. Reappraisal of the
       Peruvian and Brazilian lower titer tetravalent rhesus-human reassortant rotavirus
       vaccine efficacy trials: analysis by severity of diarrhea. *Pediatr Infect Dis J* 1999; **18**:
       1001–6.

177   StataCorp. Stata Statistical Software: Release 13. 2013.

178   Nathanson N, Kew OM. From Emergence to Eradication: The Epidemiology of
       Poliomyelitis Deconstructed. *Am J Epidemiol* 2010; **172**: 1213–29.

179   Török TJ, Kilgore PE, Clarke MJ, Holman RC, Bresee JS, Glass RI. Visualizing
       geographic and temporal trends in rotavirus activity in the United States, 1991 to

1996. National Respiratory and Enteric Virus Surveillance System Collaborating Laboratories. *Pediatr Infect Dis J* 1997; **16**: 941–6.

180   Fann N, Brennan T, Dolwick P, *et al.* Air Quality Impacts. In: Program GCR, ed. The Impacts of Climate Change on Human Health in the United States: A Scientific Assessment. Washington, DC: Global Change Research Program, 2016: 69–98.

181   Xu Z, Etzel RA, Su H, Huang C, Guo Y, Tong S. Impact of ambient temperature on children's health: A systematic review. *Environ Res* 2012; **117**: 120–31.

182   Grace K. Considering climate in studies of fertility and reproductive health in poor countries. *Nat Clim Chang* 2017; **7**: 479–85.

183   Fang H, Beaudoing HK, Rodell M, Teng WL, Vollmer BE. Global Land Data Assimilation System (GLDAS) products, services and application from NASA Hydrology Data and Information Services Center (HDISC). 2009: 151–9.

184   Hamm NAS, Soares Magalhães RJ, Clements ACA. Earth Observation, Spatial Data Quality, and Neglected Tropical Diseases. *PLoS Negl Trop Dis* 2015; **9**: e0004164.

185   Kirkwood B, Sterne J. Essentials of Medical Statistics, 2 edition. Malden, Mass: Wiley-Blackwell, 2001.

186   Nash JE, Sutcliffe JV. River flow forecasting through conceptual models part I — A discussion of principles. *J Hydrol* 1970; **10**: 282–90.

187   Wang F, Wang L, Koike T, *et al.* Evaluation and application of a fine-resolution global data set in a semiarid mesoscale river basin with a distributed biosphere hydrological model. 2011; **116**. DOI:10.1029/2011JD015990.

188   Ahmed M, Sultan M, Yan E, Wahr J. Assessing and Improving Land Surface Model Outputs Over Africa Using GRACE, Field, and Remote Sensing Data. 2016; : 1–28.

189   de Oliveira G, Brunsell N, Moraes E, *et al.* Use of MODIS Sensor Images Combined with Reanalysis Products to Retrieve Net Radiation in Amazonia. *Sensors* 2016; **16**:

956.

190 Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. *Int J Forecast* 2006; **22**: 679–88.

191 Lee RM, Lessler J, Lee RA, *et al.* Incubation periods of viral gastroenteritis: a systematic review. *BMC Infect Dis* 2013; **13**: 446.

192 López-Carr D, Pricope NG, Aukema JE, *et al.* A spatial analysis of population dynamics and climate change in Africa: potential vulnerability hot spots emerge where precipitation declines and demographic pressures coincide. *Popul Environ* 2014; **35**: 323–39.

193 Gasparrini A. Modeling exposure-lag-response associations with distributed lag non-linear models. *Stat Med* 2014; **33**: 881–99.

194 Gomwalk NE, Umoh UJ, Gosham LT, Ahmad AA. Influence of climatic factors on rotavirus infection among children with acute gastroenteritis in Zaria, northern Nigeria. *J Trop Pediatr* 1993; **39**: 293–7.

195 Colston JM, Ahmed T, Mahopo C, *et al.* Evaluating meteorological data from weather stations, and from satellites and global models for a multi-site epidemiological study. *Environ Res* 2018; **165**: 91–109.

196 McCormick BJJ, Lee GO, Seidman JC, *et al.* Dynamics and Trends in Fecal Biomarkers of Gut Function in Children from 1-24 Months in the MAL-ED Study. *Am J Trop Med Hyg* 2016; : 16-0496.

197 Tjur T. Coefficients of Determination in Logistic Regression Models—A New Proposal: The Coefficient of Discrimination. *Am Stat* 2009; **63**: 366–72.

198 Hashizume M, Armstrong B, Hajat S, *et al.* Association between climate variability and hospital visits for non-cholera diarrhoea in Bangladesh: effects and vulnerable groups. *Int J Epidemiol* 2007; **36**: 1030–7.

199   Park SH, Kim EJ, Yun TH, *et al.* Human Enteric Viruses in Groundwater. *Food Environ Virol* 2010; **2**: 69–73.

200   Dennehy PH. Transmission of rotavirus and other enteric pathogens in the home. *Pediatr Infect Dis J* 2000; **19**: S103-5.

201   Prince DS, Astry C, Vonderfecht S, Jakab G, Shen F-M, Yolken RH. Aerosol transmission of experimental rotavirus infection. *Pediatr Infect Dis J* 1986; **5**: 218–22.

202   Jones RM, Brosseau LM. Aerosol Transmission of Infectious Disease. *J Occup Environ Med* 2015; **57**: 501–8.

203   Fernstrom A, Goldblatt M. Aerobiology and Its Role in the Transmission of Infectious Diseases. *J Pathog* 2013; **2013**: 1–13.

204   Gralton J, Tovey E, McLaws M-L, Rawlinson WD. The role of particle size in aerosolised pathogen transmission: A review. *J Infect* 2011; **62**: 1–13.

205   Arundel A V, Sterling EM, Biggin JH, Sterling TD. Indirect health effects of relative humidity in indoor environments. *Environ Health Perspect* 1986; **65**: 351–61.

206   Payne DC, Currier RL, Staat MA, *et al.* Epidemiologic Association Between FUT2 Secretor Status and Severe Rotavirus Gastroenteritis in Children in the United States. *JAMA Pediatr* 2015; **169**: 1040–5.

207   Altizer S, Dobson A, Hosseini P, Hudson P, Pascual M, Rohani P. Seasonality and the dynamics of infectious diseases. *Ecol Lett* 2006; **9**: 467–84.

208   Badr HS, Zaitchik BF, Dezfuli AK. Climate Regionalization through Hierarchical Clustering: Options and Recommendations for Africa. *Am Geophys Union, Fall Meet 2014, Abstr #GC53B-0538* 2014.

209   Vaisala. Vaisala HUMICAP HUmidity and Temperature Probes HMP45A/D. Helsinki, 2006.

Curriculum Vitae

Josh Michael Colston
Curriculum Vitae - May 21, 2018

The Johns Hopkins Bloomberg School of Public Health
615 N. Wolfe Street
Baltimore, MD 21205
josh.colston@jhu.edu
Born February 9th, 1981, Redwood City, California, USA

604 Columbia Road NW, Unit 1
Washington D.C., U.S., 20001
202-679-7723
joshcolston@gmail.com

EDUCATION

2014 - Present
Baltimore, MD

The Johns Hopkins Bloomberg School of Public Health (JHSPH)
Ph.D. in Global Disease Epidemiology and Control

Thesis applies advanced statistical methods to climate variables and biomarkers of environmental enteropathy and immune function to assess their potential in predicting statural growth faltering, enteric infections and other morbidities in infants. Completed required coursework and comprehensive and oral examinations. Anticipated date of completion May 2018. Adviser: Prof. Margaret Kosek.

2008
London, U.K.

The London School of Hygiene & Tropical Medicine (LSHTM)
M.Sc. in Demography and Health

2004
London, U.K.

King's College London
M.A. in European Studies

2002
Birmingham, U.K.

The University of Birmingham
B.A. in Philosophy

PUBLICATIONS

Colston JM, *et al.* 2018 "Evaluating meteorological data from weather stations, and from satellites and global models for a multi-site epidemiological study", *Environ. Res.* 165:91–109.

Colston, J. *et al.* 2018 "Seasonality and within-subject clustering of rotavirus infections in an eight-site birth cohort study", *Epidemiology & Infection*.

Exum NG, …**Colston J**… *et al.* 2018 "A Longitudinal Study of Household Water, Sanitation, and Hygiene Characteristics and Environmental Enteropathy Markers in Children Less than 24 Months in Iquitos, Peru", *American Journal of Tropical Medicine and Hygiene.*

Colston, J. *et al.* 2017. "A methodological framework for modeling and assessing biomarkers of environmental enteropathy as predictors of growth in infants - an example from a Peruvian birth cohort", *American Journal of Clinical Nutrition*, American Society for Nutrition; ajcn151886.

Wang, W., M. Winner, C. Burgert, and **J. Colston**. 2014. "Influence of Service Readiness on Use of Facility Delivery Care: A Study Linking Health Facility Data and Population Data in Haiti". 114. DHS Working Papers. Rockville, Maryland: The DHS Program, ICF International.

Colston, J., and C. Burgert. 2014. "Using Geospatial Analysis to Inform Decision Making in Targeting Health Facility-based Programs: A Guidance Document". MEASURE Evaluation.

Colston, J and Saboya, M, I. 2013. "Soil-transmitted helminthiasis in Latin America and the Caribbean: modelling the determinants, prevalence, population at risk, and costs of control at sub-national level", *Geospatial Health* - Volume 7, Number 2, May 2013, Pages 321-340

Colston, J., 2012. The Neglected Tropical Diseases (NTD) initiative for Latin America and the Caribbean of the Inter-American Development Bank and the role of geospatial analysis in health programmes. *Geospatial Health* - 6, S11-14. doi:10.4081/gh.2012.117

AWARDS AND HONORS

2017 - Present	**Group on Earth Observations Work Programme**	
	National Aeronautics and Space Administration	
2015 - 2017	**Water, Climate and Health Integrative Graduate Education and Research Traineeship (IGERT)**	
	The National Science Foundation and Johns Hopkins University	
2014	**International Health Department Scholarship**	
	Johns Hopkins Bloomberg School of Public Health	

INVITED TALKS

2017 Baltimore, MD	**GnosisGIS 2017 11th International Symposium on Geospatial Health** Evaluating meteorological data from weather stations, and from satellites and global models for a multi-site epidemiological study	
2017 Boston, MA	**Tufts University School of Medicine, Global Health Seminar Series** Assessing pathogen-specific seasonality of pediatric enteric infections in an eight-site cohort study	

CAMPUS TALKS

2016 Baltimore, MD	**IGERT External Knowledge Transfer Board Annual Meeting** Climatic determinants of enteric disease seasonality	
2015 Baltimore, MD	**International Health Student Group Journal Club** Soil-transmitted helminthiasis in Latin America and the Caribbean	

TEACHING ASSISTANT EXPERIENCE

2016 - 2017	**Foundations of International Health (on-site and online offerings)** International Health Department, JHSPH	
2016	**Global Disease Control Priorities and Programs (online)** International Health Department, JHSPH	
2015 - Present	**Global Disease Epidemiology and Control MSPH Program Seminar** International Health Department, JHSPH	

COMMUNITY OUTREACH

2016 - 2017	**STEM Achievement in Baltimore Elementary Schools (SABES)**	

Baltimore, MD   After-school program mentor

PROFESSIONAL EXPERIENCE

2015 - Present   The Johns Hopkins Bloomberg School of Public Health
Baltimore, MD   Research Analyst

Analyze and manage longitudinal data from the Interactions of Malnutrition & Enteric Infections: Consequences for Child Health and Development (MAL-ED) project

Analyzed Demographic and Health Surveys data on reproductive health indicators for situational analyses of Guinea and the Democratic Republic of Congo

Conducted a geospatial study to estimate population at risk of insecticide-treated net misuse in nine malaria-endemic countries

2013 - 2014   The Demographic and Health Surveys Program, ICF International
Rockville, MD   Associate Health GIS Analyst

Used advanced geospatial methods to provide data validation support to household survey projects and to contribute to data analyses; writing reports and papers for peer review.

Contributed to numerous publications; wrote a guidance document on how to quantify target populations for facility-based interventions using advanced geospatial techniques.

Organized several expert workshops on applying geospatial methods to household surveys.

2009 - 2012   Inter-American Development Bank
Washington, DC   Research Fellow in Epidemiology and Public Health
Served as technical advisor to the Neglected Tropical Diseases (NTDs) Initiative for Latin America and the Caribbean and the *Salud Mesoamérica* 2015 Initiative.

Led an epidemiological study to model and map the burden and costs of controlling Soil-Transmitted Helminthiasis in the LAC region; published in a peer-reviewed journal.

Coordinated an inter-institutional working group to develop guidelines and define indicators for monitoring and evaluating NTD projects.

2008   Options UK Public Health Consultancy Firm
London, UK   Research Intern

2007   Action Against Hunger UK
London, UK   Operations intern

2006   LEPRA Society India
Hyderabad, India   Programs intern

2005 – 2006   Various educational institutions
Beijing, China   Teacher, lecturer and assistant manager

OTHER TRAINING AND CERTIFICATIONS

2015   Certificate in Clinical Vaccine Trials and Good Clinical Practice
Online   Intensive eight-week distance education course, JHSPH

2015   Basic Human Subjects Research Course
Online   Collaborative Institutional Training Initiative (CITI Program)

2014   Responsible Conduct of Research

Baltimore, MD	Certificate from Johns Hopkins School of Public Health
2010	**Mapping and Modelling for Neglected Diseases in Latin America**
Washington, DC	Workshop at the Pan American Health Organization
2009 - 2011	**Inter-American Development Bank, Knowledge and Learning Sector**
Washington, DC	Various courses in health program planning, monitoring and financing

LANGUAGES

English	Native
Spanish	Full proficiency
Chinese	Basic proficiency
Portuguese	Basic proficiency
French	Basic proficiency
German	Basic proficiency

PROFESSIONAL MEMBERSHIPS AND AFFILIATIONS

2016 - Present   **American Society of Tropical Medicine and Hygiene**
Member

2016 - Present   **Royal Society of Tropical Medicine and Hygiene**
Member

2012 - Present   **Gnosis GIS – Global Network for Geospatial Health**
Member

2011 - 2014   **LSHTM Alumni Network, Washington DC Chapter**
Board member