Part 2 Questions & Answers Session A

Please type your questions in the Question Box. We will try our best to get to all your questions. If we don't, feel free to email Amita Mehta (amita.v.mehta@nasa.gov), Dr. Antarpreet Jutla (ajutla@ufl.edu), or Bailey Magers (bmagers@ufl.edu).

Question 1: Are PACE data Available worldwide?
Answer 1: Yes, but many datasets are still emerging. Right now, datasets for chlorophyll-a are available at a 4 km spatial resolution. In the future, algorithms will be improved to provide greater spatial resolution (I believe up to 300 m) and include data on different microorganism concentrations, including phytoplankton and primary productivity.

Question 2: How can we apply this to mosquito-borne diseases?
Answer 2: Many mosquito- and vector-borne diseases are climate-sensitive diseases, meaning that the distribution and emergence of these diseases is heavily influenced by climate factors, just like cholera. For cholera, the main drivers of our model are temperature and precipitation. These parameters (as well as others, like humidity) are also drivers of vector-borne diseases, including West Nile virus, Zika, malaria, dengue, and many more. Higher temperatures and stagnant waters typically benefit mosquito breeding, helping us to identify periods when disease vectors may be more prevalent. Prediction modeling for West Nile virus in the US is being done by several researchers, and significant progress is being made.

Question 3: Is it possible to use PACE data, which is based on algorithms designed for the ocean, for inland waters like Lake Okeechobee?
Answer 3: At this moment, I am not sure if it is possible due to the current spatial resolution. Lake Okeechobee will only have one PACE pixel so it is not usable. However, in the future when algorithms are improved and spatial resolution is maximized, datasets should include some larger inland water bodies and estuaries.

Question 4: Can we reciprocate the datasets used for cholera to other waterborne diseases like diarrhea or dysentery?
Answer 4: Potentially. The datasets can absolutely be used to predict the emergence and presence of other waterborne diseases. However, each pathogen and disease are

unique. The exact parameters and thresholds we use for our cholera model will likely not have as much success for other diseases. Therefore, it would be best for research on specific diseases to guide dataset use on them for prediction.

Question 5: Can I use a water scarcity map, adding LST data with precipitation data, to develop a waterborne disease risk map?
Answer 5: There is potential to accomplish this, but general waterborne disease risk can be tricky. For instance, *Cryptosporidium* and *Giardia*, which are protozoa, exhibit different climatic sensitivities than *Vibrio spp*. or *Salmonella*, which are bacteria. Grouping together large groups of diseases to predict general risk can be misleading. However, there may be potential to use climate variables to identify regions most at risk of water scarcity, almost like a social vulnerability index for water insecurity.

Question 6: How do you make this water scarcity map?
Answer 6: This requires assessment of water availability (supply) and water demands at a watershed level. There is Water Risk Atlas available from Aqueduct (World Resources Institute).

Question 7: The data proves precipitation is a mediator to Vibrios bacteria, but then, is it also swayed by wind in high temperature, or summer?
Answer 7: We have not necessarily found a direct relationship between wind speed and vibrio presence. One of the main reasons we do not include wind is that it is a cofounder for other variables. However, cholera and proliferation of other *Vibrio spp*. is highly seasonal, with most cases occurring in the summer months (typically late summer after bacteria have had a chance to proliferate).

Question 8: The threshold of cholera, can I use the same one for any waterborne disease?
Answer 8: Most likely no. Each pathogen behaves differently under different environmental stressors. While many waterborne diseases have general correlation with similar environmental parameters, thresholds of temperature and other environmental factors differ.

Question 9: Do you conclude that the shift of *vibrio* North and increase in *vibrio* cases in the Southern US are due to climate impacts?
Answer 9: We are currently studying the impact of environmental factors on vibriosis cases along the east coast of the US, including the southeast. Generally, vibriosis

cases are significantly associated with several parameters that are environment-sensitive. As vibriosis distributions have shifted, there are several explanations. One is, of course, fluctuations in environmental conditions, including inter-annual variability from events like ENSO. However, other factors may contribute as well, including increasing trade and distribution of seafood. Increased surveillance of vibrios may also be a factor, but this is less likely due to mandatory reporting of cases by healthcare providers since ~1990.

Question 10: Is there a sustainable plan to mitigate the risk of cholera before 2030?

Answer 10: This depends on the definition of mitigation. Will we essentially eliminate cholera, as we have in the US, globally by 2030? No. I do not think that is realistic. This is also because vibrios are always present in the water systems. However, is there a path to distributing a cholera decision support system to countries afflicted with cholera, which would help with policymakers in planning mitigation? **I would say yes**. Much of this depends on increased global collaboration and partnerships abroad.

Question 11: For the study done in Yemen, would the 3-year period be sufficient to validate the association between climate and cholera? Just asking this as for other vector-borne infections, it generally uses 10-15 years of data. It would be helpful to understand an existing rule of thumb for the typical length of the data periods required for diarrheal infections like cholera.

Answer 11: The difficulty with cholera and many other diarrheal infections is limited availability of data. For instance, in Yemen, the weekly data we received at a county-scale is almost unheard of until recently. Many other countries do not have these datasets, and if they do, they are considered sensitive to national security and cannot be shared. Also, the outbreak of cholera in Yemen is approaching 10 years, meaning we do not even have the amount of data used in vector-borne studies yet. At this stage, we are working with what data is available. However, while most studies cover smaller periods than other infectious disease studies, the global coverage and diversity of studies coming to similar conclusions provides sufficient evidence that cholera is a climate-sensitive disease.

Also, point to note that we have used data from several other parts of the world where we have tested the climate-microbiological hypothesis.

Question 12: What is the correlation between the presence of water hyacinth and waterborne diseases? What are the common models used to analyze their relationship?

Answer 12: I am not well-informed on research about water hyacinth and waterborne disease. However, the optimal growth conditions of water hyacinth, including thresholds for salinity and temperature, are similar to some waterborne diseases, so there is absolutely potential for this to be explored.

Question 13: What is the reason for the strong cholera presence in Guinea/Sierra Leone/Liberia?

Answer 13: There can be a few reasons for the strong presence of cholera. One is that temperatures in that region are quite high nearly all year round. Abnormally high temperature in some months may lead to water shortages and poor WASH conditions. Assessing environmental, WASH, and social conditions would help determine exact reasons for high cholera presence in Guinea/Sierra Leon/Liberia.

Question 14: *Vibrio cholera* affects only humans or all mammals?

Answer 14: To my knowledge, *Vibrio cholera* is not pathogenic to other mammals. However, vibrios are zoonotic pathogens, meaning there are animal species, mostly shellfish, that host the bacteria, which humans can consume and can cause illness.

Question 15: Any Cholera cases reported in Arctic regions and Antarctica?

Answer 15: We are not aware of this and have not conducted any experiments. Vibrios are autochthonous, so it is likely that the pathogens will be present in the water system. Vibrios also hibernate, meaning at certain conditions, they may not be viable but still present in the waters.

Question 16: How can we assess skin disease due to land surface temperature? Which observatory can we use for this?

Answer 16: I believe that you are referring to *V. vulnificus*… LST is not directly related to skin disease. LST serves as the proxy variables for cholera and represents the warming conditions for water.

Question 17: If there is a PACE pixel available for an inland water body like the Great Lakes, is it scientifically appropriate to use that data—considering that the algorithms are designed specifically for ocean conditions?

Answer 17: PACE data are likely to be available for the Great Lakes, and should be research ready at some point of time.

Part 2 Questions & Answers Session B

Please type your questions in the Question Box. We will try our best to get to all your questions. If we don't, feel free to email Amita Mehta (amita.v.mehta@nasa.gov), Dr. Antarpreet Jutla (ajutla@ufl.edu), or Bailey Magers (bmagers@ufl.edu).

Question 1: Is it correct that the Vibrio spp. outbreaks are mainly close to the coastal areas? Therefore, freshwater systems near the coast are more likely to have these Vibrio spp. than in the deeper mainland?

Answer 1: Yes and no. Cholera, caused by the naturally occurring bacteria Vibrio cholerae, often thrives in coastal waters, making coastal regions prone to outbreaks. Diseases like cholera remain endemic in various coastal regions, such as Bangladesh and Mozambique. However, *Vibrio cholerae* can also be found in inland water bodies and survive adequately in terrestrial waters, leading to large epidemics, such as those observed in Malawi and Yemen.

Other infections, like septicemia, can occur through direct contact with water containing pathogenic *Vibrio spp*., typically found in brackish waters or seawater with high temperatures and sufficient microbiological communities. These infections are almost exclusively observed in coastal areas.

Freshwater bodies, including rivers and lakes, have varying salinity profiles, but not much is known about the distribution of vibrios in riverine systems. Another mode of infection is through the consumption of raw or undercooked seafood contaminated with the bacteria. This can happen almost anywhere, especially with the global trade of seafood becoming more common. However, these infections are still usually more concentrated in coastal communities, likely due to dietary and cultural preferences. It should be noted that most foodborne infections are not as severe as those acquired through direct contact, so foodborne infections are often underreported.

Question 2: Is sanitation ranked?

Answer 2: Please refer to Environmental Performance Index. Also, some information can be found at this UNICEF data website. Many countries have internal documentation of regional WASH infrastructure vulnerabilities or WASH indexes, but many of these datasets are considered sensitive and not shared with the public.

Question 3: Is the model you generated a raster-based model? And is it available for other researchers to use?
Answer 3: The model is raster-based, and can be recreated for other diseases quite easily. The model is patent-pending right now, and we are unable to share the model. However, the underlying hypothesis and methods are explained in detail in many of our publications, if any are interested.

Question 4: Is there such a thing as a legal threshold for these risk maps?
Answer 4: Not necessarily. The thresholds we have defined as high, medium, and low risk are based on our previous efforts and research in other countries. For some regions, it may be possible that these thresholds do not hold as true, and so, in those cases, we might need to adjust the thresholds. We do not want to strictly stick to any threshold of risk, as each region and outbreak might present unique and different challenges.

Question 5: Which collaboration institutions are you working with in Malawi?
Answer 5: We have a consortium of collaborators all over the world. This includes several academic institutions in the US (such as the University of Maryland, University of Rhode Island, University of Alabama) which handles the bulk of our microbiological, communication and app development work. For certain regions, we generally collaborate through partners such as USAID, US State Department, FCDO, UN and NGOs who connect us with regional health ministries. Contact us for more details for Malawi (cholerapredict ion_users@lists.ufl.edu).

Question 6: South Sudan is a prone area for cholera outbreak in some of the state. Do you have any information for South Sudan in particular that can help to predict the risks and mitigate the impact?
Answer 6: We are currently investigating the environmental conditions leading to the ongoing outbreak in South Sudan. Preliminary results show that drought may have been a significant factor in disease emergence, but heavy rainfalls may have contributed to the worst parts of the outbreak. Contact us for more details (cholerapredict ion_users@lists.ufl.edu).

Question 7: What are the wash issues? R.E., is there any use to offer infrastructure under X circumstances?
Answer 7: We primarily advocate for increased access to WASH infrastructure under all circumstances to prevent human contact with cholera and all other waterborne/foodborne pathogens in water and wastewater supplies. However, in our reports to governments or health agencies or NGOs, we typically stress certain times when increased supplies should be distributed (chlorination tablets, for example), which would be when model risk is highest in the region.

Question 8: I am a Python developer and working with machine learning. Which technologies did you use to develop the UFWRA model?
Answer 8: This model is mostly simple and was developed using only ArcGIS software and toolboxes. However, python has been used to automate the process for data extraction, data processing, and computation to quickly generate the raster files. We do have a Cholera Decision Support System in making, Contact us for further details.

Question 9: Is there some relationship between cholera outbreak and heavy metal pollution in water? Perhaps when analyzing mining operations and their environmental impact.
Answer 9: Generally, *Vibrio cholerae* is a resistant bacteria, surviving even in waters with heavy metal pollution. As far as I am aware, the metals do not contribute to their presence, but pollution may be an indicator of water quality, which would indicate *Vibrio cholerae* presence.

Question 10: What validation techniques do you use to validate your model predictions? And are there any statistical analyses that you use or generate from your model predictions for the significance of your results?
Answer 10: We use typical model performance statistics, including accuracy, sensitivity, precision, F1 score, and AUC to evaluate the success of the model in the Yemen study. The testing dataset was the 2017-2019 data we were provided with.

Question 11: What are the best practices to transfer these models to other regions?
Answer 11: Generally, our model is transferable to all regions. If we notice one region behaves uniquely, we may have to recalibrate the model for that region, but it is quite

rare. As for transferring it for other diseases, it would depend on the environmental conditions that promote the proliferation of whatever pathogen we are considering.

Question 12: What would be the difference between modeling cholera using MGWR and your model?

Answer 12: Cholera modeling has a rich history, beginning with Codeco's foundational compartmental model in 2001. Since then, numerous approaches have been developed to model cholera time series. One notable method is Multiscale Geographically Weighted Regression (MGWR), which effectively establishes associations between time series data and environmental variables. However, MGWR still relies on cholera time series data to build the model and interpret coefficients contextually, which can limit its simplicity and generalizability.

Our technology offers a significant advantage in this regard. Once our model is calibrated and validated, it no longer requires cholera time series data to be applied to new regions. This is because the variable weights are automatically adjusted for different regions, allowing us to identify environmental factors that are likely to be critical in developing cholera risk. This capability enhances the model's parsimony and broadens its applicability, making it a powerful tool for cholera risk assessment across diverse geographical areas.