Part 1 Questions & Answers Session

Please type your questions in the Question Box. We will try our best to answer all your questions. If we don't, feel free to email Assaf Anyamba (assaf.anyamba@nasa.gov) or Tatiana Loboda (loboda@umd.edu).

Question 1: How does satellite imagery help us analyze the specific distribution of vectors?
Answer 1: What we are able to do with satellite observations is to characterize the availability and quality of the habitat, and from that, we infer the likelihood of vector presence and population dynamics. Ideally, to get more detailed information, you would pair your habitat availability and quality observations with field surveys for vector species and population dynamics. The actual specific species distribution may not be important - it is likely that all you need to know is, "are there vector species in this area that host plasmodium parasites?" If you know that in your region you have only forest-dwelling mosquitoes, then you could assume that if there are no forests around, these species are not present. Alternatively, if you know that you have a very large and diverse group of malaria-carrying vectors, you know you have to monitor forests, and grasslands, and shrubs, and cropped areas - everything. Vector density or population dynamics are not possible with satellite observations but field surveys can be combined and modeled.

Question 2: What kind of species distribution model was used while doing the Myanmar case study? Can you share a link to the study?
Answer 2: We did not focus on species distribution per se. What we established is that we have a lot of various mosquito species spanning all land covers, they are present and biting throughout the year, and we used this to inform our strategy for monitoring extent and frequency. The study was conducted in Vietnam.

Question 3: Can this approach (exposure, vulnerability and hazard) be applied to animal diseases like anthrax? If yes, what kind of datasets might be needed?
Answer 3: It most certainly can. I do not know enough about anthrax to give you the answer. But I very strongly urge you to think through first conceptually what metrics are for hazards, exposure, and vulnerability, and then look for proxies or direct observations from satellite data. Vectors are just one component.

Question 4: What are the possible household survey data that can be combined with remote sensing data to examine service delivery and access-related barriers?

Answer 4: In our work, we focused on understanding the exposure through household surveys rather than access to care. Complementing earth observations with local or regional surveys will monumentally improve both your ability to better conceptualize the monitoring framework (in your case, vulnerability) as well as incorporate specific Earth observations into the system.

Question 5: How does satellite imagery help us in analyzing the specific distribution of areas?

Answer 5: I am not sure I understand this question. It depends on what we are trying to characterize the distribution of; land use was a topic covered in this training.

Question 6: Was the case study mainly based in primary or secondary data sources? Were local insights of the people living in the landscape taken into consideration? What was the impact assessment of this work? Did it address any existing government policy?

Answer 6: This study was linked to an International Center of Excellence in Malaria Research centered on Myanmar. We worked very closely with local and international experts and healthcare professionals. A large part of our work included knowledge transfer and capacity building. And honestly, the knowledge transfer went both ways - we learned just as much from in-country experts. I would not have been able to conceptualize this model without their input. However, unfortunately, Myanmar was a data-scarce region, and in many ways, we had to rely on secondary data and inform the decision space from primary data or expert guidance.

Question 7: Is it possible to use remote sensing to distinguish breeding sites that are more likely to spread dengue compared to those that spread malaria?

Answer 7: In most of the world, malaria has been a rural disease vs. dengue, which was considered more urban. Different breeding sites for the *Anopheles* vs. *Aedes* were thus determined based on the habitat preference between the two species. However, in Africa, *Anopheles stephensi* is now recognized as the urban vector for malaria, raising the concern about the rapid changes in epidemiological patterns of malaria occurrence towards urban malaria. Identifying larval sites within urban settings has been done but, to the best of my knowledge, nobody has been able to distinguish between *Aedes* (i.e., dengue) and *Anopheles* (i.e., malaria) breeding sites.

Question 8: What types of remote sensing data (vegetation, rainfall, temperature, surface water) are most useful for disease surveillance?
Answer 8: There are many types of RS data available, including but not limited to MODIS (vegetation and land surface temperature), GPM (rainfall), LANDSAT (vegetation and land use land cover characterization), SMAP (soil moisture), Sentinel, VIIRS, etc. However, the determination of the most valuable metrics is determined by the specifics of the disease we observe and the environmental settings. One has to match the data with the topic of investigation. Cholera, for example, is highly related to extreme rainfall. Tick-borne disease surveillance will be strongly informed by winter temperatures. Mosquito-borne diseases in dry ecosystems may be more sensitive to moisture availability, and the corresponding models would benefit most from precipitation (volume and rate) as well as the landscape's moisture-holding capacity. Mosquito-borne diseases in areas with extreme heat might be more sensitive to daytime and nighttime temperature observations. There is no single prescriptive metric. That is why pre-development conceptualization of factors driving the disease in specific geographic settings is so very important.

Question 9: How is biting behaviour-related information for a specific vector collected in a reliable way?
Answer 9: We rely on collaboration (or peer-reviewed publications) with expert entomologists. There are numerous methods and studies. An excellent example is Vantaux et al 2021. Anopheles ecology, genetics and malaria transmission in northern Cambodia. Nature Scientific Reports, 11: 6458.

Question 10: For the Myanmar Malaria Case Study, I'm curious with the large number of remote sensed imageries available for the various hazards, exposures and vulnerabilities, what are your suggestions to group together different EO variables that are likely to be highly correlated? For example, you have MODIS LST Day, LST Night, and LST Diurnal. Naturally they come from the same imagery and there might be high correlations between them. How would you deal with this?
Answer 10: MODIS LST Day and Night come from different satellite overpasses. They are collected by the same sensor but at different times. They are correlated to a degree but not very closely as cloud cover, weather fronts, elevation, and other parameters

weaken such correlation. Ultimately, the important (for malaria in this case) issue is whether daily maximum temperature or daily minimum temperature falls outside the range of optimal temperatures for mosquito breeding and parasite development. Both excessive heat and insufficient heat present a barrier, and thus, minimum and maximum daily temperatures are good metrics regardless of their potential spatial autocorrelation. Another, separate from this example, option is to use models whose predictive capabilities are not strongly impacted by spatial autocorrelation. These frequently include Machine Learning approaches (Random Forest or XBoost) and Deep Learning algorithms.

Question 11: Is remote sensing useful for only infectious diseases? How about the sicknesses like Itai-Itai and Minamata, which are earth and elemental based?
Answer 11: I am not familiar enough with either disease to speculate. As a geospatial data science and EO expert, I expect that remote sensing can help identify the environmental variable proxies that can help in identifying areas with a higher probability of disease occurrence. But my work is limited to malaria and dengue, and I would need to learn a lot more about each of those diseases before I could say definitively whether EO data are a helpful resource.

Question 12: Can remote sensing be used for early warning or rapid response to infectious disease outbreaks?
Answer 12: Yes, and there are many examples - just do a Google search on Early warning and Infectious Disease Outbreaks or Early Warning and Vectorborne Diseases. We will also be presenting a specific example in the second part of this training. There is a gap between the current changing landscape (env. conditions), but a 1-2 month warning can help with monitoring sites and supply locations.

Question 13: How do you deal with different EOs at different spatial resolutions for risk assessments?
Answer 13: This really depends on the focus of a given investigation - if one is carrying out investigations at global, continental or regional scales, then coarse resolution data (1km - 10km) is appropriate, otherwise at country scale and local scale, EO data at 500m or less is appropriate. There are many factors to consider including the temporal resolution needed etc and one could end up using different resolutions to address different aspects of a research/application problem. It will depend upon your reporting unit. But in any case, you will need to bring your data to the same reporting unit scale. If you have input data coming at different spatial resolution, you will either need to

aggregate up (from high resolution to coarse resolution) or downscale (from coarse resolution to high resolution). There is a very large variety of methods for both, ranging from straightforward to incredibly complex. The choice should be driven by a conceptual framework of your model.

Question 14: Is there a comprehensive list of the existing datasets? Is there a way to identify common uses/utility of existing datasets?
Answer 14: This was covered towards the end of the presentation. NASA Earthdata (www.earthdata.nasa.gov) is the central access point for NASA's earth observation data. For different topics, there are summary pages describing what data are available and used for different applications, accessible from: https://www.earthdata.nasa.gov/topics.

Question 15: What are your suggestions for the decommissioning of MODIS? What types of methods are being discussed to match what VIIRS would produce to connect to MODIS, especially when monitoring long term climatic trends?
Answer 15: Here is an Earthdata article which includes a table of similar products between the systems: https://www.earthdata.nasa.gov/news/blog/from-modis-viirs-continuing-legacy.

Question 16: Which metrics best represent vector breeding site suitability in different landscapes (urban, rural, forested, wetland)?
Answer 16: This is a very complicated question. There is no prescribed range of variables. It will depend on a specific situation. For example, for urban, there will be very different answers if we consider business centers of large established cities (e.g., Hong Kong) or slums of Mumbai. Another example, for wetlands, are we considering coastal tidal wetlands or seasonal monsoonal wetlands?

Question 17: How will the environment and climate change the genes of a single kind of animal? As you said, malaria depends on the climate and environment. How will these two factors change the genes of a single animal?
Answer 17: I am not a geneticist and I cannot answer this question. What I can say is the study that I am currently preparing for publication has shown that over the past 20 years malaria econiche has changed both due to the influence of climate change and human activities. Some regions of the world have developed more favorable settings for malaria spread while others became less suitable. There are also several entomological studies that have demonstrated the change in mosquito biting behavior

over time. But I am not at all qualified to speculate on the expected changes in genetic traits.

Question 18: For gap-filling, it is not clear to me how the grid value for the center point of the grid cell is assigned. Does this come from observations on the ground? If so, is that not a problem in areas with a low density of weather stations?
Answer 18: The value for the center of the gridcell is assigned from the satellite image. All high-quality observations from satellite images are used. The missing values are interpolated (at the sea level) to create a continuous surface.

Question 19: How do you work with the seams between images or the gaps between them? Sometimes there are "blacks" in different satellite images linked together?
Answer 19: The “blacks” I presume are the orbital gaps, yet another common source of missing observations. These are frequently filled by compositing values from multiple images over a certain time. For example, almost every MODIS land product comes with some sort of compositing option. There are 8-day or 16-day composites which create gap-filled mosaics to compensate for the orbital gap.

Question 20: When using the gap-filled and predicted products, how would you suggest propagating the uncertainty into other statistical models? For example, for your example on vector habitat suitability, how did you propagate uncertainty?
Answer 20: This is extremely difficult, especially when you are trying to do that at a per-pixel level and you are trying to make a quantitative prediction of the number of cases. Our model was developed in a data-poor environment (on the epi side) and we focused on thresholded/qualitative risk assignment (very low - very high). One of our goals was to use these rankings to stratify for the region for actual case load assessment which we intended to use later for quantitative prediction.

Question 21: What is the reason for using NDVI as opposed to EVI? I understand the latter is better for very "green" (tropical) areas. Is this a resolution or historical time series issue?
Answer 21: It really depends on the ecosystem. NDVI does saturate but only in very dense tropical forests. EVI offers a more nuanced assessment of very high biomass accumulation of green vegetation. For our purpose, the gradient of greenness in very high density vegetation is not particularly important because those areas will uniformly

provide good habitat for vectors. Whereas the variability in lower vegetation densities are well captured by NDVI. But you could really use EVI for preference.

Question 22: Why did you choose a 3-year period (vs. another number) for predicting vegetation greenness?
Answer 22: There is a reasonable stability in climatic conditions and subsequently undisturbed vegetation in our region of interest. The interannual variability is generally low. That makes a 3-year mean a reasonable estimate. If an area has large interannual variability in conditions, a different assessment period, or even a stratification by climatic conditions might be necessary.

Question 23: There is the problem of each "line" of satellite image being taken at different times. When you link each other, there is sometimes a line in the image/tile with different phenomena split by a line. How do you work with it?
Answer 23: I am not sure I understand this question, sorry.

Question 24: Can you explain more about the NDVI prediction, or share information about this technique?
Answer 24: NDVI on Normalized Difference Vegetation Index is not a predicted variable, rather it is a remote sensing metric that quantifies vegetation health and density by measuring the difference between near-infrared (NIR) and red light reflection (Refer to Fundamentals of Remote Sensing Training). High NDVI values (closer to 1) indicate dense, healthy vegetation, while low or negative values are associated with barren land, snow, or water. NDVI is widely used in agriculture, environmental monitoring, vectorborne disease early warning systems and climate change studies to track vegetation growth, identify plant stress, predict fire risk, and monitor agricultural management practices. There are 1000's of publications on NDVI since 1979. A good foundational reference is *Tucker, C.J. (1979) Red and near-infrared reflectance: An indicator of the amount and photosynthetic activity of plant vegetation. Remote Sensing of Environment, 14(1), 36-49.* Its use in global agricultural monitoring is demonstrated at https://glam1.gsfc.nasa.gov/ and there are many other applications you can find.

Question 25: In countries like Myanmar which have many rivers and canals, should waterways not be included in accessibility analysis?
Answer 25: In the areas we surveyed, travel by boat was extremely limited and motorcycles were the dominant mode of transportation. Based on the input of

in-country experts, we considered but did not include rivers and streams in the accessibility assessment. However, in other geographic settings, they may represent a common and necessary component to include.

Question 26: Do you ever use other data sources to gap-fill or validate satellite images? For example, historical GIS maps, community surveys, ground observations?

Answer 26: Validate – definitely. Gap-fill – never. We use ground observations to build and validate satellite products routinely. We used community surveys to gain insight into the conceptualization of our models, particularly to estimate *exposure*. Historical maps have limited use because they are not contemporary, sometimes they are incomplete, and finally, they are highly stylized. But we frequently include other types of digital information – we tried to use everything that was available at the time from the Mynamar Information Management Unit (MIMU).

Question 27: Could you please share the references for the specific NDVI and LST gap-filling approaches? Either for Malaria monitoring in Myanmar or the methodologies used in this study?

Answer 27: Techniques for gap-filling can be found in the references below.

a) https://www.sciencedirect.com/science/article/abs/pii/S0924271621002215 A practical approach to reconstruct high-quality Landsat NDVI time-series data by gap filling and the Savitzky–Golay filter
b) https://www.mdpi.com/2072-4292/13/3/484 Gap Filling for Historical Landsat NDVI Time Series by Integrating Climate Data
c) https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MOD09Q1G_NDVI MOD09Q1G_NDVI - MODIS/Terra Gap-Filled, Smoothed NDVI 8-Day L4 250m SIN Grid
d) https://www.nature.com/articles/s41597-021-00861-7 Worldwide continuous gap-filled MODIS land surface temperature dataset
e) https://www.mdpi.com/2073-4433/9/9/334 Gap-Filling of MODIS Time Series Land Surface Temperature (LST) Products Using Singular Spectrum Analysis (SSA). There are many more.

For the application to Malaria monitoring in Myanmar, this can be covered in a future training in detail.

Question 28: How were the model results validated? What ground truth data did you use? What is the model performance in space and time?
Answer 28: We worked in the data-poor region. Partially, our intent was to use our "expert-driven" qualitative system to identify areas in space and time that public health and medical professionals would visit to collect the necessary data. We intended then to use the data iteratively to validate the existing model and drive subsequent model refinement.

Question 29: How does the temporal burden potential compare to the public health surveillance data?
Answer 29: It compared well, but the data were available only at the state level (the largest administrative unit of the country) and at the annual level. Our system, in contrast, was implemented at 250m resolution at 8-day repeat frequency. No data was available to verify this prediction. We intended to collect those data as part of the iterative validation/refinement process.

Question 30: Has the malaria risk model been used to predict burden in current years, and any idea of how predictions compare to observed cases?
Answer 30: This will be covered more in Part 2.

Question 31: Why do you use the 30m resolution when you can use 10m resolution? What are the benefits of this resolution?
Answer 31: If you do not need the information at 10m resolution, going to 30m resolution can save you on almost an order-of-magnitude of data volume, speeding up your analysis. It also depends on the length of historical record you might need. We have 30m data globally going back to mid-1980s.

Question 32: I have found the DAAC website sometimes really hard to navigate. Is there an API, Python, or R package we can use, maybe developed by NASA?
Answer 32: NASA Earthdata or DAACs may provide example codes to access the data. For example, for the Harmonized Landsat-Sentinel dataset, this Data Access and Tools page has links to examples:
https://www.earthdata.nasa.gov/data/projects/hls/data-access-tools.
Similar pages are available for many NASA missions and datasets. NASA Earthdata is the central landing page for data. Example codes on access are available on the Data Access and Tools page.

AppEEARS is another portal that is very useful. Navigate to the dataset landing page and resources should be available.

Question 33: Doesn't the spatial resolution of satellite imagery (for example, Landsat at 30 m, MODIS at 250 m, or meteorological data such as ERA5 at around 27 km) sometimes represent a limitation when conducting risk studies at a smaller scale where higher precision is required?

Answer 33: Absolutely. I cannot imagine an effective risk model that needs to run at greater resolution than 30 m even at the local scale. The amount of variability you would need to account for if you tried to model at household level is mind boggling. But sometimes, we do need to involve very high resolution imagery <30 cm; and even drone data.

Question 34: Considering that the class terminologies used in Land Use/Land Cover products are not standardized across countries, how can such issues be overcome to ensure consistency in the analyses?

Answer 34: That is partially why we compiled our map from a variety of sources which map a discrete parameter rather than a "class". We used one source for "croplands", a completely different source for "impervious surfaces", A third one for "trees", etc. See Chen et al. (2021). A disease control-oriented land cover land use map for Myanmar. Data (Basel), 6 (6):63.

Question 35: Is it possible to get air quality data through satellites and how can we link AQ to diseases such as tuberculosis?

Answer 35: There are specific missions that are focused on assessing air quality (e.g., TEMPO and MAIA). Other approaches use Aerosol Optical Depth (AOD) from MODIS or other polar orbiting and geostationary satellites to map particulate matter pollution. I don't know enough about tuberculosis to offer an informed opinion.

Question 36: You mentioned that it is very difficult to obtain occupational health data related to vector-borne diseases. Would you have any recommendations for how to proceed when working on this topic?

Answer 36: We have collected very detailed surveys (see for example Hoffman-Hall, A., Puett, R., Silva, J.A., Chen, D., Baer, A., Han, K.T., Han, Z.Y., Thi, A., Htay, T., Thein, Z.W., Aung, P.P., Plowe, C.V., Nyunt, M.M., Loboda, T.V., 2020. Contextualizing Malaria Exposure in Myanmar by Combining Satellite-Derived Land Cover and Use Observations with Field Surveys. GeoHealth, 4 (2): e2020GH000299.)

Question 37: Would it be feasible to model climatic and NDVI data spanning 23 years in relation to leprosy cases? What type of time series analysis would you recommend for such a long period?
Answer 37: We definitely have data records for climatic and NDVI variables spanning an even much longer period. The selection of statistical techniques needs to be fine-tuned to the datasets, their granularity, etc. There are many ways to build a robust analysis.

Question 38: Are the remote sensing data sets equally captured and analysed around the world? I use the tabulated data for my disease prediction, but also different sources give different values for the same indicator, like temperature.
Answer 38: This is one of the advantages of satellite observations – you get consistent and comprehensive coverage from a single sensor around the world. The instrument and the collected data are continuously screened, calibrated, and validated. It can still be biased (but generally the bias is understood), and the datasets are never uniformly good everywhere. But they provide a much better density and consistency of observations. To be absolutely clear, operational quality local meteorological stations provide more precise and accurate information at the exact point, where they are located, but beyond that, the accuracy of their estimate varies more.

Question 39: What resources are needed for the different steps (creation of local datasets, gap-filling, preprocessing), in terms of computing resources and person-time?
Answer 39: As you can imagine, it varies greatly and many of these variables are codependent (e.g. computing resources and person time). We were able to build this with only 3 people at about 50% commitment. On the one hand, these people have PhDs in geospatial data science and satellite image analysis. On the other hand, we had to develop methods from scratch, test, validate, etc. When we executed the model at 30m resolution, it required parallelized runs on our High Performance Cluster. Once we moved to the 250m implementation, we no longer had to parallelize the processing.