Part 3 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 Sitian Xiong (SXiong@clarku.edu) or Jamon Van Den Hoek (jamon.vandenhoek@oregonstate.edu).

Question 1: Has a seasonality of conflict/movement of refugees ever been seen, maybe related to the seasonality of crops or climatic seasons?
Answer 1: Keep in mind that refugee movements are tied to conflict patterns, so refugee movements would be tied to the pulses in armed conflict. There are certainly seasonal factors of armed conflict associated with lean season crop harvest raiding, for example. However, a brief review of the literature doesn't point to climatic conditions or the crop cycle driving being causally linked to the timing of refugee migration, but it doesn't look like this has been thoroughly studied yet.

Question 2: What advancements are being made in the field of noise reduction for Earth observation data, particularly with the integration of Artificial Intelligence and machine learning?
Answer 2: First, the major challenge might be the lack of sufficient data or observations. When there are enough observations, noise reduction would not be a major problem and might not even require advanced methods. AI and machine learning can be used for gap filling – using spatial-temporal relationships to estimate the value of a certain pixel at a certain place and time. There is also a trend towards using multi-sensor fusion to create gap-filled and higher spatial resolution time series data, with machine learning algorithms playing a key role in the fusion process (e.g., Planet's fusion product). For models, Generative Adversarial Networks and other self-supervised training methods are being used for gap-filling, including newer foundation models, such as Prithvi (see here), which is based on a Vision Transformer model. So, if the goal is noise reduction only, then I believe we can follow a similar approach: by leveraging a larger dataset and longer observation periods, we can better distinguish between noise and signal. Some noise was from clouds as well, and those computer vision based models would be useful.

Question 3: Could you provide case study examples or refer to effective climate resilience strategies that have emerged from your research or experience, specifically tailored to enhance agricultural livelihoods in refugee settings?
Answer 3: The two main ingredients needed are access to land and access to markets. Only a handful of countries allow refugees access to arable land to grow crops for consumption or sale, but land access and policies that support the development of agricultural livelihoods for refugees are essential. Access to markets to buy and sell agricultural commodities is also required to give refugee farmers flexibility in decision-making around what they do with their harvest – consume, sell, trade, etc. Behind both of these is socio-economic integration with the host community where refugees engage with local communities, contribute to the economy, and of course benefit from the economic opportunities as well. This is an emergent area of research, but I'd point you to these articles for further discussion:

- Refugees, food security, and resilience in host communities: Transitioning from humanitarian assistance to development in protracted refugee situations https://ebrary.ifpri.org/utils/getfile/collection/p15738coll2/id/128135/filename/128346.pdf
- Refugees as Actors? Critical Reflections on Global Refugee Policies on Self-reliance and Resilience https://academic.oup.com/jrs/article-abstract/33/1/22/5607339
- Building resilience via increased livelihoods opportunities and strengthened social cohesion https://globalcompactrefugees.org/good-practices/building-resilience-increased-livelihoods-opportunities-and-strengthened-social

Question 4: What is the definition of NDVI anomalies in this analysis?
Answer 4: The anomalies here are really the residuals from the linear trend you have fit to the data, so that you can better isolate the annual effects. We also convert the NDVI time series into crop productivity metrics, specifically the NDVI seasonal maximum and NDVI seasonal integral, measured on an annual basis.

Question 5: Did you perform collinearity tests for the multivariate analysis?
Answer 5: We haven't done a multicollinearity test for the regression since we only have two predictors from separate weather variables. Multicollinearity is crucial when dealing with large numbers of closely related variables, especially if you aim to understand the impact of specific variables on the dependent variable. While more

advanced models could be implemented, this is just a simple demonstration of potential relationships between crop productivity and weather variables.

Question 6: The guide says: "To access these geojson files, please create a shortcut to the NASA_ARSET2024 data folder shared with you, and add the shortcut directly under your MyDrive folder." Was it shared with us or should we download it from GitHub and then upload it to Drive?
Answer 6: The data is shared through GitHub. You need to either place it in a Google Drive folder or directly upload it to the Colab directory.

Question 7: Did you apply any filtering of NDVI based on the quality band?
Answer 7: Yes. We applied filters for both MODIS and Sentinel-2 using their QA bands, which you can see in the code. You can see this "`.map(s2_mask_clouds)`" for Sentinel-2 and "`.map(mask_modis_qa('NDVI'))`" for MODIS in the data processing pipeline.

Question 8: Is there any key difference in the analysis when applied to humanitarian applications compared to other applications?
Answer 8: The technical approach may be the same but the social context, availability of other data, etc. will vary.

Question 9: With the r value at less than one, what does this mean correlation-wise, with regards to the series of CAs done in the notebook?
Answer 9: r is a correlation coefficient that indicates the strength of the linear association between two variables. It ranges from -1 to 1, with -1 indicating a perfect negative relationship, and 1 indicating a perfect positive association (as variable X increases, Y moves the amount with no deviation. 0 means no association between the variables.

R^2 is the square of r, and ranges between 0 and 1. It provides an estimate of how much variation in the Y (dependent) variable is explained by the X variable(s), and can be interpreted as the proportion of the variation in Y explained by X. So, if $R^2 = 0.5$, then the independent variables account for 50% of the variation in the dependent variable.

Question 10: How have the recent EO advancements improved the pre-disaster preparedness and hotspot analysis? Can you give recent examples?

Answer 10: With the amount of sensors we have, we are able to see more of what is happening on Earth. A lot of the benefits are also coming through more coverage and scales of visibility.

One example is provided by the recent availability of the high spatial resolution, daily imagery provided by Planet. These imagery provide the ability to develop cloud-free seasonal composites in cloudy regions (e.g. Republic of Congo, Southern Ghana) that have a high enough spatial resolution to detect many smallholder fields. This allows the ability to develop more effective cropland maps, which can then be used to more effectively identify where farming is occurring, and which fields are threatened by Hazards such as floods, droughts, etc.

Part 3 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 Sitian Xiong (SXiong@clarku.edu) or Jamon Van Den Hoek (jamon.vandenhoek@oregonstate.edu).

Question 1: Is there a way we could exploit SAR data for crop health detection that would eliminate the cloud cover issue?
Answer 1: Yes, the major advantage of SAR is that it penetrates clouds, thus having fewer or no cloud cover issues. We did not include SAR in this training, but there are existing ARSET training sessions for using SAR. We will post a link to the related training in our answer document.

ARSET training related to this:

- Mapping Crops and their Biophysical Characteristics with Polarimetric SAR and Optical Remote Sensing

However, current SAR (e.g. Sentinel-1) is often less effective than optical time series for key crop monitoring methods, such as crop type mapping and possibly productivity estimation as we have done here. One of the major reasons is that in smallholder croplands, many of the fields are quite small, and SAR has speckling artifacts that make its effective resolution somewhat coarser than the typical field size. Previous work has shown that time series provided by Sentinel-2 often convey the majority of the signal needed for mapping crop types, and our current work shows that SAR is less effective for mapping crop phenologies.. SAR does provide some useful information when used jointly with optical time series, but on its own models tend to perform poorly. One potentially useful role for SAR is to use it in models (e.g. Conditional Generative Adversarial Networks) that can learn relationships between optical and SAR data, in order to create gap-filled synthetic time series.

Question 2: Can we normalize the temperature data and compare it with NDVI values to get better insight? What could such a comparison suggest?
Answer 2: Normalization is a related but different concept from detrending. Normalization adjusts the scale of the variables to e.g., 0 to 1, so no variable dominates the analysis because of its scale. But normalization itself is not needed in this simple regression-based analysis to understand the relationship between temperature and NDVI-derived productivity or precipitation and NDVI-derived productivity. Scatter plots of each variable will show the same pattern, whether normalized or not. So normalization would help as inputs for particular models, and actually it is a standard preprocessing for many analyses (and choice of normalization method is particularly important for neural networks), but it may not give the same insight that detrending gives you in this relatively simple analysis, where detrending helps separate annual impacts from longer-term changes.

Question 3: Are there any studies done on the feasibility of growing dates through remote sensing around refugee settlements, as dates are commercial fruit crops which can be a potential source of livelihood. Also, they are grown in warmer climate regions, which is consistent with refugee camp areas.
Answer 3: The FAO crop calendars are essential information here. Growing dates in refugee areas should not be meaningfully different than dates for nearby, non-refugee or broader, regional croplands (https://cropcalendar.apps.fao.org/#/home). As far as fruit crops, these are not as widely documented as vegetable crops, but similar crop calendar information is available.

On the topic more generally, there is wide literature on using remote sensing time series to estimate crop phenologies/land surface phenologies more generally. Much of the agricultural work focuses on areas with larger field sizes, and relatively less work focuses on regions where smallholder agriculture is the most common form. We have actually done recent work, led by Dr. Mike Cecil, that is coming out in the next few weeks in which phenology measures estimated from several satellite sensors (including Sentinel-2) are compared to those from ground-based sensors (see earlier slide for photo of sensor in field) in maize fields in Zambia and Kenya, with a focus on

understanding how well key dates (planting, maturity, harvest) can be estimated in smallholder fields. We will provide a link to the paper when it is available.

Regarding the phenology of fruit crops, we don't have much experience with those, but from a general perspective there may be other factors more informative than VI curves in many cases, such as number of days above or below certain temperature thresholds. VI values may be informative in many cases for capturing general fruit-bearing plant productivity variations that also impact fruit production, but that might not always be as informative. One example, tomato plants might do very well in terms of high growth rates with increasing rainfall, but the tomatoes themselves might yield poorly when there is too much rainfall because of split skins, etc.

Question 4: I don't really understand what the NDVI integral is and how it's calculated. Could you please explain more?
Answer 4: The NDVI integral is the area under the NDVI curve in an NDVI time series. It measures the cumulative NDVI value over time, so if the NDVI has a low value at any point, the NDVI integral will capture it. For example, if the crop experienced any adverse events, the NDVI integral will reflect the impact. And it does not matter when this adverse event happens; it serves as a summary of the growing season. A simple way to calculate NDVI integral is to multiply the NDVI value to the time step in your time series, assuming you have a regular time series. Otherwise, you use the NDVI values to multiply the corresponding time steps to the next observation.

Question 5: Can we follow a similar approach to understand how crops could be damaged during seasonal floods or floods in general?
Answer 5: Yes, we can apply this general method for evaluating crop productivity to evaluate flood-related damage. One challenge is that floods are often associated with heavy cloud cover, making it difficult to obtain a good NDVI time series and calculate the NDVI integral. Having a crop metric based on a specific time point rather than the entire season may be more feasible.

You can also obtain fairly good insight into likely flood damage to crops if you 1) have a good cropland map, and 2) use flood mapping techniques, such as those based on

radar datasets. Overlaying flood maps with cropland maps can give you an insight into the likely extent of crop losses.

Question 6: How can we know what kind of crops are being affected if we want to do an impact-based forecast?

Answer 6: Ground data is generally the best way to do this. The types of crops growing in fields can be classified using satellite remote sensing techniques, and the data and models to do this are getting better. However, you still generally need to have large amounts of data collected from the ground or near ground to show what crop is in what field for the season of interest, and these data are expensive and hard to collect in a way that is representative of the crops growing in a region (drones, which we have applied for this purpose, can help collect near-ground crop observations in a more representative way, but that poses its own challenges with image processing, etc.), so training, or at the very least assessing, crop type models, which generally require time series of images to distinguish crops based on their phenological curves, may often be too challenging.

An alternative approach that is more effective is to use the crop surveys that many governments collect to get a sense of the relative frequency of different crops growing in particular administrative districts. Although these data are not collected in some countries, they are available in many countries, at least for some years, so this is a valuable source of crop type knowledge that can be linked to generalized crop productivity-climate metrics of the type we are demonstrating here to make simple estimates of impacts across various crop types at administrative levels. These can be combined with statistical disaggregation measures to map crop-specific information onto cropland maps (e.g., Monfreda et al, 2008).

Question 7: Can we resample or use pan-sharpening techniques to make satellite imagery with lower resolution consistent with higher-resolution data for our analyses to get a longer time series with finer details?

Answer 7: There is an algorithm called STARFM to fuse MODIS and Landsat to make a higher spatial and temporal resolution product. We have to pay attention to the potential temporal mismatch between sensors

(https://modis.gsfc.nasa.gov/sci_team/pubs/abstract.php?id=07653) and newer implementations that are sensor-agnostic (https://github.com/nmileva/starfm4py). Simple pan resharpening often introduces artifacts that lead to greater noise in long term time series analysis, and not all sensors have a pan band to extend the time series backwards. Super resolution approaches are another direction for increasing the nominal spatial resolution of a dataset (https://www.nature.com/articles/s41597-023-02538-9), but this has not been well explored in a long term time series scenario where stability in the signal is of paramount importance.

Planet has also developed a new fusion product that provides Sentinel-2 optical information at daily time steps with PlanetScope resolution (~3 m). This is based on their own multi-sensor fusion model. Although commercial, this approach highlights a trend towards combining sensors to provide more consistent image time series at moderate to high spatial resolutions that will likely continue. Newer foundation models such as Prithvi are also effective in helping to generate gap-filled time series.

Question 8: Do the crop productivity metrics obtained from NDVI have the same meanings as the ones that could be obtained from other VIs like SAVI or EVI, for example? Will NDVI max mean the same thing as EVI/SAVI max in terms of crop productivity?

Answer 8: NDVI, EVI, and SAVI are highly related, and each might be suitable for assessing productivity. However, they will likely give different values for specific vegetation at a given time and will therefore be unlikely to have a 1:1 relationship, and the same holds for metrics derived from them. EVI is designed to improve sensitivity in high biomass areas and reduce atmospheric effects by including the Blue band and applying coefficients that adjust for canopy background signals and atmospheric influences. According to these characteristics, EVI may reach its maximum later as it is less likely to saturate in dense vegetation, and the value may be influenced by background and atmospheric conditions. Similarly, SAVI reduces the influence of soil brightness in sparsely vegetated regions, where NDVI may be more influenced by the soil. In the notebook, we showed how to calculate the NDVI for Sentinel-2, and we showed how to select the VI band for MODIS. You can explore EVI by using the EVI

formula for Sentinel-2 and select the EVI band in the MODIS VI product. For SAVI, the MODIS product may not have it as a default band, so you will need to calculate it using the MODIS surface reflectance product.

Question 9: As far as food security assessment is concerned, a concern that we usually face in Cameroon is assessing the level of agricultural productivity during the last agricultural season. How could this approach be used to face this issue?
Answer 9: You can use the notebooks provided for this training to analyze this problem by updating the location and dates of interest.

Question 10: What are the spatio-temporal dynamics of crop phenology and biodiversity interactions under different climate scenarios, and how can advanced predictive modeling techniques (e.g., machine learning, agent-based modeling) be used to forecast these changes?
Answer 10: There are many possible interactions that could be examined, and the methods needed to examine such interactions would vary widely. One example could be a crop that is dependent on being pollinated by a wild species that is migratory in nature. Climate change could impact both the phenology of the crop and the migration timing of the pollinator, causing a mismatch that undermines the pollination of the crop species. You will need to look at techniques to model the planting of the crops (e.g. remotely sensed estimates of crop phenology) and analyze the movement of species for pollination (e.g. citizen science observatories that track the arrival of different species).

Question 11: Are there crop datasets available for analysis?
Answer 11: Yes, and there are a growing number of those. Please see the Radiant Earth Foundation/Source Cooperative, as well as AWS's Open Data Registry. PULA's BirdsEye data is an open crop dataset covering several southern and eastern African countries. The Open Science Foundation may also host a number of datasets. We will include some links here.

Question 12: How can machine learning algorithms be employed to analyze large-scale, multi-temporal remote sensing datasets for monitoring biodiversity

changes in agricultural landscapes, and what are the limitations and uncertainties associated with these methods?

Answer 12: This question involves many details and components to think about. However, one important development is growing efforts to map essential biodiversity variables (e.g. see here and here), which can be followed to gain insight into the related methods.

ARSET trainings related to this topic:

- Fundamentals of Machine Learning for Earth Science
- Large Scale Applications of Machine Learning using Remote Sensing for Building Agriculture Solutions