Title

Climate Effects on Food, Agriculture and the Environment

Authors

Tse, Justin
Zarate, Alina
Bruellman, Ryan
et al.

Publication Date

2024-11-21

UC GIS Week
Thursday, November 21, 2024
10 AM – 11 AM
Climate Effects on Food, Agriculture and the Environment

Sessions:

Welcome to GIS week day three. It's great to see everybody coming in this morning for this wonderful session on climate effects on food agriculture in the environment. My name is Erin Mutch. I'll be your moderator for this session on what UC Merced. We're going to go ahead and go ahead and please read through the top slide here. Please feel free to ask any questions in chat. Time permitting, we're going to have all questions at the end. We have four sessions this morning, three lightning talks and one for about 15 minutes. Again, the focus is on climate effect on food agriculture in the environment. Again, happy GIS day week. I guess everybody, I don't know, was it yesterday? I guess it was. It's great to see everybody. So I'm going to go ahead and get started with Justin Tse from, UC Berkeley, and their presentation today will be comparing GIS and WUDAPT approaches for local climate zone mapping, a case study in Denton, Kelly, thank you for presenting. Yeah, let me share my screen.

Comparing GIS-Based and WUDAPT Approaches for Local Climate Zone Mapping: A Case Study in Denton County

Speaker: Justin Tse – UC Berkeley

Abstract:

The Local Climate Zone (LCZ) classification scheme offers a standardized framework for characterizing the local thermal environment, revolutionizing urban climate studies by moving

beyond the traditional urban-rural dichotomy. This method classifies the landscape into 10 built types and 7 land cover types to provide a better representation of the urban fabric and morphology. While machine learning approaches using satellite imagery have gained tremendous popularity in LCZ mapping, they often require high-quality training samples and can introduce uncertainties depending on model performance and data quality.

In this study, we applied a GIS-based approach to map LCZs in Denton County, TX at a 100-meter resolution. Utilizing 1-meter resolution land cover data and LIDAR-derived products, we extracted key indicators such as building height, building surface fraction, and impervious surface fraction, as outlined in the LCZ framework. Our approach highlights the benefits and trade-offs of using selective indicators to optimize the mapping process. This work contributes to a more comprehensive understanding of LCZ classification and further benefits urban climate research.

Transcript: Video Timestamp: 1:38

Hi everyone, I am Justin Tse, a second year master's student in the Department of Landscape Architecture and Environmental Planning at UC Berkeley. So today I'm going to be presenting, comparing the GIS based and well-read database as a photo to approach for local climate zone mapping, a case study of Denton, Kelly, and Telsus.

So here is my agenda. I'll first talk about the current challenges and introduce local climate zone and then I'll talk about the GIS mapping, evaluation and comparison and limitations.

A majority of existing urban GIS studies rely on the differences of the temperature differences in the urban area and temperature in the rural areas. However, this equation is problematic because first, urban has no single objective meaning and secondly, the urban rural division is actually not quite sufficient to devise cities and the countryside. And to address this problem, server and okay developers local local climate zone classification system to categorize the environment into 10 biotypes and seven link types using a list of different, a list of indicators showed you and then the factors, the indicators highlighted in red are some of the most common indicators that people have been using to map the local climate zones, which include the sky to factor aspect ratio, building surface friction, impervious surface friction, perfect surface friction, height of worthless and terming worthless. And so there are currently two mainstream LCC mapping approaches. One is for machine learning base where users would define their training samples and one a machine learning algorithm to map the LCC and another one is a GIS based as you can, as you can imagine both of these approaches have their pros and cons. And currently the WUDA PT has a global LCC datasets derived from using a machine learning based approach that it's very available to the public. But how is its performance compared to a GIS based map wall? And so to answer the question, I chose Denon County as a study area. So Denon County is located in state of Texas. As you can see here, it's part of this urban abomination in this region adjacent to Terrain County and Dallas County. And, and, and compared to the, to the adjacent region, Denon County actually doesn't really have a lot of urban areas and with over 50% of the land cover is grass.

So here is a general LCC mapping process. First, we would collect the data and then define the basic spatial unit and they will calculate the urban canopy parameters and using these parameters to classify the local climate zones and they will have processing and lastly, visual evaluation.

So here are the three datasets that I've used in the LCC mapping process. So I have the records of building polygons. These are the building for pre and then I also have the building high roster derived from the LUTR dataset and I also have used the LUS, the length of the dataset provided from by the urban watch database and the building high roster and also length of the data roster has a one meter resolution. And so one common way to define this basic spatial unit is by looking at the spatial autocorrelation of the building heights to looking to look at what distance can, at what auto mode distance will capture similar building features. However, this record is actually not applicable in Denon County because most of the buildings had a very homogeneous building heights. So I picked a 100 times, 100 meter resolutions as the basic spatial unit which is also consistent with the global data set.

And here are the three, so using these three big datasets I derived seven indicators. As you can see here, these indicators are actually not consistent with the indicators that I introduced in the first slide. Because, probably because some of the indicators were posted by Oki, they are actually quite computational intensive such as the Skyfield factor. So I introduced some new indicators as an alternative such as the Lumper of Buildings and the Mesmer Building for pre.

And in terms of the LCC classification, I used a modified standard rule-based approach which means modifying the range of values proposed by, proposed in Oki's paper. And in the first step, I used the building surface friction at the number of buildings to distinguish the build types and the length of the types. And then in second step, when classifying the build types, I incorporate the rest of the parameters, indicators such as the mean building height, Mesmer Building for pre, and length of the curve, and also previous surface friction to further distinguish the different build types. And then for the length of the types, I mainly used the length of the data set to identify the majority length of the beam, the 100 meter cell size.

And I also used the tree canopy surface friction to further distinguish the tree type and also dense forest. So, and in the pro processing, so during the pro processing, I, so this is a, in the first stage, I applied a feed by feed moving window to only select the majority, to majority length type within that window. And then in the next stage, I eliminated isolated pixels. So as you can see, the original roster, we have a lot of noise initially, we have a lot of noise, but after the pro processing, we were able to eliminate those losses. So here is the result. So using, so I I pick, I generated 150 samples using a stratified, a stratified random sampling approach. And I referencing the high resolution satellite data to develop the cloud truth data and then develop a confusion matrix to objectively compare the overall accuracy. As you can see here on the left is the GIS based method, which has an overall accuracy of 80% as compared to the global data sets, which only has a 47% as you can see, the GIS based approach has a much higher accuracy rate compared to the global data sets. And if you look at the roster here, we at the regional scale, it looks kind of similar, but when we zoom into it, you can see at a local scale, we can see that the world database, the global data set actually discuss a lot of the large low rise area. And these areas are typically the big boss areas, big boss commercial zones such as cost code and yeah, yeah, commercial areas. And as you can see, the producer accuracy, the global data set only has a 13% accuracy compared to the GIS based.

Another way to perform a devaluation is by looking at the land surface, by looking at the temperature, it can be land surface temperature or air temperature. And here is just an example of the boss part using producer from the BODIS data set using the BODIS land surface temperature, comparing the daytime and also lifetime land surface temperature.

But ideally, if you want to perform a ball robust evaluation, you would have to look at the intra and inter-summer variability by calculating the differences between each of the local climate zones.

But here is just an example to show you how we can also utilize temperature data to compare their performance. So there are several limitations in my study. First of all,

is the data inconsistency. As you can see, the urban watch data set and the Microsoft Building for P this, the basic space review need to degree that I applied that doesn't really fit the raster data set as you can see here, the outer edge only containing a portion of the of the land cover data, which means these areas doesn't have the complete information to perform the classifications. And so these areas are oftentimes misclassified in this GIS based approach. Both slightly, they will require the band new classifications to further enhance the accuracy.

And here are my references. And if you're interested in using the data set or learning more about the process, I can send you over the script and the data set. And I will also try to make all this information available on GitHub in the future. Thank you.

Thank you, Justin. I'm happy to see that you use the Microsoft Building database. I've showed that to some students as an encouragement to try to use that as that data years ago was really difficult to find. So great work. And thank you for your presentation. Feel free to putting any questions in chat. Again, we'll do Q&A at the end. Our next speaker is Alina Zarate from UC Berkeley, assessment of multivariate drought impacts on agriculture in central Chile for the enhancement of sustainable adaptation. Thank you for presenting.

Assessment of Multivariate Drought Impacts on Agriculture in Central Chile for the Enhancement of Sustainable Adaptation

Speaker: Alina Zarate – UC Berkeley

Abstract:

Over the past decade, Chile has suffered under “Mega-drought” conditions placing significant pressure on agricultural production. In response to the crisis, the Chilean government has moved to promote sustainable agricultural practices to increase agricultural resilience in a changing climate. This research focuses on multivariate drought impacts on agriculture in Chile’s Región Metropolitana de Santiago and Región O’Higgins in order to spatially characterize priority areas for sustainable adaptation based on overall vulnerability to drought. To address this topic, I pose the following sub-questions: 1) How have meteorological conditions, streamflow levels, and vegetation conditions in agricultural areas changed from non-drought years (2000-2009) to drought years (2010-2020)? 2) Based on meteorological, streamflow, and vegetation condition factors, which agricultural areas are most susceptible to drought conditions and where should Chilean sustainable agricultural development be focused? The Palmer Drought Severity Index, the Standardized Streamflow Index, and the Vegetation Condition Index were used to calculate changes in meteorological, hydrological, and agricultural drought severity respectively from 2000-2020 derived from the Catchment Attributes and Meteorology for Large Scale Studies, Chile Dataset and from Landsat 5, Landsat 7 ETM+, and Landsat 8 satellites in the Google Earth Engine platform. Index values were compared by running a suitability analysis in ArcGIS Pro. Overall, the results suggest that high vulnerability agriculture is located primarily in the Región Metropolitana de Santiago and did not correspond to areas identified by the Chilean government as being in severe drought. This research will help improve our understanding of the potential of combining drought indices to determine agricultural vulnerability and inform the implementation of sustainable adaptation in Chile.

Transcript: Video Timestamp: 13:27

Hi, everyone. Thank you all for being here. My name is Alina Zarate. I'm a first year PhD student in the energy and resources group at UC Berkeley. But I'm going to be presenting on some work that I conducted as an undergraduate here at Berkeley. So growing up in Southern California, drought has always been a presence in my life and something that I've been really concerned about. So when I had the opportunity to study abroad in Chile in fall 2022, I was acutely aware of the similarities when it comes to drought conditions. For those that don't know, Chile has been experiencing a mega drought or an exceptionally dry period lasting many years since 2010. And you can actually visually see this in the photo that I have here. The snowpack in the Andes around Santiago is incredibly low. And so this has resulted in extreme water shortages and competing water needs, particularly between rural and urban areas.

And while I was living in Chile, the Chilean government actually released announcements for new efforts to mitigate drought impacts in the agricultural sector, which led me to wonder, is there a way to use drought analyses to actually inform sustainable development? And so that led to my research questions, which were how can sustainable agricultural development be best implemented in central Chile to address the impacts of drought? And I sought to answer this question by looking at how meteorological stream flow and vegetation conditions have changed from non-drought years, so 2000 to 2009, to drought years, so 2010 to 2020. And then based on these factors, to see where Chilean sustainable development should be focused. Now, drought is a complex phenomenon that can be defined and analyzed in a number of different ways. So meteorological drought refers to atmospheric causes of dryness, so precipitation, evaporation, surface temperature. Hydrological drought refers to abnormal stream flow, groundwater levels, and agricultural drought focuses on agricultural cultivation, so production and crop health. So typically, when researchers analyze drought, they will analyze one type of drought. But my plan for this project was to actually analyze all three different types together, using three different drought indices. So using the Palmer Drought Severity Index, the Standardized Stream Flow Index, and the Vegetation Condition Index, using remotely sensed and in situ data that I processed in Google Earth Engine, in RStudio, and in ArcGIS Pro.

And the goal was to take all of these spatial results of the various drought analyses and overlay them using a suitability analysis. And so based on these combined results, the goal was to identify which areas were most vulnerable to drought impacts overall. And then the final step would be to actually compare these results to analyses conducted by the Chilean government and their classifications of which areas are in the most extreme drought and are therefore most vulnerable. So for my study site, I selected 14 watersheds that contain the majority of agricultural production in the Metropolitan and O'Higgins region, so kind of the area surrounding Santiago where I was living. And so I analyzed each of these different types of drought. So this is the meteorological drought analysis. So I calculated the average monthly PDSI value for each region. Positive values indicate wet conditions.

Negative values indicate dry conditions. And these red arrows indicate when the start of the mega drought occurred in 2010. And so as we can see, there's pretty significant differences in dryness, except for kind of this anomaly in 2016 when they happened to get a lot of rain. And then the map indicates the difference in average PDSI values from pre-drought years to drought years for each watershed. So basically showing which ones have seen the greatest amount of change during this period of time. And as we can see for meteorological drought, this occurred primarily in the metropolitan region around Santiago. Very similar process of analysis for hydrological drought. So I again, you know, calculated the average with positive values being wet conditions. Negative being dry conditions. Red arrow indicating the start of mega drought. But as we can see, when comparing which areas have changed the most, it's a little bit different with more areas in the central region experiencing greater differences in stream flow.

And then the final analysis that I conducted was a vegetation condition analysis. So this looked at peak agricultural growth during the wettest five years and compare that to the driest five years. And really looking at the greenness of the vegetation during that time. And so from those results, I found that all of the watershed exhibited mixed vegetation values, but that these northern watersheds had lower overall BCI values indicating drier conditions. And so more drought impacted vegetation.

And so putting it all together, I conducted a suitability analysis in ArcGIS Pro and found that one vulnerability is not evenly distributed within the study site, but rather these northern watersheds, particularly around Santiago exhibited high vulnerability while those in southern regions exhibited comparatively low vulnerability. So vulnerability seemed to decrease with latitude. And then I also found that these results did not align with the classifications made by the Chilean government. So the watersheds outlined kind of in this purple color indicate the areas that they had identified as being completely depleted of surface water and in extreme drought.

And so what does that tell us? One, I think it's very clear that sustainable adaptation should be prioritized in watersheds that are adjacent to Santiago compared to what the Chilean government had previously classified. It also shows that an important factor to include in future analyses is socioeconomic drought. So conflicting water needs and water management infrastructure. The watersheds around Santiago are going to have the biggest conflict between urban needs and rural needs, showing that that is a huge factor in dealing and adapting with drought. And then finally, this shows that there are new approaches to conducting this type of analysis. So drought is complicated and incorporating more variables and factors can be effective, not just in measuring drought, but also in forming adaptation. So those were kind of the results. This is something that I hope to continue working on as I move forward in my PhD. But with that, thank you all.

Thank you very much for sharing your research and your maps look great. I enjoy your color choices. I always appreciate that. And again, any questions you want to ask, you can do in chat and hopefully you can capture those at the end of our session. Next up, we have Ryan Bruellman, who's with UC Riverside. And his presentation is a recipe for health disparity, quality of neighborhood grocery stores. Very important research. Thank you.

A Recipe For Health Disparity: Quality of Neighborhood Grocery Stores

Speaker: Ryan Bruellman – UC Riverside

Abstract:

Access to fresh, healthy food is a crucial component of adopting and maintaining a healthy diet that may offset aging-related diseases such as Type 2 Diabetes. Supermarkets and grocery stores consistently serve as a mainstay for Americans' main food source, particularly in urban settings. Classifying supermarkets and grocery stores based on factors such as cost and quality could be an important consideration in understanding health impacts across regions and within cities. This study offers classification of grocery stores and supermarkets across the United States based on the overall cost and quality of food. We performed network analysis within ArcGIS Pro to elaborate on access to these food sources by census tract in major urban areas. Regression models of many large urban areas across the United States showed positive associations between the presence of low-quality grocery stores and Type 2 Diabetes rates,

whereas presence of higher-quality grocery stores was associated with lower Type 2 Diabetes rates. In many urban areas, these associations were still significant even after factoring in overall socioeconomic factors. Future analysis aims to further parse apart quality metrics, socioeconomic factors (i.e. income and vehicle access), as well as the impact of overall cost of these grocery stores and supermarkets. This study can offer insights into public health policy on reducing the incidence of aging-related diseases exacerbated by a poor diet and compounded by lack of access to higher quality, affordable food sources.

Transcript: Video Timestamp: 21:44

Thank you so much. Hi, everyone. My name is Ryan Bruellman. I'm a PhD candidate at University of California Riverside. And today I kind of want to emphasize why thinking about one's grocery store quality is important when you talk about different types of health outcomes. So we well know that access to things like fresh produce as well as fresh proteins and baked goods is essential for maintaining a healthy diet. And that's obviously important for us to prevent the effects of impacts of aging and disease prevalence and as well as exacerbation as well. So within the overall households, individuals can go to fast food locations, restaurants, convenience stores to get different types of food. However, the vast majority of US households, regardless of income, do get a lot of their food from a grocery store. Now, how they get that varies based off of the different things available to them. Most individuals will use their own or someone else's vehicle. If that's not the case, then obviously they need to rely on walking, biking, or using public transit. It's important to note that the research has shown that individuals, while they do shop at a grocery store traditionally within their vicinity, it might not be the closest one. So it's important to look at not necessarily the closest grocery store, but just kind of the grocery stores in the general area. Now with my project, some of the different hypotheses that I'm working with, again, kind of getting at that high quality aspect of things, I'm expecting that unhealthy outcomes, and for this talk, I want to focus on type 2 diabetes rates, are going to be negatively associated with areas that are with having access to a lot of high quality level grocery stores. And then on the other side of things, when there is a huge prevalence of low quality grocery stores throughout one's neighborhood, they kind of overpower that and really potentially exacerbate these unhealthy outcomes such as type 2 diabetes rates.

So my materials and methods that I'm working with, I'm working with Timepoint roughly between 2015 and 2017. So pre-COVID, and I'm focusing in on urban metro areas. So I didn't necessarily want to focus on rural areas. I grew up on a farm, we grow a lot of our own food, and also getting to the grocery stores varies a lot depending on the rural area. So for this, I'm focusing in on generally cities. For this, I'm bringing in the 500 Cities Project from the CDC, which looks at different disease rates across the US, and I'm working within a subset, so roughly 60 metro areas across the US. It's very important to bring in socioeconomic or SES factors to this as well, because obviously income and access are a huge component. So for this, I'm bringing in another component from the CDC known as the Social Vulnerability Index, which has those different types of variables. So for those that are not so familiar with it, I'll just kind of give a basic overview of what's all included within a Social Vulnerability Index. So they have different buckets of categories, one of which is known as the socioeconomic status bucket, which you can see here, those relate to overall income as well as education levels. You have household composition as well, which is related to kind of age range of individuals, disability as well as single parent households. There is a minority status bucket as well, and then finally housing and transport, which this kind of gets at your basic composition in the census tracts of the overall housing in that area, as well as importantly for this no vehicle access that's obviously going to be a major component for individuals getting to a grocery store or their means. Now a lot of these different variables within here, and the Social Vulnerability Index has metrics where they look at each individual bucket, but they also have one where they combine these together and create what's known as a Social Vulnerability Index, and that's what I'm going to be working with and presenting on for this talk itself. So higher levels of Social Vulnerability Index generally mean lower levels of access as well as lower socioeconomic status, those type of things. So my supermarkets and grocery stores, I've coded, geocoded these locations across all 50 states, and as you can see, I have roughly over 65,000 of these locations. Now these are, these range from your traditional supermarkets and grocery stores, but these also I tried to gather those that were not necessarily just regional or national, but also statewide or kind of more localized ones. These were broken down by cost and quality. I'm going to focus more so on quality, but I do want to touch on the cost metrics just so that way it's clear in kind of my future directions, but I did break down costs from standard and high, and then quality from low, medium, and high as well. Now the cost metrics were based on 100 common items that were across this study that was done in Seattle, so I looked at these different items and took the base price. This was generally speaking for at least the process foods, typically the store brand, and I usually got them from the store website these prices, but sometimes there was no store website, so I had to use things like Instacart or DoorDash, and I had to adjust that appropriately. Obviously those are going to be higher level prices.

So what this looks like, and sorry I'm going to go through this relatively quickly because again I'm focusing on quality here, but just to kind of give you an idea of what this would look like, here are some different examples of some of those stores. I coded different items relating to different kind of groups. You can see here dairy, there's also proteins as well, so these would be your meats, nuts, different things like that. Grains are in here as well, and then also fruits, vegetables, sorry again for going through quickly, oils and sweets, as well as beverages, and from there I was able to kind of parse apart a little bit of the cost aspect of things. Now what this looks like, kind of in the large scale of things, I broke down to standard cost low quality as one of the categories. So this is representing roughly over half of the locations. Now what these are typically are your dollar stores, so these would be your dollar general, family dollar, where they really when you walk into the store it's a huge influence, it's a huge kind of process foods. You might have a few little areas of produce like apples and bananas and things like that, but it's not necessarily a major focal point. Typically the focal point in those stores would be your process foods. Now moving on to medium quality, so standard cost medium quality, these are roughly around a quarter of the locations. These are going to be where fresh produce is available, but there's no counter service, so you're going to not have a butchery, deli, seafood counter, bakery, that type of thing. Generally those type of items, they'd have to package them before and bring them in. So while there is some options, you do have a department like generally produce in that area, it's considered medium quality. So this would be your Walmart's targets, and I know I'm going to get flack for this, but Trader Joe's would be considered medium quality as well.

Standard cost high quality, this is kind of the final area within the quality metrics. So this is representing roughly just over 20% of the locations. This would be where you have that counter service, you have their seafood department, things like that. A lot of these are going to be your large chain grocers, so Vons, Albertsons, Ralphs, Kroger, Publix, those type of things. I do want to mention I'm not going to again focus on cost for this talk, but I did break down high cost high quality, because I hypothesized that eventually if you plop a high cost grocery store in an impoverished area, it might not necessarily have the same influence that a standard cost would. So it's not a lot of locations, but it's just over 2,000. These are ones like your Whole Foods, as well as Safe Way, that are generally higher cost relating compared to some of the other chains. Now my methods, I've done some overall network analysis and regression, and I've done this by tract. So driving buffers of roughly 10 minutes, getting to the different access within 10 minutes, how many grocery stores can you get to in that amount of time. The counts of these grocery stores were what I have done for each of the different quality metrics.

So what this looks like running the overall generalized linear regression protocol in ArcGIS, I have my outcome as type 2 diabetes rates, and I build from the univariate model. So first looking at low quality grocery stores, and then for now I've combined medium and high quality. So again, this is kind of a story right now of low quality versus the others. I then combine the two in the next model, and then finally in the next model, I bring in the socioeconomic factors to see how that plays a role here. So what this looks like across the US, again these are roughly 60 cities across the US. The median low quality grocery stores that one would have access to kind of what that looks like across the US, roughly a third of them would be low quality in your general vicinity, versus two thirds of them would be medium quality, high quality. And the median diabetes rates across these tracts was around 9.9 percent. So these regression models on its own low quality was significant, and as you can see it was positively associated with higher diabetes rates, so higher levels of low quality access, the entire levels of diabetes rates. On the other side, medium quality and high quality was associated significantly with negative, with driving down diabetes rates. And then incorporating the two together kept that significance. Now interestingly, when you bring in socioeconomic, as you can see here in the bottom right, those significances do hold. So it's important to note that even accounting for things like the SES metrics, these significances and impacts on the diabetes rates do hold. Now I've done a lot of different cities here, and in kind of lack of time I want to just focus in on where probably most of us are in the Pacific region.

So within the Pacific region here, you have different cities such as Los Angeles metro, so this would include LA County, Orange County, as well as the Inland Empire, Portland, Sacramento, San Diego, San Francisco Bay Area, and then Seattle. Now it's interesting to note that for the Pacific region, you can see that the medium quality, high quality grocery stores are much more prevalent. So not surprisingly that I, the medium diabetes rates are much lower compared to the rest of the U.S. Now looking at those same type of models, again low quality is significant on its own. Interestingly on its own, medium high quality is not, but when you bring them together, the significances do come, are present for both of them, and it's looking a lot like the trends at least as far as the U.S. showed, and these significances do hold when accounting in for socioeconomic factors across the Pacific. Again, pressed for time a little bit, but I want to look at one specific city, so I'll focus in on San Francisco here. So San Francisco, a population of roughly 4.5 million, you can see here that overall low quality grocery stores are quite low in San Francisco, and therefore not surprisingly the median diabetes rate is also quite low underneath the U.S. median that we saw previously. So just kind of orient everyone.

Some of you may be familiar with the Bay Area, but just for those that might not be, the red area is a community known as Hayward. You have further up north UC Berkeley campus, and then over, of course, you have downtown San Francisco in the green box with the Bay Bridge. Now, doing a cluster kind of approach with these different grocery stores, the different census tracts, the pale yellow areas indicate relatively low diabetes rates compared to the U.S. average. The orange areas would be kind of around that U.S. average, and then the red areas are over the U.S. upper limit estimate for diabetes rates of 14.1 percent. It's interesting to note, so the low quality grocery stores are indicated by the red clusters, and then medium quality, high quality are by the blue clusters. And you can see that an area like Hayward, and just north of it where you have kind of the orange clusters of higher diabetes rates, you don't have as big of medium quality, high quality grocery stores, but also the red dots are almost in line with what you see with the blue dots, whereas up by Berkeley and then over in the downtown San Francisco area, you get those relatively larger medium quality, high quality grocery stores. So visually, you can kind of see a little bit of the impact going on here. However, if we look at the overall kind of SES factor of things, this is almost operating like a heat map, if you will, where darker red indicates higher levels of SVI overall. So kind of this is getting at probably higher poverty, as well as vehicle access and things like that. You can see that the Hayward area is generally slightly darker, as well as north of the Hayward area, with the red. So what does this look like overall for a city like San Francisco in those models? You can see that low quality is significant on its own, medium quality, high quality is as well. Combining the two, everything holds similar to what we saw with the general Pacific region and the US too. And then also factoring in overall socioeconomic factors, you have those significances holding for low quality and medium quality too.

So just kind of summarize some of the next steps that I'm planning to take with this, doing some relative sensitivity models, especially the variance inflation factor for the regression is quite high, which doesn't happen often, but it has happened for a few of the cities. I want to eventually break down medium quality, high quality. Again, look at cost metrics as well, and then specific socio SVI factors. So kind of breaking down that overall thing. And this might be dependent on the region in the city. So I want to ideally avoid over complication within my models, but it's got to be kind of, again, region and location specific.

And then ultimately, I want to also look at kind of a neighborhood's approach of things by doing geographically weighted regression. I like to acknowledge my labs at UCR and then also lab I work with at UC Boulder. And with that, that's it for me. Thank you.

Thank you, Ryan. Yeah, Trader Joe's isn't high quality. Oh, how dare you say that. I'm kidding. Thank you for sharing your research is very interesting to look at those correlations and yeah, food availability. We're in California, most of us so we should have quality access to food for everybody. Anyway, thank you again for your presentation. Again, any questions you have for any of the three previous presenters, please put them in chat or you can direct message and hopefully you can share some of those question and answers at the end of the session. So to close out today's session, we are we have a presentation from Bo Yang and talking about GIS and remote sensing for wildfire and coastal monitoring.

GIS and Remote sensing for Wildfire and Coastal Monitoring

Speaker: Bo Yang – UC Santa Cruz

Abstract:

Dr. Bo Yang from UCSC will discuss the integration of multi-source remote sensing data, UAV mapping, and spatio-temporal modeling to monitor critical environmental changes. Covering applications from California wildfire tracking to coastal ecosystem assessment, this presentation highlights the use of drones, LiDAR, and high-resolution satellite data to enhance detection, prediction, and management of wildfire behavior and seagrass health. Discover how cutting-edge GIS tools and data fusion methods provide actionable insights, driving more effective environmental monitoring and resilience planning.

Transcript: Video Timestamp: 36:55

Oh, thank you, Erin. Let me share the slides.

Okay, great. Yeah, hello, everyone. I'm Bo Yang. I'm a new faculty at UC Santa Cruz. So before I joined UC Santa Cruz this year, I worked three years at San Jose State. That time I also am a GIS faculty and I'm working on a lot of the GIS project and we hold the GIS day every year. So I'm very glad to join the UC system and join the UC GIS day this week. So today I'm going to share like the summer or like how we use the GIS and remote sensing, especially the dual remote sensing, both the wildfire and the coastal seagrass monitoring.

So let's start with the data. I think we talked a lot about the vector data today and we know that in GIS we have both vector data and raster data. The previously raster data always kind of collected by the satellite remote sensing and the aerial remote sensing. The satellite is like very convenient but it's not very high resolution. The aerial data is high resolution but it's kind of always need a preparation. You need to rinse the air plant, hair the pallet to collect the data. And the drone remote sensing recently has become more and more trendy because it's convenient, can collect your own data at high resolution. We need to do a lot of processing using the GIS remote sensing techniques for our own drone data. So for the past few years supported by the NSF and NASA, we have been undergoing a bigger project, collaborated with some universities across the entire west coast of the North America, including both China and USA.

So we're trying to use the drone and the marine fieldwork to model the seagrass habitat. Because you know the seagrass habitat, it has roots and creates a habitat that anchors this area to provide lots of ecosystem functions for marine animals, algae and provide photosynthesis for carbon dioxide and other other kinds of global climate impact. But in recent years the seagrass habitat has facing a disease called wasting disease that has killed lots of the seagrass habitat along the west coast. So previously we always use manual sampling to monitor these kinds of diseases and do the quarantine and management effort. But from a few years ago we started using drone with optical remote sensing and GIS trying to model that. Because you know the previous remote sensing and GIS method, for example this is a landscape image, we can barely see the seagrass and this is San Francisco Bay Area. So you can see we can barely see external seagrass. But from the drone we can say because we use a very low optical mapping we can collect the image at very high resolution. We can see all the blades you can see very clear. And if we put them into GIS to also mosaic stitch to a bigger map we can have very high resolution and we can model them together. So for the past five years we have been working on collecting all the drone images. You can see for each kind of the area along the west coast we select a few kind of representing sites. We collect seagrass, select the imagery. So for each side we collect thousands of imagery and we stitch them together to a big giant map. And this map has very high resolution. We can see everything very clearly. And simultaneously because you know we collect some of the GCP points using our high performance GPS.

You know the GPS also has about half meter to like the one to two meter arrow level. So we also use some technique. For example we use transact tapes to put it on the inside to ground sampling locations. So we can see them from the drone image. So we will generate a big large map. It is very high resolution. You can see all the transact tapes. So you can know where we collect the ground truth data to validate our remote sensing data. So this is how we do the like the image collection field work and how to put them together. So you can see from each part we collect lots of the image from the drone using the low altitude mapping. And simultaneously over here you can see our field like the partners the marine team collect samples allows the transact. So you can see them. We can see here the wasting disease when it is healthy seagrass it's turning all too green. But when it's starting catching the disease it's starting losing the greenness and become like a darken and a liaison and a diet. So based on this color change we can use a remote sensing way to model the disease. It's much more efficient than the human manual sampling and also it can access somewhere that people cannot go. So we can see here here the overall workflow.

So the UAV team the GIS team will collect the UAV data do the image analysis each to a big map and generate a map based on the texture and also the greenness. We do the object orientated analysis generate a map and also we will use the inside sample. It's kind of a few points not the entire coverage but with every set we collect 30 points of the ground sampling try to validate this data and see how accurate we can get. So we can say we use a machine learning algorithm to do the orientated object orientated segmentation first. So we can segment into the pieces to know which kind of the plant of the seagrass we are working on and also it will be a similar plant of the one plant of the seagrass. We can know what the green is changed over the time. And we're also using some like the algorithm to calculate what we call the green leaf error index G-L-A-I. So using like G-L-A-I we could know like how much percentage the green color like the seagrass, healthy seagrass versus the total coverage of the seagrass and how much percentage are likely to catch a disease.

So here you can see when we generate a map we have the other points the sampling point marked on the zone imagery which can return as a zone imagery. So we do not just worry about the GPS like the BIOS, the GPS error about a half meter, but also we also collect the GPS points for like the referencing and the georegistration. So based on that we can see all the points. For example, if you look at here you can see here has a bigger point, bigger circle is representing the higher kind of disease level collected by the inside to the ground like sampling and they collected the laser sending into lab for analysis. And we can see the higher like area has higher like disease and we can see barely see it lose lots of greenness. So we run the static analysis. I will be explored but you can see all the p-value is statistically significant and we got pretty higher like the R-square value from 30 points to the 50 likes a percentage. So it's pretty good prediction result but still there is about half of the model on certain takes on expand. We are further working on it if there are some other reasons. We have a couple of the publications on this so if we want to learn more like we are happy to hear it's all open access. So if you want to learn more about our method you can refer to the publication. Also besides that we are also our lab program we are also using a drone to model the California welfare. So you can see here because usually if you see the welfare it will be all smoke or visible I see it's all smoke. But we use also use a low altitude through mapping for example you can see from the thermal mapping because the thermal can directly heat the since the heat. So we know where is the welfare and we can combine with some traditional JS method. For example we map the before and after fare multi-spectral at the lidar. So we know how much like so it is healthy before the fare and healthy after the fare. So we can see before the fare is quite a bit healthy from both RGB and near-infer for color and afterwards it's turning too dark most of them. And also we can use the lidar to map like how much volume biomass has been burned through this kind of fare. This is a prescribed fare like so implemented by the cow fare. So we know it's prescribed fare so we know that where and when they will be burned they will clean a buffer area. So we can do this pre and after in stock to the modeling and to help the future laps of more JS analysis.

Yeah that's about my next presentation and last I want to thank everyone and take this opportunity to show our UC San Marcos right now is working on new JS degree which I'm come here to to lead and it's a two-year degree and as a JS it's becoming more and more trending technology in the fare area. So we would like to cultivate more like the students to jump into this area and also like to do more kind of remote sensing JS, drone mapping, this kind of training and also to satisfy the basic need and the trending need for in the silicon van. Yeah and thanks very much that's my presentation.

Thank you very much for sharing your research. I'm just I'm reading the slide I'm really interested and I'm glad to see that this program is developing at UC Santa Cruz. So congratulations on that. So we do have a little bit of time for question answers. I'm trying to go through the chat to see to start with Justin's questions or any questions for Justin. But I am open if anyone wants to ask any questions verbally too. We can talk about that. If you want to raise your hand and raise any questions there's been a few. So the first question was regarding wildlife safety research from Alec about building footprint data. I think this is a pretty detailed question regarding this relation to landscape vegetation and so on and so forth.

Q&A:

Justin do you have anything any questions that came up that you want to clarify to the group at this time? No I think I pretty much answered the questions. I guess the only thing that's kind of annoying with the data sets is that they are only available to download at this date level. So whenever you try to download the data you have to download the entire state which has billions of building polygons. So we just kind of it will be kind of irritating to have to load the data if you process that into your study areas. That's kind of just one of the challenges of this data set. Yeah that's a really good point. These are really large data sets and it's difficult to subset those. So maybe somebody can come up with an API or script to pull those out per county or per region. But again, everybody geographic study is always different. So it could be county level, it could be Sierra Nevada versus San Joaquin Valley versus Northern California and are those really boundaries. So yeah these are interesting problems but I guess they're good problems to have that data is available. Okay and I'm just apologies for trying to get through all the questions.

All these are great presentations. Again thank you so much to all four of you and to the planning committee especially Amy and Danielle for keeping this going for the fourth year. Is this our year four? I don't even know. I should know because we start in 2020 so I guess this is year four. Okay so we have downloads here. I'm just going to start from here. Oh you're five. Okay there's some volunteer mapping questions from our solicitation to support some projects. So there's your link in the chat if you're interested in that. There's a discussion about open street map which is also kind of like a building data set. You can extract from that but it's just not easy but the data is actually pretty rich and useful once you figure out your study area. And then there's a question about geocoded locations of the supermarket grocery stores and I don't mean to skip anybody. I'm just kind of randomly going through these questions. Ryan do you have anything you want to clarify to the group about the questions that you received in the chat regarding the locations and any data sharing or other questions that came up? Yeah so I recently finished the geocoding of those grocery stores and actually I have convenience stores fast food and farmers markets as well. I honestly thought farmers markets would have a little bit more old but they didn't. Yeah I'm hoping to make those available to people because when I was doing it I was like I don't all these I have to you have to pay for access and things like that. It was quite frustrating so I do hope to have that. Unfortunately it will be like 2015 to 2017 so at this point it's like 10 years old and covid did alter a lot of things but yeah I'm hoping to have that stuff ideally in a repository next year and someone gave me a link for the UC GIS kind of base one so I'll definitely shoot for that one.

Well I have a question and I know that I don't want to derail your research at all. This is just more of a I'm asking this question from like just a random citizen on the street question. I've noticed in the last year or two a lot of fruit stands on the side of the road and I live in a really rural area and I'm like what I mean this is like off the road just nobody around and there's a fruit stand there. How I'm not aware of the quality of the food or that like how would that play into this type of research because it seems really accessible and it seems pretty popular.

Yeah so I think kind of fruit stands and the farmer's markets were ones where I was like oh those I think having access to those would be great but what I actually found was the seasonality of them but also just the fact that they're not open seven days a week and kind of limited hours was problematic especially for those that were lower income individuals you know working multiple jobs or shift work of 12 hours plus it seems like it seemed like the trends were just kind of going with the grocery stores in general. So there I definitely think that there is some good influence from them in general but I would have to do a little bit more of a localized approach which the study that I work with the sample set that I'm eventually going to bring in is more so based in Colorado so I'll maybe have like a kind of more localized approach there but right now I'm kind of half starting broad then slowly focusing in. Yeah that's again a challenge data and how you categorize these things and how they come into play. I'm trying to come up with questions for Bo for Alina again I thought Alina's research was great on the drought studies and congratulations on that and for both again I'm excited to see what you do as you said especially with the inflammation of drone data collection and integration into research so

if there aren't any other questions I'm trying oh there's something else here okay there's one specifically for regression analysis

um okay there's kind of a random question um about utilizing gi uh ai and gis work so does anyone wanna um I presented something about what what we did with some ai um on Tuesday I don't know if anyone else has anything to share about use of ai for gis yeah we'll use some ai for the image processing because right now Alexa uh we use some traditional method for image segmentation and classification and there is more and more deep learning model especially with the ai flourish this year so with right now we are trying to use a mental ai has a segment anything model which we call as a is a pre-trained model so it has been trained for billions of image likely to recognize the image which part so it's ready to use right now uh so if you can yeah if you can search for a mental ai as ai model is very useful for the image segmentation that we call that which is building which is a zone it's very useful

well great yeah that's um there's a lot to keep on with so we all know but I know these tools these deep learning tools are can be very useful for replicating research well if there aren't any other questions again thank you for sticking that out for day three on our first session our next session is at 1 p.m. I forgot to mention please fill out the survey if you can there is a link that came through here if you want to fill out the survey help us keep this going and um to support the GIS community and the University of California again thanks again happy GIS day and we'll see you all or GIS week I don't know we'll see you all soon