Title: BMSB final project
Name 1: Yaxuan Zhang
Name 2: Maochuan Wang
Date: 05/01/2023
Project Github repository of all code:
https://github.com/MaochuanW/GIS5572/tree/main/final_project
Link to Final ArcOnline MapViewer:
https://umn.maps.arcgis.com/apps/mapviewer/index.html?webmap=9bd531509d5d499e89c29ae3d3900228
Link to Video Overviewing Each Aspect of Your Project:
https://drive.google.com/file/d/1UWIAICbD0yDIlzetAVIZS6cJ4NQTbfEq/view?usp=sharing
Link to Your Lab 4 Peer Reviews:
https://drive.google.com/file/d/1u9_ANNPiSFNA5o8XEjlz0eUWrYHkrKyP/view?usp=sharing

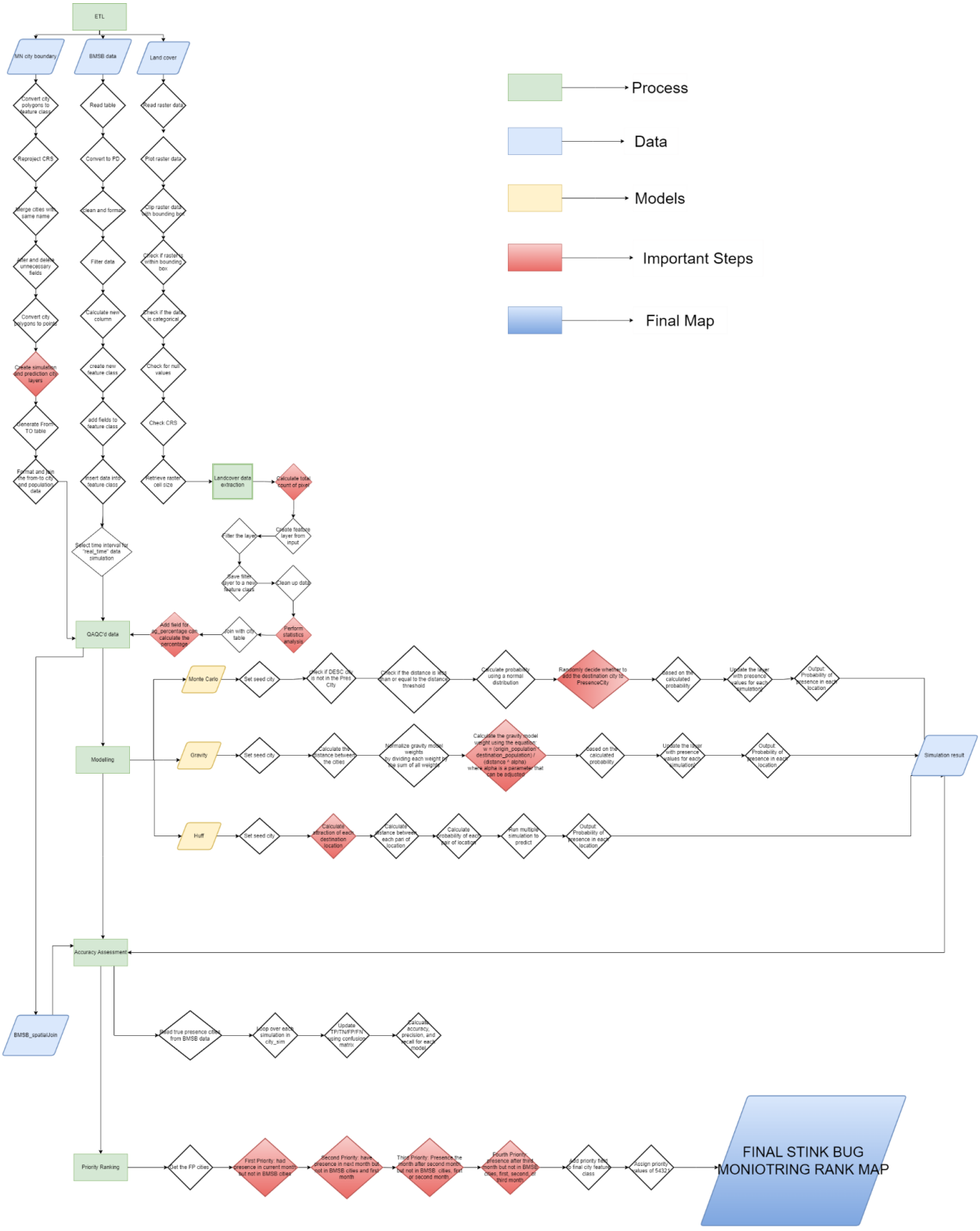
Abstract (50 words)
This project aimed to develop a geospatial decision support system for predicting and monitoring the spread of brown marmorated stink bug across the state of Minnesota. The system integrated spatial models to simulate and predict the spread of BMSB. For the end results, this project uses different simulation models to create a prioritized list of cities for BMSB monitoring.

Problem Statement (100 words max)
Invasive species are detrimental to our ecosystem. It affects the biodiversity, food chain, and local biological niche. To preserve the natural beauty of Minnesota and its ecosystems' stability from attacks from invasive species, building a real-time update of invasive species monitoring web application could be a crucial step to preventing ecosystem destruction. The objective of this project is to build a pipeline to prioritize Brown Marmorated Stinkbug Invasive species monitoring. The online system needs to show a map and a ranked list of cities that we should prioritize for monitoring and detection of spread on a regular basis across Minnesota.

Input Data (no text, only table)

#	Title	Purpose in Analysis	Link to Source
1	City, Township, and Unorganized Territory in Minnesota	Raw input dataset for city boundary	https://gisdata.mn.gov/dataset/bdry-mn-city-township-unorg
2	NLCD 2019 Land Cover, Minnesota	Raw input dataset for land cover to calculate the percentage of agriculture by each city	https://gisdata.mn.gov/dataset/biota-landcover-nlcd-mn-2019
3	Brown Marmorated Stink Bug Survey Data, Minnesota	Raw input dataset for stink bug records as the ground truth value used for model evaluation	https://gisdata.mn.gov/dataset/biota-bmsb

Data Flow Diagram(s) for System (0 words)
This should be extremely detailed and characterize the full end-to-end behavior of your system.


ETL
MN city boundary
BMSB data
Land cover
Convert city polygons to feature class
Reproject CRS
Merge cities with same name
Filter and delete unnecessary fields
Convert city polygons to points
Create simulation and prediction city layers
Generate From-TO table
Format and join the from-to city and population data
Read table
Convert to FD
Clean and format
Filter data
Calculate new column
create new feature class
add fields to feature class
Insert data into feature class
Select time interval for "real_time" data simulation
Read raster data
Plot raster data
Clip raster data with bounding box
Check if raster is within bounding box
Check if the data is categorical
Check for null values
Check CRS
Retrieve raster cell size
Landcover data extraction
Calculate total count of pixel
Create feature layer from input
Filter the layer
Save filtered layer to a new feature class
Clean up data
Perform statistics analysis
Join with city table
Add field for ag_percentage can calculate the percentage
QAQC'd data
Process
Data
Models
Important Steps
Final Map
Modelling
Monte Carlo
Set seed city
Check if DESC_City is not in the Pres City
Check if the distance is less than or equal to the distance threshold
Calculate probability using a normal distribution
Randomly decide whether to add the destination city to PresenceCity
Based on the calculated probability
Update the layer with presence values for each simulation
Output Probability of presence in each location
Gravity
Set seed city
Calculate the distance between the cities
Normalize gravity model weights by dividing each weight by the sum of all weights
Calculate the gravity model weight using the equation: wi = (origin_population * destination_population) / (distance ^ alpha) where alpha is a parameter that can be adjusted
Based on the calculated probability
Update the layer with presence values for each simulation
Output Probability of presence in each location
Huff
Set seed city
Calculate attraction of each destination location
Calculate distance between each pair of location
Calculate probability of each pair of location
Run multiple simulation to predict
Output Probability of presence in each location
Simulation result
Accuracy Assessment
BMSB_spatialJoin
Read true presence cities from BMSB data
Loop over each simulation in city_sim
Update TP/TN/FP/FN using confusion matrix
Calculate accuracy, precision, and recall for each model
Priority Ranking
Get the FP cities
First Priority: had presence in current month but not in BMSB cities
Second Priority: have presence in next month but not in BMSB cities and first month
Third Priority: Presence the month after second month but not in BMSB cities, first or second month
Fourth Priority: presence after third month but not in BMSB cities, first, second, or third month
Add priority field to final city feature class
Assign priority values of 5432
FINAL STINK BUG MONIOTRING RANK MAP

Model Comparison

#	Model Name	Evaluation Metric 1 (Accuracy)	Evaluation Metric 2 (Precision)	Add more Eval metrics as needed (Recall)	Rank Score of Model
1	Monte Carlo	0.92974	0.19512	0.22857	3
2	Gravity	0.94379	0.31429	0.31429	2
3	Huff	0.95082	0.40541	0.42857	1

<Monte Carlo> (50 words max)

Monte Carlo model: We first use a distance threshold (0.005 quantiles) to decide if stinkbugs could travel between two cities. For cities within the threshold, we set the transition probability as a normal distribution (mu = 0.05, sd = 0.005).

<Gravity> (50 words max)

Gravity model: The transition probability between two cities is calculated as:
Probability = (from_population * to_population)/(distance ** alpha). From city A (presence) to city B (absence), we randomly generate a number, if the number is smaller than the probability, stink bugs can travel to city B.

<Huff> (50 words max)

Huff model: The transition probability between two cities is calculated based on the attractiveness of the destination city:
Probability = (to_population * percentage_of_agriculture_land)/(distance ** alpha). From city A (presence) to city B (absence), if the random generated number is smaller than the probability, stink bugs can travel to city B.

Recommendation for Decision-making (50 words)
What model do you recommend be used for decision-making and why?

Based on the evaluation of the three models, the Huff model showed the highest accuracy, precision, and recall. Thus, we recommend using the Huff model for decision-making. The other models can also be viable in certain situations.

Reflection: What did you learn? What would you do differently if you did this again?
Yaxuan: I learned a lot about the spatial interaction model and how to run a spatial simulation. I also learned how to use cloud run to run the codes and publish to arconline.
If I did this again, I would have a more clear understanding how the whole pipeline works. So I will preprocessed my data and did qaqc based on what data i need for the following simulation

and analysis. I would also include more external datasets to define the attractiveness of city to get a better simulation model.

Maochuan: I learned a lot about spatial data modeling using python and arcpy, google cloud run set up and dockerfile..etc. One thing that could be done differently is to consider using more recent data and updating the model parameters regularly to improve its accuracy. Additionally, it may be helpful to incorporate data from other sources, such as weather data, to further refine the predictions. Finally, using more advanced machine learning algorithms could be explored to potentially improve the accuracy and efficiency of the models.

References

Minnesota Geospatial Information Office. (2021). Boundary of Cities, Townships, and Unorganized Territories in Minnesota [Data set]. Minnesota Geospatial Information Office. https://gisdata.mn.gov/dataset/bdry-mn-city-township-unorg.

Minnesota Department of Natural Resources. (2020). NLCD 2019 Land Cover, Minnesota [Data set]. Minnesota Department of Natural Resources. https://gisdata.mn.gov/dataset/biota-landcover-nlcd-mn-2019.

Minnesota Department of Agriculture. (2021). Brown Marmorated Stink Bug Survey Data, Minnesota [Data set]. Minnesota Department of Agriculture. https://gisdata.mn.gov/dataset/biota-bmsb.

Appendix (include your original spec)

Copy of Software Requirements Specification (SRS) Template.docx (1).pdf