Certificate I: Understanding AI and Machine Learning in Africa

Course AIML02: AI and Machine Learning in Africa

Module 02: Application Case Studies

Lecture 03: Agriculture

Carnegie Mellon University
Africa

Learning Objectives

1. Identify the advantages of using mobile devices to collect survey data
2. Explain how computer vision can be used to diagnose different types of disease
3. Explain how samples of diagnostic data can be interpolated to generate maps showing the density of disease
4. Explain how diagnosis and mapping can be combined to yield more effective ways of monitoring and tracking crop disease
5. Explain how this approach can be used to optimize the use of survey resources

Lecture Contents

Crop Disease Monitoring in the Developing World

This agricultural case study is based on an article by Quinn (2013):

Computational Techniques for Crop Disease Monitoring in the Developing World

It demonstrates how data analysis techniques can be used to improve the

- Speed
- Accuracy
- Cost-efficiency

of conducting surveys of crop disease in developing countries using cassava and banana crops in Uganda as examples

Crop Disease Monitoring in the Developing World

The economies of many developing countries are dominated by small-scale agriculture

- Smallholder and subsistence farmers are responsible for most production
- Utilizing relatively low levels of agricultural technology

https://www.ifad.org/thefieldreport/

Crop Disease Monitoring in the Developing World

- Disease among staple crops presents a serious risk, with potentially devastating consequences
- Monitoring the spread of crop disease is essential
  - Allows for targeted interventions to be planned
  - Provide early warning of risk of famine

https://www.icco-cooperation.org/en/wp-content/uploads/sites/2/2019/09/Final-stars-rwanda-smallholder-farmer-report-gecomprimeerd.pdf

Crop Disease Monitoring in the Developing World

Survey teams are sent out to farms to make assessments of crop health

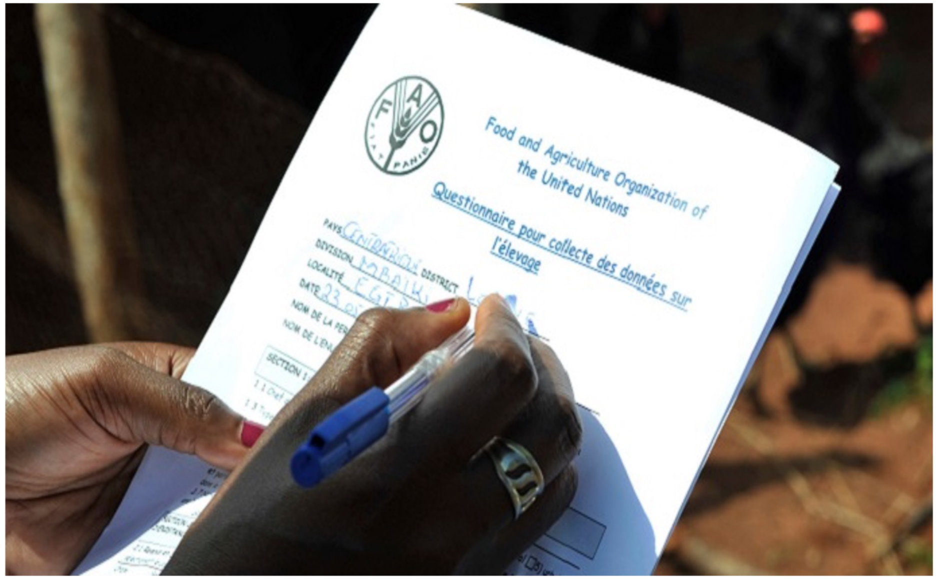
- Visit multiple sites
- Carry out disease diagnosis
- Fill in paper surveys
- Use the survey data on returning to base to infer the incidence and spread of the diseases

https://blogs.worldbank.org/opendata/working-hard-and-not-being-counted-evidence-sub-saharan-africa-redefining-employment

Crop Disease Monitoring in the Developing World

Sending teams of trained agriculturalists to visit farms and assess of crop health is expensive and inefficient

- Suitably-trained trained staff are scarce
- Transport is difficult and expensive
- Coordination of paper reports is time-consuming

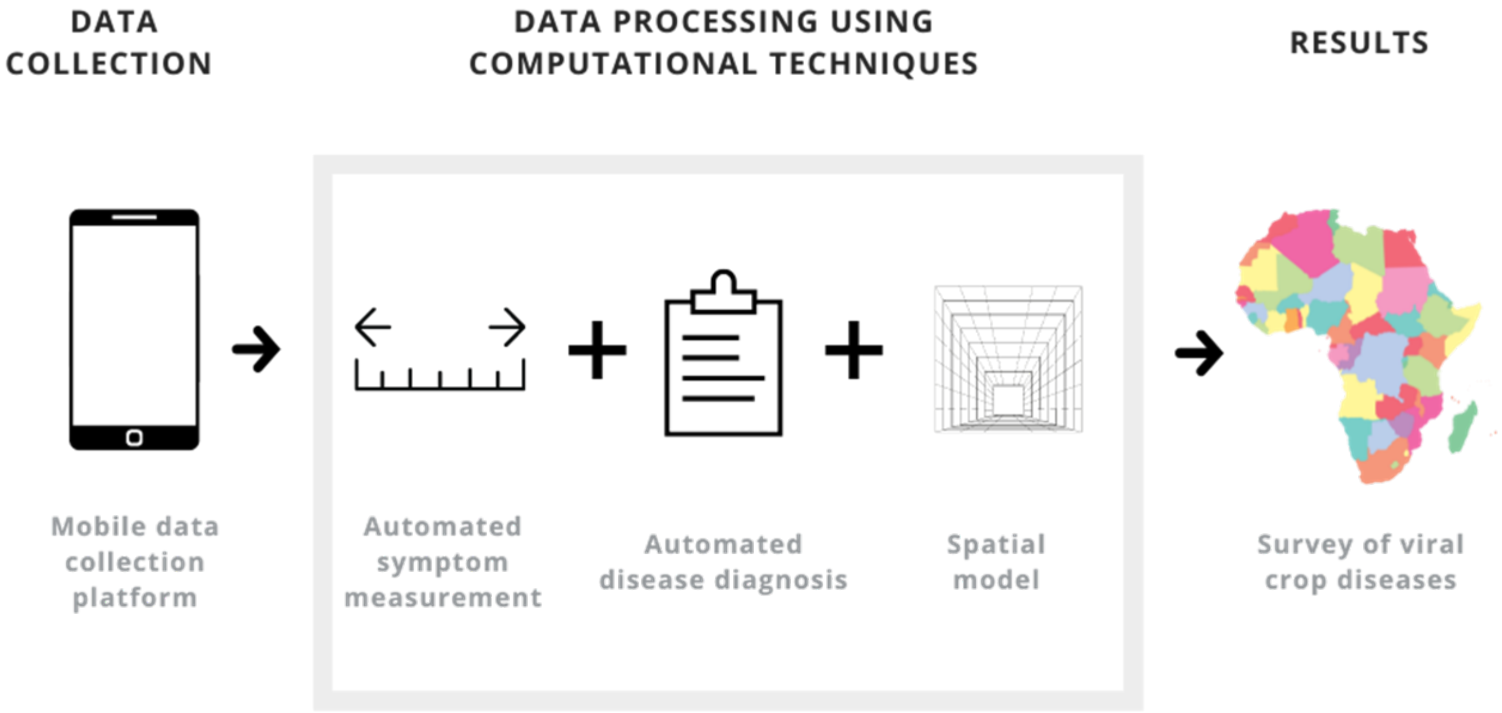


https://www.fao.org/statistics/data-collection/en/

Crop Disease Monitoring in the Developing World

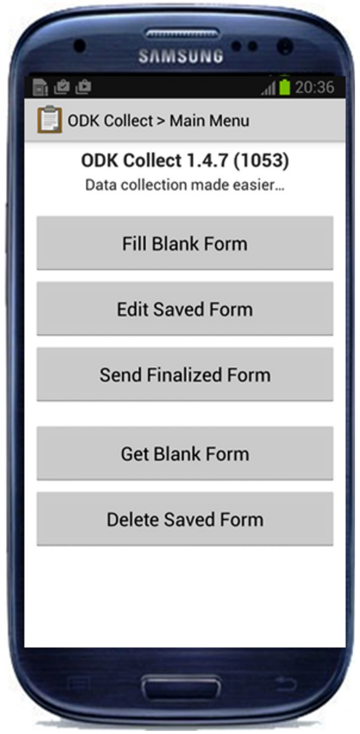
- The challenge is to identify and track the spread of viral crop diseases in a manner that is cost-effective, timely, and accurate
- The target article sets out to demonstrate that data collected using digital surveys on inexpensive Android phones can be used to provide the required information
  - Automated disease diagnosis
  - Monitor the incidence and spread of viral crop diseases
  - Optimizing the use of scarce survey resources

Crop Disease Monitoring in the Developing World



Mobile Data Collection

- Data is collected on low-cost Android phones
- The Open Data Kit (ODK) is used to create digital surveys
  - Similar to data found on paper-based surveys
  - Allows for collection of images and GPS coordinates



https://kopokuasare.wordpress.com/2015/11/26/data-collection-with-odk/

Mobile Data Collection

Advantages

– Reduced time needed for data-entry

– Results are immediately available

– Data collection by workers with only basic training

– Images can be assessed remotely by experts or by software on the phone

Collecting data using a mobile phone for a survey of cassava plants

Automated Diagnosis and Symptom Measurement

A national survey would require expert inspection of around 20,000 plants

Assessing of the status of the disease and the level of symptoms

Cross-sections of cassava roots showing necrosis with scored denoting severity

https://www.sciencedirect.com/science/article/pii/S2352340920310647#fig0004

Automated Diagnosis and Symptom Measurement

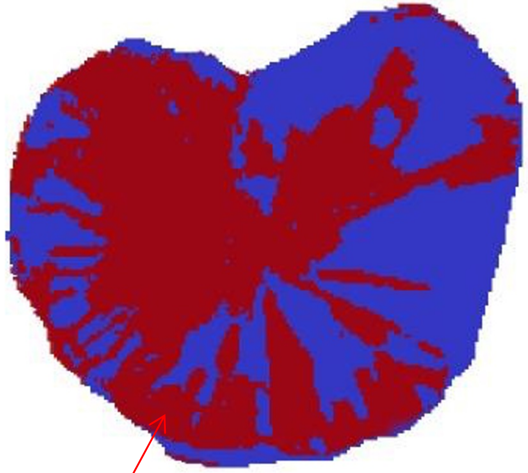
A national survey would require expert inspection of around 20,000 plants

Assessing of the status of the disease and the level of symptoms

Whiteflies on cassava leaves

https://plantvillage.psu.edu/diseases/cassava-whitefly

Automated Diagnosis and Symptom Measurement

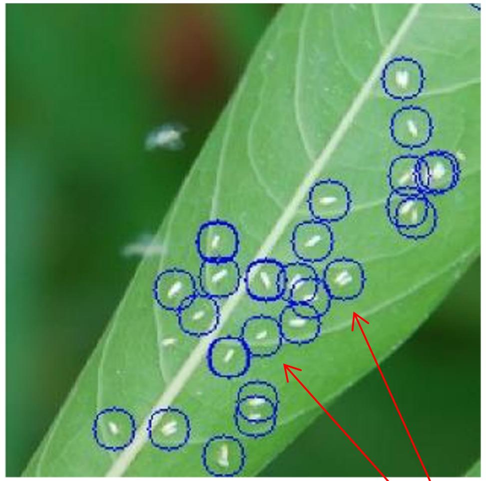


(Quinn, 2013)

Cassava root with necrosis
caused by cassava brown streak disease

Classification of pixels to measure the
degree of necrosis

Automated Diagnosis and Symptom Measurement



Whitefly count on cassava leaf

Automated Diagnosis and Symptom Measurement

(Quinn, 2013)

Healthy banana leaf

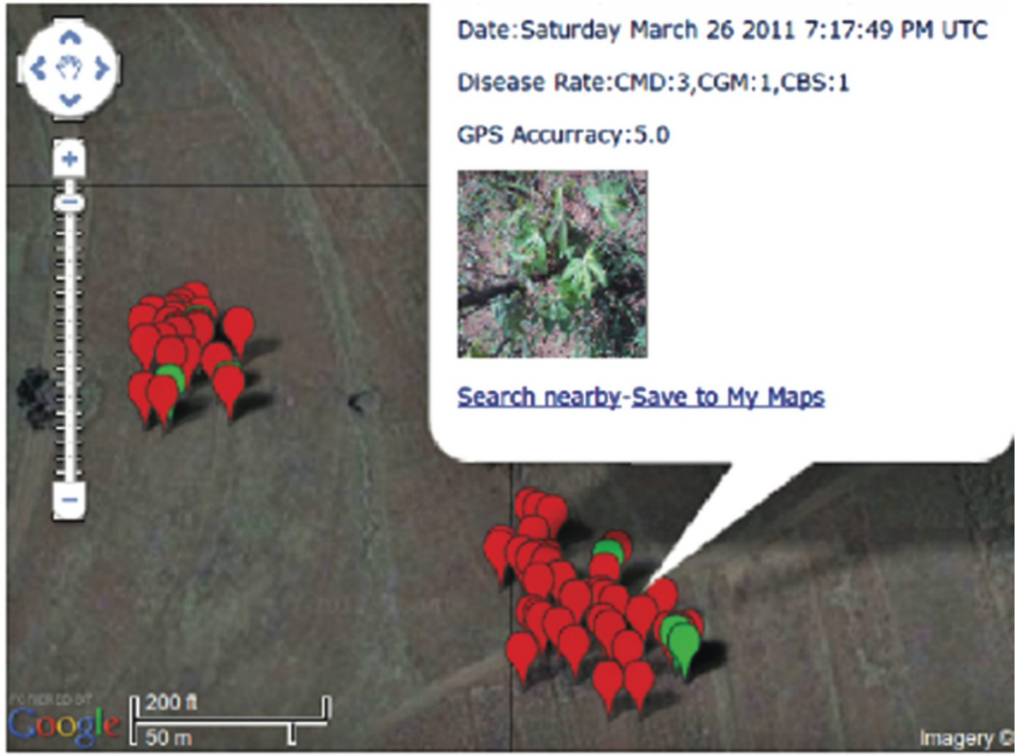
Bacterial wilt

Black sitagoka disease

Automated Diagnosis and Symptom Measurement

Software on the phone detects cassava mosaic disease from leaf appearance

Constructing Spatial Maps of Crop Diseases

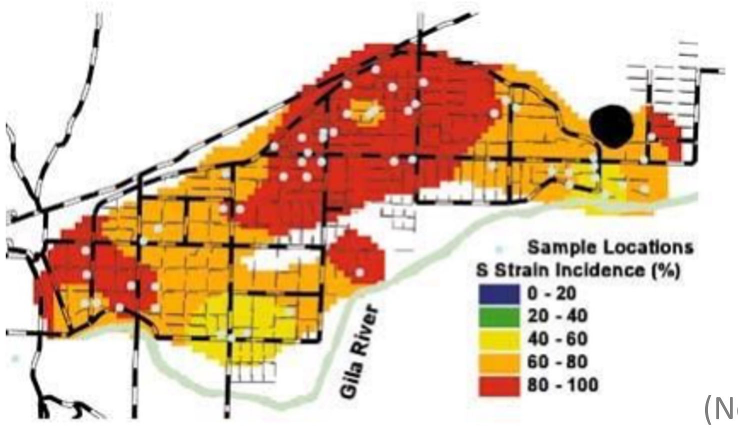


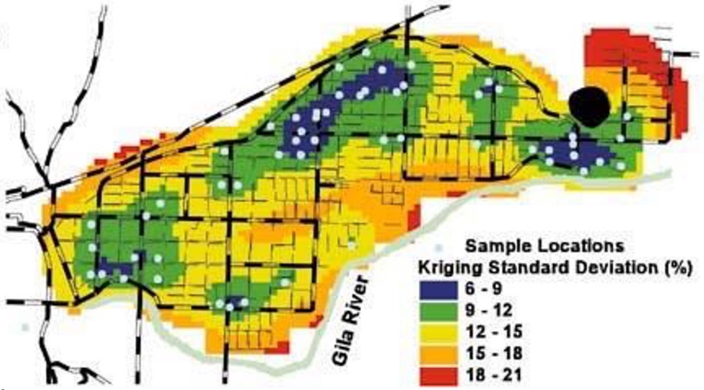
Data collected with the phone is used to create a risk map using the GPS coordinates collected as part of the survey

Constructing Spatial Maps of Crop Diseases

Interpolation between samples using Gaussian process regression

Gaussian Process Regression (GPR) is a form of non-parametric machine learning, i.e., it is not limited by a functional form.
It calculates the probability distribution over all admissible functions that fit the data. For example, when provided with geographical weather data that is incomplete, a GPR model can be used to generate weather data for unobserved locations





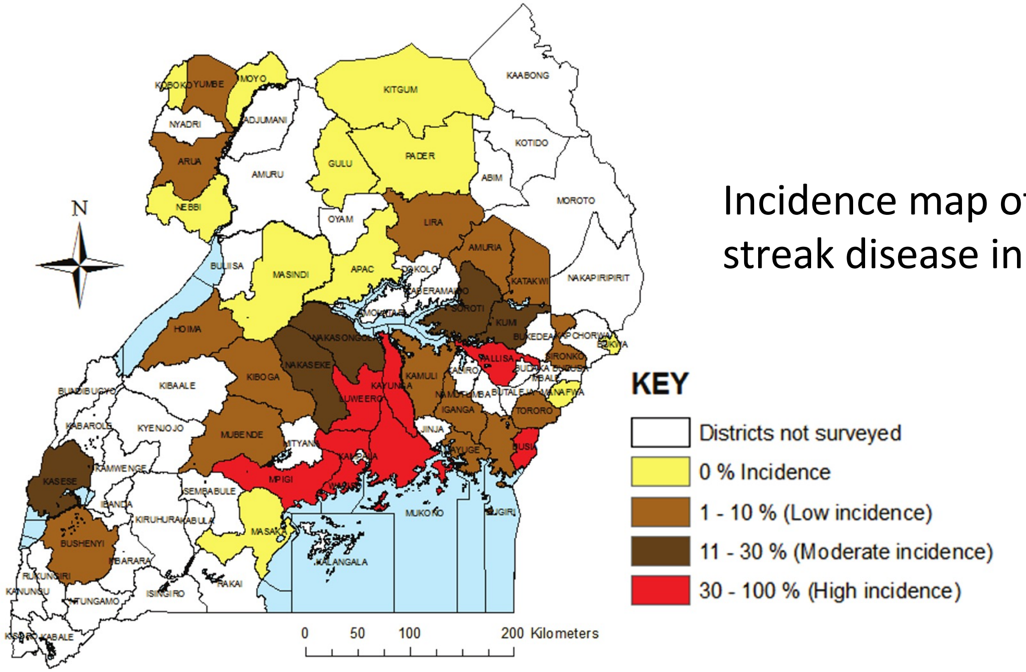
(Nelson et al., 1999)

A map showing estimates of the incidence of *A. flavus S* strain, Texas Hill area, Yuma County, Arizona

An estimate confidence map of *A. flavus S* strain, Texas Hill area, Yuma County, Arizona

Constructing Spatial Maps of Crop Diseases

Interpolation between samples using Gaussian process regression



Incidence map of cassava brown streak disease in Uganda, 2009

(Quinn, Leyton-Brown, & Mwebaze, 2011)

Combining Diagnosis and Mapping

- Diagnosis and mapping are normally performed independently
  - The diagnosis is not usually formally coupled with estimates of the risk of the disease
- However …
  - The risk estimate from the map provides a very useful prior
  - Individual diagnoses can be used to update the risk map

Known information that can impact of the diagnosis.

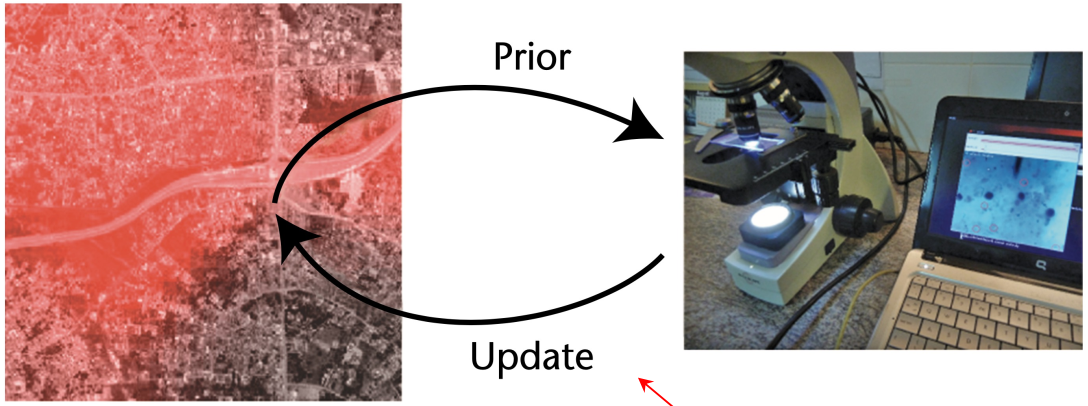
Combining Diagnosis and Mapping

- They can be combined here because
  - The data collection devices are networked
  - Their locations are known
- This combined inference of spatial disease density and diagnosis in individual cases can be done with multi-scale Bayesian models

Recall, we met the concept of probabalisitic Bayesian models in course AIML01, Module 2, Lecture 3 on statistical machine learning

Combining Diagnosis and Mapping

This is a win-win situation: it improves the accuracy of the risk map and of individual diagnoses because the uncertainty in both tasks is jointly-modeled



(Quin et al, 2014)

Recall, we mentioned this in Module 1, Lecture 2

Optimizing Survey Resources

- A probabilistic spatial or spatio-temporal model can also be used to determine which locations would be most informative for collecting new data
- Not possible in the traditional paper-based survey system
  - Data entry happens after the surveyors return home from their trip to the farms
- The approach here allows models to be learned in real-time, as data is collected

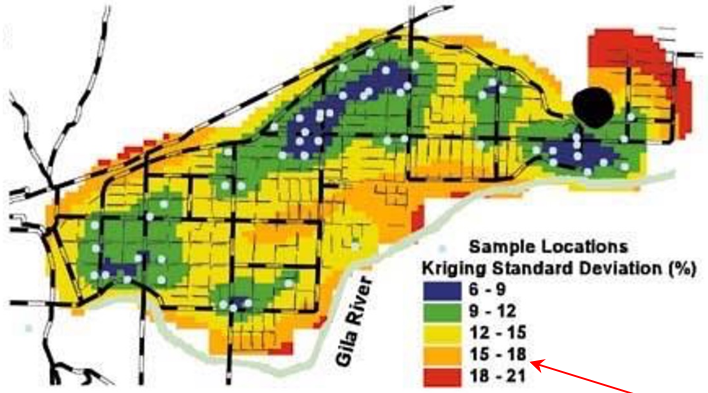
Optimizing Survey Resources

This is essentially active learning:

Collect data from locations in which the model has the lowest confidence

The locations that provide the most information

In information theory, information corresponds to reduction in uncertainty



(Nelson et al., 1999)

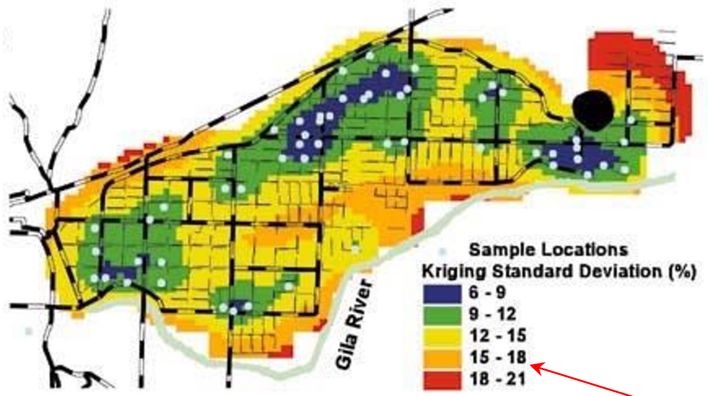
An estimate confidence map of *A. flavus S* strain, Texas Hill area, Yuma County, Arizona

Collecting data in these low confidence locations provides the most useful information for updating the risk map

Optimizing Survey Resources

This would suit situations where phones are given to agricultural workers across the country

Rather than by experts travelling in the field, effectively crowd-sourcing the survey



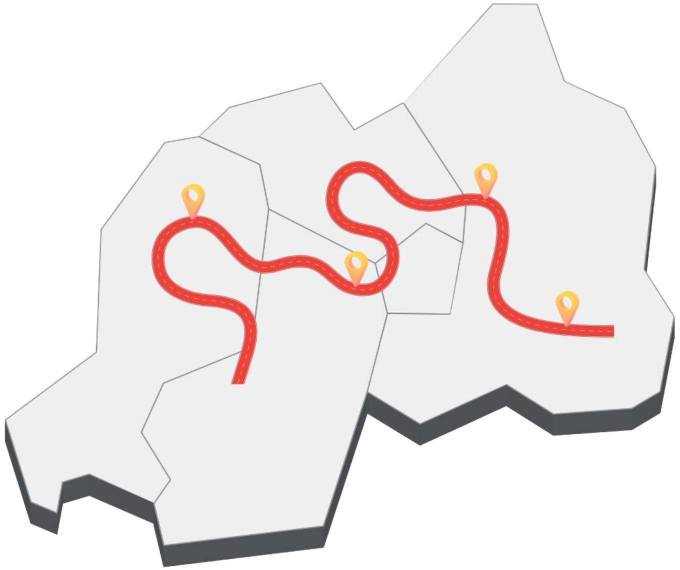
(Nelson et al., 1999)

An estimate confidence map of *A. flavus S* strain, Texas Hill area, Yuma County, Arizona

Collecting data in these low confidence locations provides the most useful information for updating the risk map

Optimizing Survey Resources

- The model can also be used to guide experts in the field
  - Limited amount of time
  - Limited budget
  - Limited number of survey visits
- So that they collect more valuable data
- While keeping fixed their budgeted number of samples or visits

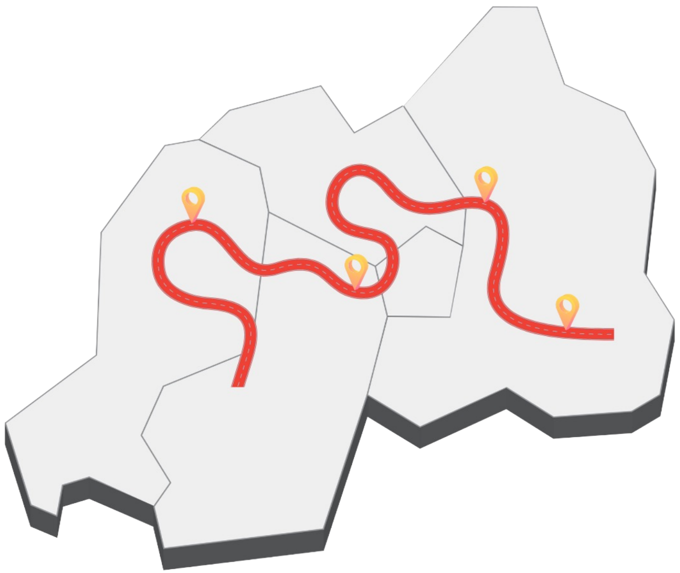


Optimizing Survey Resources

- Difficult optimization problem if the experts can travel any path in a given road network with some given budget

- However, the road network is often sparse in rural parts of the developing world

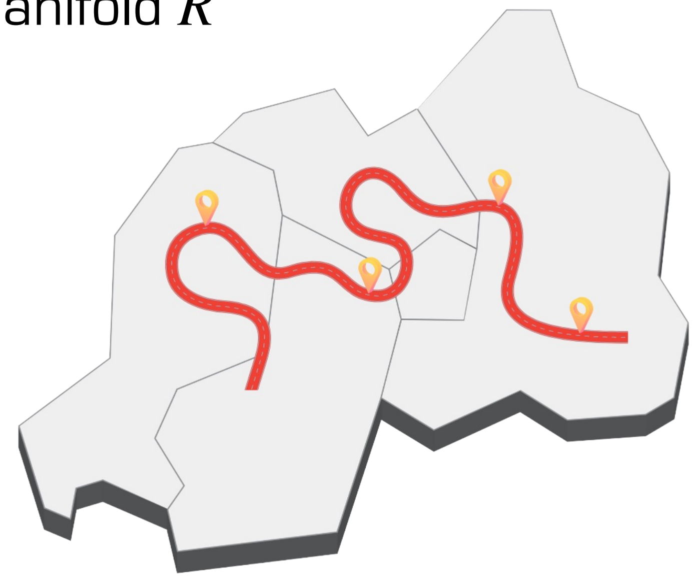
- This makes it reasonable to assume that survey teams will follow a given route



Optimizing Survey Resources

×W !"#$%&'()*+,-./0123456789:;ÿ<=>?@ way of representing some high-dimensional structure in a lower-dimensional space, thereby simplifying the analysis of that structure

- This route corresponds to a one-dimensional manifold R within the spatial field



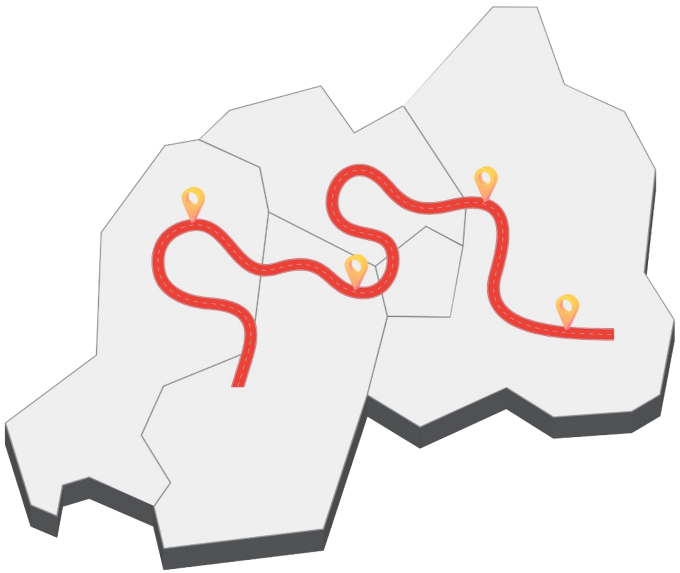
- With a fixed survey budget allowing k stops, we want to identify the points along R that maximize the information from the survey

- Under the sparse network constraint, optimization is tractable with a Monte Carlo algorithm

A Monte Carlo algorithm is a mathematical technique based on statistical random sampling to estimate the optimal solution to some problem.

Optimizing Survey Resources

- The optimal location for the next sample
  - Is recomputed after each stop
  - Given on the spatial model
  - Given the most recent observation



Discussion

1. Inexpensive data collection method compared to use of paper-based surveys
2. Automated disease diagnosis means that experts don't need to be on site for crop evaluation
3. Probable diagnosis of plants with known locations can be accomplished using the risk map
4. Optimized data collection by survey teams in the field

Discussion

A spatial model is not able to provide disease forecasts that could be used to inform decision-making to implement disease prevention measures

However, the approach can be extended to include temporal information, yielding a spatio-temporal model

A spatio-temporal model is one which contains data elements of time and space, making it possible to describe an event as having occurred at a given time t and in location x

Lecture Summary

1. The ability to track the spread of viral crop diseases in developing countries can help governments to put in place measures to mitigate disasters such as famine, and to do so effectively and in a timely manner

2. The approach described in this case study makes use of digital surveys administered using Android phones to collect data which is then analyzed using novel data analysis techniques to improve the speed, accuracy and cost efficiency of crop disease surveys

3. The model makes it possible to generate real-time diagnosis of crop diseases, map the location of crop disease outbreaks, and make optimal use of fixed budgets given to survey teams by identifying optimal survey locations

Recommended Reading

Quinn, J. (2013). Computational techniques for crop disease monitoring in the developing world. In Advances in Intelligent Data Analysis XII (pp. 13–18). Berlin, Heidelberg.

https://doi.org/10.1007/978-3-642-41398-8_2

References

Nelson, M. R., Orum, T. V., Jaime-Garcia, R., & Nadeem, A. (1999). Applications of geographic information systems and geostatistics in plant disease epidemiology and management. Plant Disease, 83(4), 308–319. https://apsjournals.apsnet.org/doi/10.1094/PDIS.1999.83.4.308

Quinn, J. A., Leyton-Brown, K., & Mwebaze, E. (2011). Modeling and monitoring crop disease in developing countries. https://www.aaai.org/ocs/index.php/AAAI/AAAI11/paper/viewFile/3777/4083