Using Satellite Data for Area Yield insurance

Rose Goslinga,
Co Founder and CEO
We develop and distribute insurance and farm advisory for small scale farmers.

611,040 farmers used our products in 2017 across 9 countries.
But farmers don’t like buying insurance

I don’t trust insurance companies

Let’s wait and see

I’ve managed all these years. Why would I spend money now?
Which is why we package insurance with products farmers actually want like seeds, fertilizer and credit.
As part of our underwriting process we collect unique farmer data that allows us to advise farmers.

This proprietary data allows for synchronized, local advice that provides value to all farmers every season, and not just to those that are compensated.
Harnessing data to understand smallholders

- Customer distribution by variety and size across the country
- Distance from farm to shop
- Farmer age and sales profile
At scale, Pula is an ecosystem of agribusiness with insurance as the initial entry point.

With each party both receiving and providing value along the chain.
3 Variations of Insurance Bundles

**Seeds + Insurance** to cover the planting stage
- Weather index cover using satellite data
- Operational in Malawi, Zambia, India

**Fertilizers + Insurance** to cover the whole season
- Area Yield index cover using ground + satellite data
- Offered in Nigeria

**Credit + Insurance** to cover the season and **harvest income**
- Area Yield index cover using ground + satellite data
- Operational in Nigeria, Kenya, Uganda, Rwanda, Malawi, Zambia, Tanzania
Data is the backbone of each product

**Seed insurance** uses satellite data algorithms to price premiums and enables farmers to sign up on any day and any location in a country.

**Fertilizer and credit insurance** uses a combination of proprietary ground yield data collection processes and remote sensing data to estimate yields.

Where index insurance has struggled with basis risk, we focus on products that limit basis risk (area yield index and start of season weather index) and price in the cost of this risk as part of the premiums.
Our go-to-market strategy is through public & private partners, targeting 1.2mln in 2018

**East Africa 2018**
705k farmers (59%)
100% Yield index

**Nigeria 2018**
266k farmers (22%)
100% Yield index

**India 2018**
11k farmers (1%)
Weather Index

**Southern Africa 2018**
228k farmers (19%)
Weather Index + Yield Index
Yield index insurance has been key to our growth

Farmers and their creditors want comprehensive, simple to understand coverage
Challenges to scale:

1) Yield insurance is in demand due to **comprehensive nature:**
   - Governments want it *everywhere*
   - Fertilizer companies want it but again, they want it *everywhere*

2) Our Yield product based on CCE’s is *labor and travel* intensive
   - 92% of unit cost

3) Historical data to base pricing on is sparse

4) Key challenges to solve:
   - Manage the cost of CCE’s
   - Gather a baseline of measurements
Crop Cut Placement: A Machine Learning Model
Crop Cut Selection: The Problem

Current Operations
- Sampling minimum of 25 crop cuts per LGA (Local Government Area)
- LGA’s are not defined based on agricultural yield, so low correlation between them, and many need to be sampled

Target Issue
- A single crop cut can cost up to 36 USD
- Sampling every LGA becomes prohibitively expensive with expansion
Proposed Solution 1: Yield Data

Figure 1 individual yield prediction quality is based on Random Forest Model showing low predictive potential

Figure 2: LGA yield prediction potentially more promising
Proposed Solution 2: Evapotranspiration

EARS used the yield data provided by PULA in order to examine the possibility to assist the optimization of crop cut experiments using remote sensing data.

EARS (Environment Analysis & Remote Sensing Group) is the provider of the Meteosat Relative Evapotranspiration dataset with:

- Index Insurance experience in 18 countries, mainly in Africa
- 35 years of hourly / daily satellite data
- Full coverage of the African continent at 3 km spatial resolution
Proposed Solution 2: Evapotranspiration

Lack of signal:

For example, a country-wide dataset that shows non-drought yields is not as valuable to a model as a year with some success and some drought spots.
Target Solution
- Goal: Create districts to replace LGA’s, that cover larger areas yet are more correlated
- Fewer districts that are more representative of their regions mean fewer crop cuts needed, dramatically lowering overall costs

Methodology
- Use K-Means (unsupervised machine learning algorithm) to create districts
- Based on latitude, longitude, historic rainfall
- Weighed appropriately to create mutually exclusive regions
- Normalized for planting seasons, precipitation caps, 30+ years of data (CHIRP)
How It Works, Part 1: The Data

1) Data Preparation:
   Extraction (Climate Hazards Group InfraRed Precipitation), Normalization (daily precipitation cap, restriction to planting season, etc.)

2) Data Maps:
   Combine 30+ years of data, data clean-up (remove errors, N/A values)
Guide to Clustering:

A) Pick X number of districts you want to create.
B) Randomly choose X points, use those as the centre of districts, and allocate all points to the nearest chosen point in X.
C) Repeat Step B (for a predefined number of times) until a set of X points is found to have the minimum total amount of distance from those X points.
D) The allocation of all points to chosen X points is then the generated districts.
K Means:
Machine learning (trial and error, varying results), Unsupervised (no “correct” districts), defined number clusters, scaling.

- a) Clusters (ie. Number of Districts) needs to be decided in advance. Example on right shows results of various numbers of clusters. Decision needs to balance not having too many districts, and being able to pick up on drought.

- b) Example: For Babban Gona, our recommendation was 14 districts, and for Nirsal it was 32, since farmers were in a greater area for Nirsal. For Babban Gona, 14 districts were required to detect lower yield areas (results in later slides).

- c) Note because of the nature of machine learning, there is great flexibility of the program such as generating results for custom regions, custom years, and district shaping.

Cluster analysis for 8, 10, 12, 14 Districts, respectively:
Unit areas of insurance: Past and Future

Based on administrative boundaries and operational limitations

Based on machine leaning climatological features

Manage the number of areas → Manage the number of CCE’s → Scale faster
Validation: Statistical Studies

**Uniform Distribution:**
Would indicate area was generated effectively at random, similar to LGAs from before.

**Bimodal Distribution:**
Would indicate AEZ needs to be split further into two separate AEZs.
Validation: Empirical Studies in Kenya

Note in **Cluster 2 and 7**, the model is able to very clearly detect a payout, mirroring field experience in 2016 locations.
Application in North Central Nigeria

From 22 administrative boundaries to 5 agro ecologies
Cost was \( 22 \times 30 \text{ CCE} \times 36 \text{ USD} \)
Budget spend is now \( 5 \times 75 \text{ CCE} \times 36 \text{ USD} = 43\% \text{ cost reduction} \)
Using similar methodology, and real time vegetation estimates to understand potential poorly performing areas and ensure CCEs are placed proportionately, with the goal to limit basis risk.
Thank you.

Questions?