VITAL SIGNS

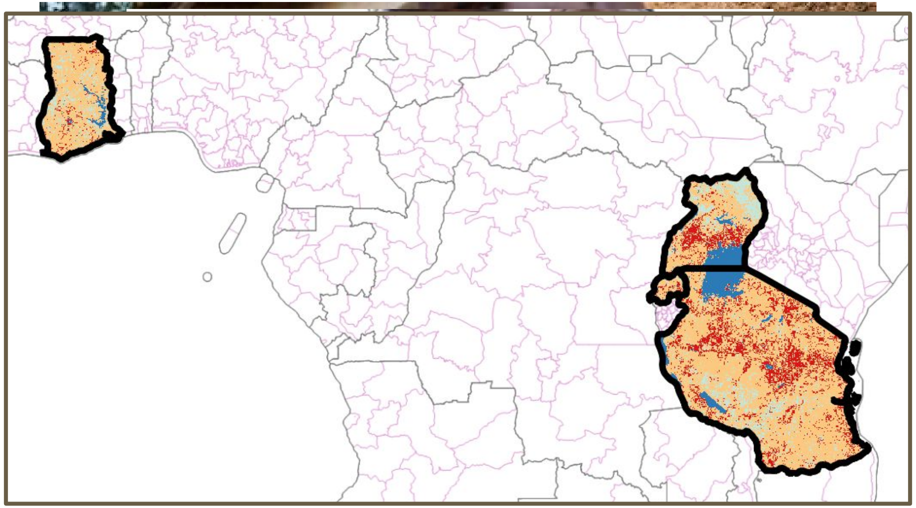
CONSERVATION INTERNATIONAL

Leads: Matt Cooper, Tabby Njung’e

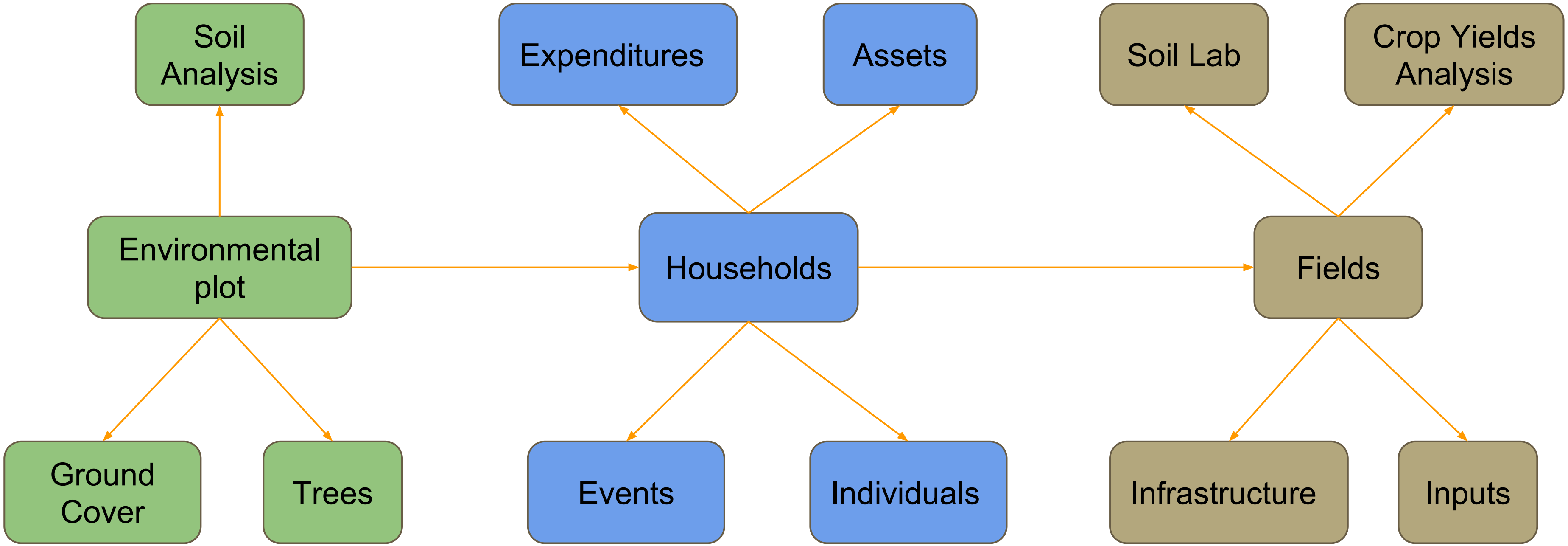
Data Scientists: Anthony Arendt, Joe Hellerstein

Fellows: Cara Arizmendi, Mitch Goist, Krista Jones, Rob Shaffer

Approach



Environmental, Social, and Agricultural Data

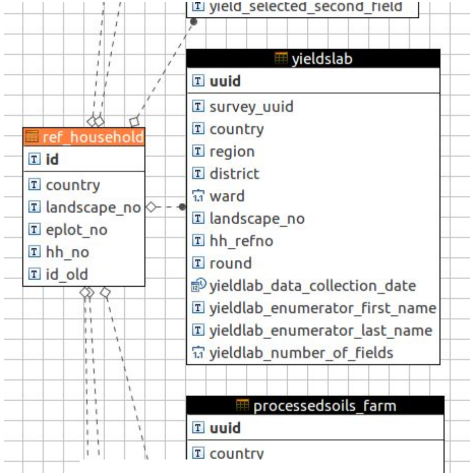


n ≈ 1350

n ≈ 750

n ≈ 2250

Data Pipeline



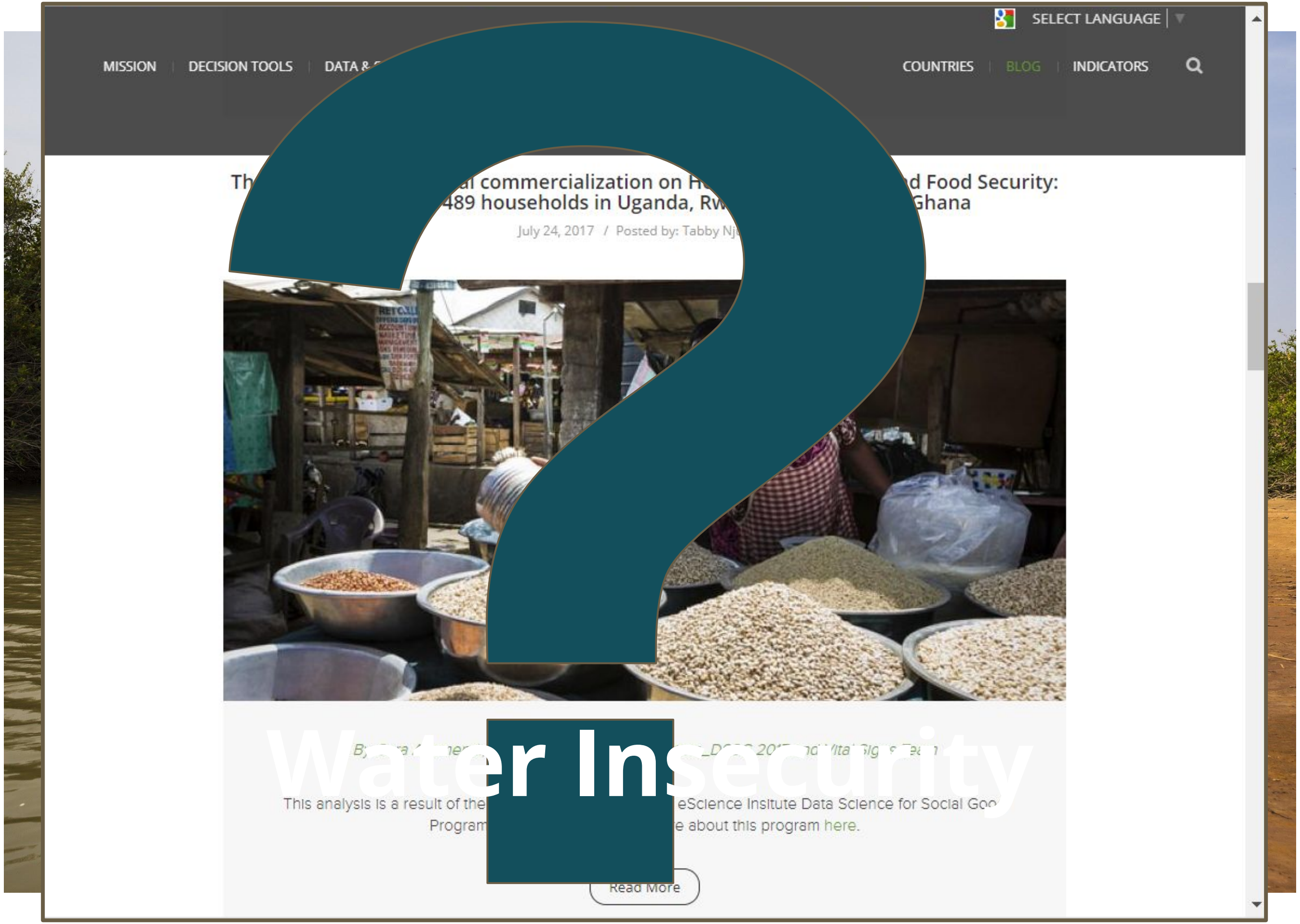
OPEN DATA KIT

PostgreSQL
the world's most advanced open source database

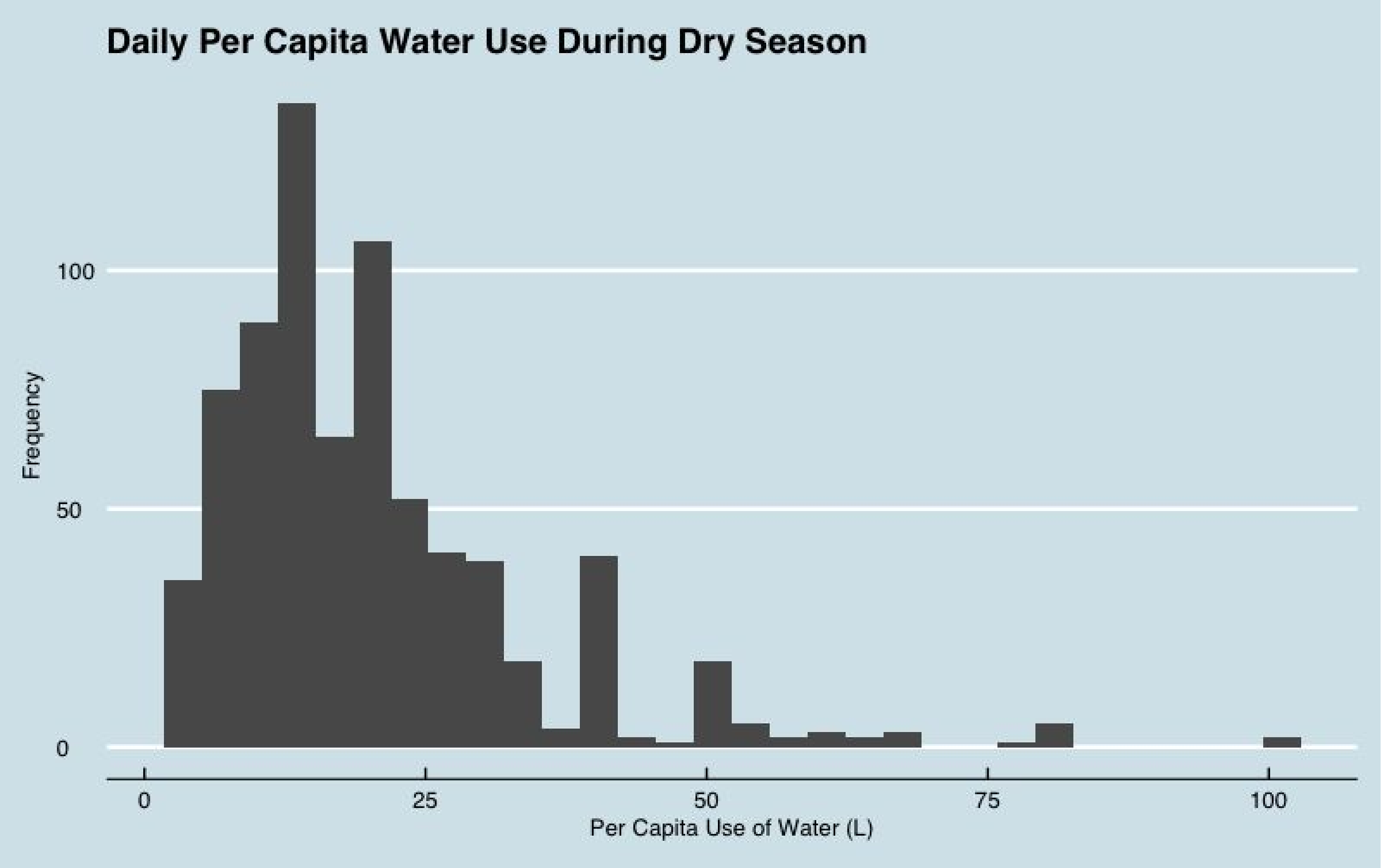
amazon web services


DECISION-SUPPORT DASHBOARD
a small set of indicators for smart agricultural investment
ANALYSIS
mathematical models and algorithms are applied
MEASUREMENT
consistent metrics are gathered on the ground and remotely via satellites

DSSG



Water Insecurity

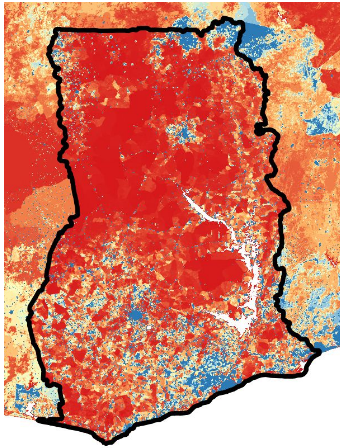

Daily Per Capita Water Use During Dry Season
Frequency
100
50
0
0
25
50
75
100
Per Capita Use of Water (L)

Water Insecurity Model

VS Survey data

- Reported water insecurity at household level

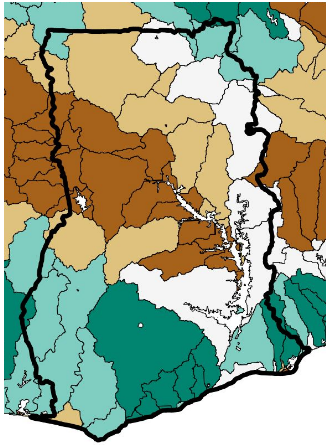
=



Rasters

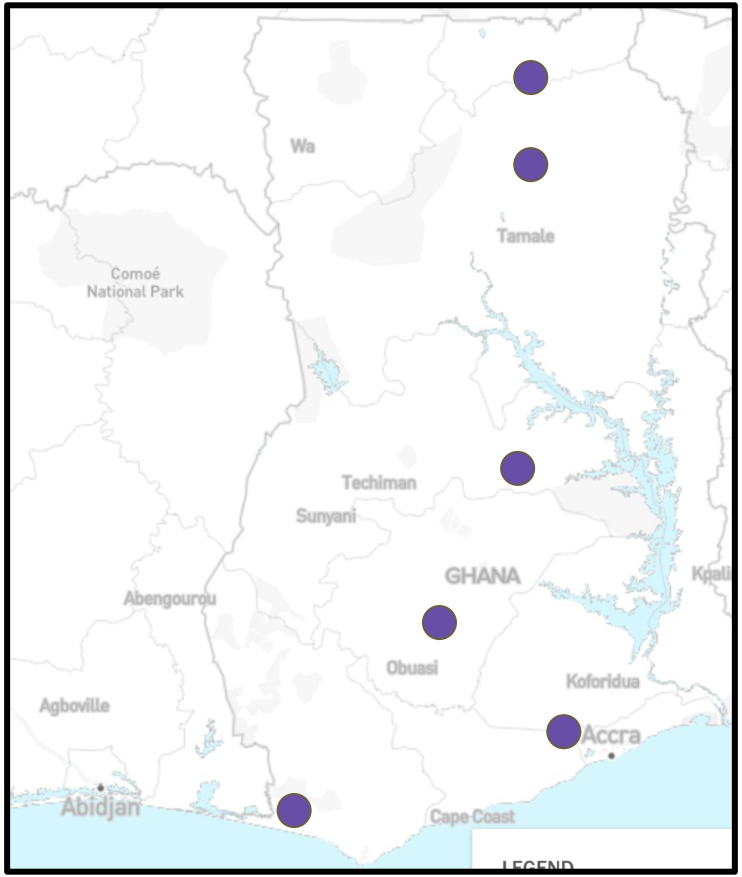
- Elevation
- Travel distance
- Demographics
- Climate

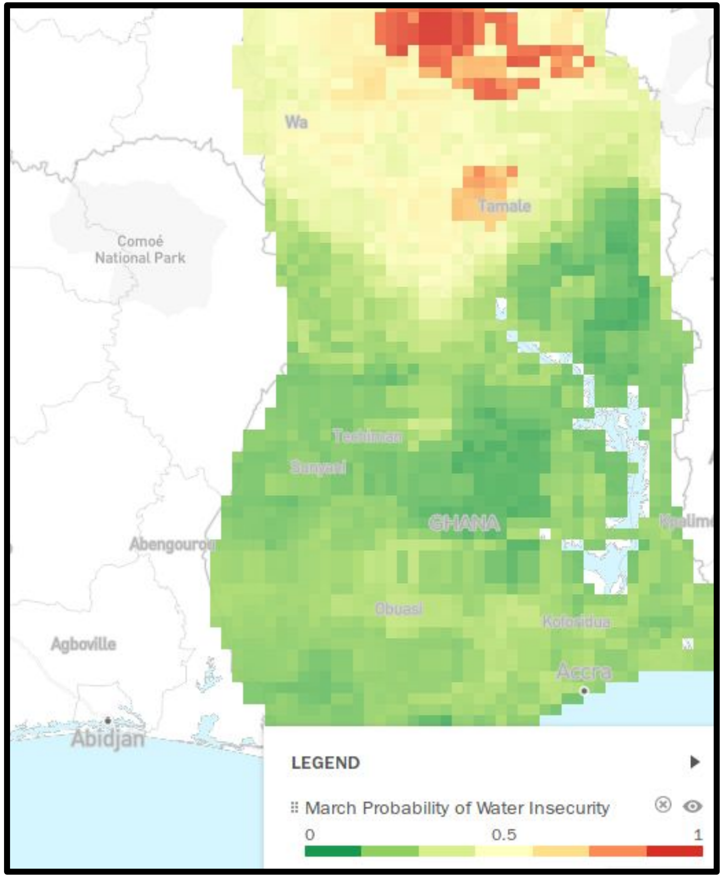
+



Watersheds

- % Forest
- % Cropland
- % Protected Area
- % Degraded Land


Wa
Tamale
Comoé
National Park
Techiman
Sunyani
GHANA
Abengourou
Obuasi
Koforidua
Agboville
Accra
Abidjan
Cape Coast


Wa
Tamale
Comoé
National Park
Techiman
Sunyani
GHANA
Abengourou
Obuasi
Koforidua
Agboville
Accra
Abidjan
LEGEND
March Probability of Water Insecurity
0
0.5
1

Modeling Strategy

Why **random forests**?

- Large feature space
- Non-parametric
- Minimize prediction error

Problem: **imbalanced classes**

- False negatives
- Misleading accuracy

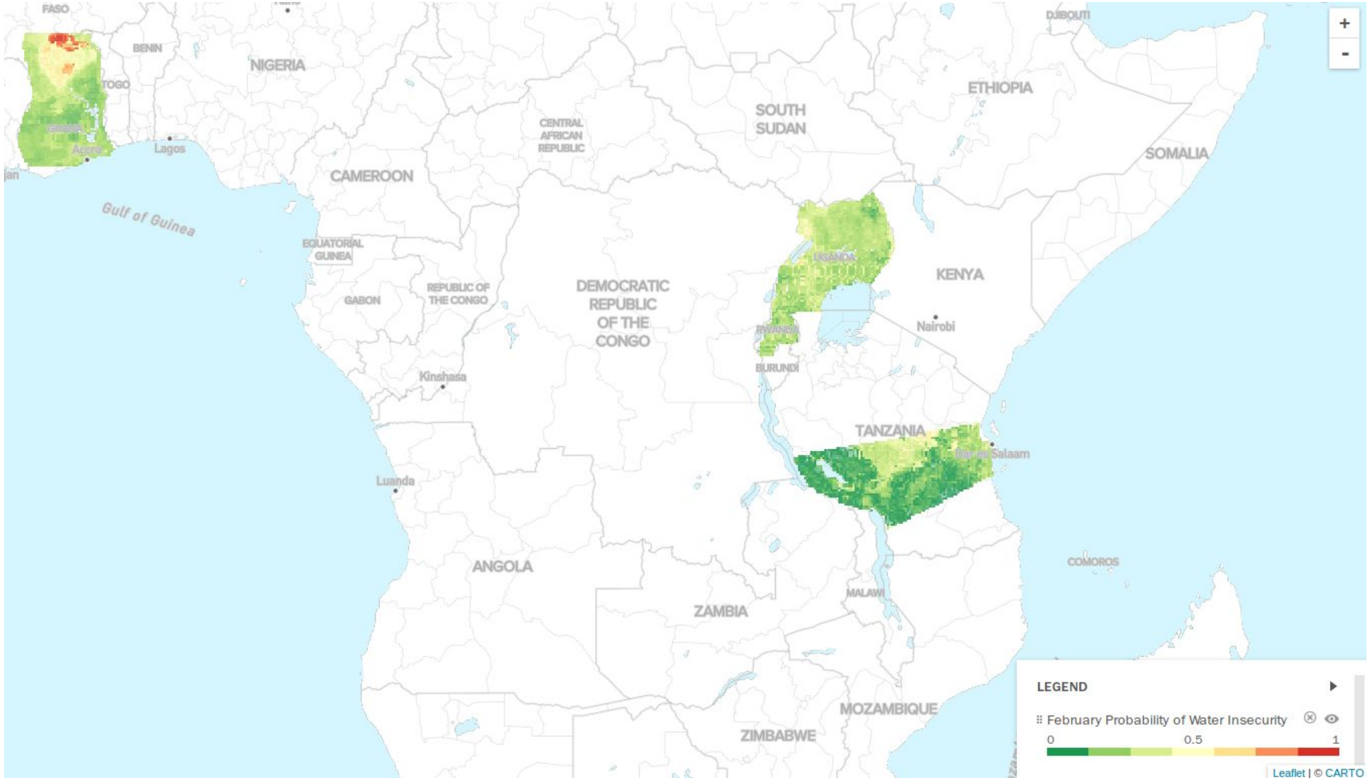
Regular Random Forest

	Actual Value	
Prediction	Secure	Insecure
Secure	8722	254
Insecure	0	0

With **SMOTE**

	Actual Value	
Prediction	Secure	Insecure
Secure	7625	45
Insecure	1097	209

Visualizing Water Insecurity

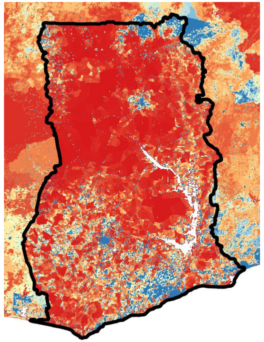


Water Insecurity Model

Water Insecurity

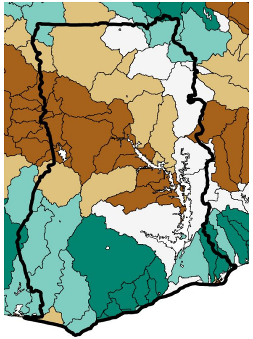
=

Rasters



+

Watersheds



+

Households

Incorporating Vital Signs Indicators

	True Positive	True Negative	F1 Score	Balanced Accuracy
With VS Predictors	**0.91**	**0.69**	**0.95**	**0.80**
Without VS Predictors	0.89	0.65	0.93	0.77

Conclusion

Broader goals for the Vital Signs summer project:

- **Identify** social, environmental, and agricultural synergies and tradeoffs in rural Africa
- Answer these questions by **combining** data from a variety of sources

Takeaway: real-world policy questions require interdisciplinary data!

CASCADIA URBAN ANALYTICS COOPERATIVE

UNIVERSITY of WASHINGTON eScience Institute

Washington Research FOUNDATION

Urban@UW

ALFRED P. SLOAN FOUNDATION 1934

Microsoft

MacArthur Foundation

BILL & MELINDA GATES foundation

GORDON AND BETTY MOORE FOUNDATION

---

http://dssg.resilienceatlas.org

http://vitalsigns.org/blogs