ESTIMATING MAIZE GRAIN YIELD IN SCARCE FIELD-DATA ENVIRONMENT:
AN APPROACH COMBINING REMOTE SENSING AND CROP MODELLING IN BURKINA FASO

Leroux L., Baron C., Castets M., Escorihuela M.-J., Diouf A.A., Bégué A., Lo Seen D.
AfricaGIS2017, November, 2017, Ethiopia
Maize (Zea mays L.): Keystone of Food Security

- Most produced crop in the world

- In West Africa:
  - Staple crop
  - Providers of health benefits and vital nutrients
  - 30 kg/capita/year
**Maize (Zea mays L.): Keystone of food security**

- Most produced crop in the world

- In West Africa:
  - Staple crop
  - Providers of health benefits and vital nutrients
  - 30 kg/capita/year

**Keystone to support food requirements and food security**
Maize (Zea mays L.): Keystone of Food Security

- Most produced crop in the world

In West Africa:
- Staple crop
- Providers of health benefits and vital nutrients
- 30 kg/capita/year

Keystone to support food requirements and food security

Maize yield trends [Ray et al., 2012]

- Stagnation of maize yields
- Socio-economic x biophysical limitations
Maize (Zea mays L.): Keystone of Food Security

- Most produced crop in the world

- In West Africa:
  - Staple crop
  - Providers of health benefits and vital nutrients
  - 30 kg/capita/year

Keystone to Support Food Requirements and Food Security

- Decline in per capita food production [Ray et al., 2013]
- Stagnation of maize yields
- Socio-economic x biophysical limitations

Maize Yield Trends [Ray et al., 2012]
Maize (*Zea mays* L.): Keystone of Food Security

- Most produced crop in the world
- In West Africa:
  - Staple crop
  - Providers of health benefits and vital nutrients
  - 30 kg/capita/year

Keystone to Support Food Requirements and Food Security

- Stagnation of maize yields
- Socio-economic vs biophysical limitations

Decline in Per Capita Food Production [Ray et al., 2013]

Timely and reliable information on maize crop yields is needed to provide timely estimates of food shortage and support decision-making.
YIELD ESTIMATION METHODS

FIELD-BASED SURVEY

- Expensive (time & labor)
- Sampling methods
- Inaccessibility
- Difficulties to upscale to large areas
YIELD ESTIMATION METHODS

FIELD-BASED SURVEY

- Expensive (time & labor)
- Sampling methods
- Inaccessibility
- Difficulties to upscale to large areas

LACK OF GROUND DATA OR UNRELIABLE DATA
## Yield Estimation Methods

<table>
<thead>
<tr>
<th>Context</th>
<th>Data</th>
<th>Methods</th>
<th>Biomass &amp; Cstr</th>
<th>Yield</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crop Growth Model</strong></td>
<td><img src="image" alt="Crop Growth Model Diagram" /></td>
<td></td>
<td></td>
<td></td>
<td>• Approximation of the reality on the ground</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Potential yields under water or nutrient limitation</td>
</tr>
<tr>
<td><strong>Field-based Survey</strong></td>
<td><img src="image" alt="Field-based Survey Diagram" /></td>
<td></td>
<td></td>
<td></td>
<td>• Expensive (time &amp; labor)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Sampling methods</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Inaccessibility</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Difficulties to upscale to large areas</td>
</tr>
</tbody>
</table>

**Lack of Ground Data or Unreliable Data**
<table>
<thead>
<tr>
<th>CONTEXT</th>
<th>DATA</th>
<th>METHODS</th>
<th>BIOMASS &amp; CSTR</th>
<th>YIELD</th>
<th>CONCLUSION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>YIELD ESTIMATION METHODS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>REMOTE SENSING</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image1.png" alt="Satellite Image" /></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image2.png" alt="Map" /></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CROP GROWTH MODEL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image3.png" alt="Crop Growth Model" /></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>FIELD-BASED SURVEY</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image4.png" alt="Field Survey" /></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**REMOTE SENSING**
- Timely and exhaustive information on vegetation cover
- Biomass = f(Vegetation Indices)
- Empirical model calibrated with agricultural statistics **BUT** available ~ 3 months after the end of the cropping season

**CROP GROWTH MODEL**
- Approximation of the reality on the ground
- Potential yields under water or nutrient limitation

**FIELD-BASED SURVEY**
- Expensive (time & labor)
- Sampling methods
- Inaccessibility
- Difficulties to upscale to large areas

**LACK OF GROUND DATA OR UNRELIABLE DATA**
<table>
<thead>
<tr>
<th>OBJECTIVES</th>
</tr>
</thead>
</table>

**IMPROVE MAIZE YIELDS ESTIMATION USING A CROP MODEL TO GENERATE DIFFERENT COMPONENTS OF YIELDS AS PROXY OF IN SITU OR AGRICULTURAL STATISTICS DATA, AND COMBINING THEM WITH REMOTE SENSING DATA**

*Uncalibrated approach* [Lobell et al., 2015, Burke et al., 2017, Sibley et al., 2014]
### Objectives

**Improve maize yields estimation using a crop model to generate different components of yields as proxy of in situ or agricultural statistics data, and combining them with remote sensing data**

*Uncalibrated approach* [Lobell et al., 2015, Burke et al., 2017, Sibley et al., 2014]

---

[1] Build a model relying on ecophysiological process
### Objectives

**Improve maize yields estimation using a crop model to generate different components of yields as proxy of in situ or agricultural statistics data, and combining them with remote sensing data**

*Uncalibrated approach* [Lobell et al., 2015, Burke et al., 2017, Sibley et al., 2014]

<table>
<thead>
<tr>
<th>Context</th>
<th>Data</th>
<th>Methods</th>
<th>Biomass &amp; Cstr</th>
<th>Yield</th>
<th>Conclusion</th>
</tr>
</thead>
</table>

1. **[1]** Build a model relying on ecophysiological process
2. **[2]** Conduct a benchmarking between linear and nonlinear models
**Objective**

**Improve maize yields estimation using a crop model to generate different components of yields as proxy of in situ or agricultural statistics data, and combining them with remote sensing data**

Uncalibrated approach [Lobell et al., 2015, Burke et al., 2017, Sibley et al., 2014]

<table>
<thead>
<tr>
<th>Context</th>
<th>Data</th>
<th>Methods</th>
<th>Biomass &amp; Cstr</th>
<th>Yield</th>
<th>Conclusion</th>
</tr>
</thead>
</table>

1. Build a model relying on ecophysiological process

2. Conduct a benchmarking between linear and nonlinear models

3. Compare the performance of remote sensing-based model in estimating and forecasting
STUDY AREA & FIELDS DATA

- Tuy province of Burkina Faso
- Sudanian climate
- Rainy season: June-September
- Agropastorale activities
- Rainfed crops: mainly maize and cotton
STUDY AREA & FIELDS DATA

- Tuy province of Burkina Faso
- Sudanian climate
- Rainy season: June-September
- Agropastorale activities
- Rainfed crops: mainly maize and cotton

Field surveys in 2014-2015-2016
- 114 maize fields
- Agricultural practices and vegetation parameters
STUDY AREA & FIELDS DATA

- Tuy province of Burkina Faso
- Sudanian climate
- Rainy season: June-September
- Agropastorale activities
- Rainfed crops: mainly maize and cotton

Field surveys in 2014-2015-2016
- 114 maize fields
- Agricultural practices and vegetation parameters

INDEPENDENT DATASET TO TEST THE ROBUSTNESS OF THE APPROACH
**DATA**

MODIS NDVI TIME SERIES

LAND SURFACE TEMPERATURE TIME SERIES

SMOS SOIL MOISTURE TIME SERIES

Disaggregated with the DISPATCH method [Merlin et al., 2013]
Phenology, Vegetation vigor and drought/heat stress related indicators

- CWSI
- TCI
- TVDI
- SMADI
- NDVI
- SSM

**DATA**
- MODIS NDVI time series
- Land Surface Temperature time series
- SMOS Soil Moisture time series

Disaggregated with the DISPATCH method [Merlin et al., 2013]
### Data
- MODIS NDVI time series
- Land surface temperature time series
- SMOS soil moisture time series

#### Phenology, Vegetation vigor and drought/heat stress related indicators
- CWSI
- TCI
- TVDI
- SMADI
- NDVI
- SSM

**Start of Season**
**Peak of Season**
**End of Season**

Disaggregated with the DISPATCh method [Merlin et al., 2013]
**Data**

- MODIS NDVI time series
- Land surface temperature time series
- SMOS soil moisture time series

Disaggregated with the DISPATCH method [Merlin et al., 2013]

**Phenology, Vegetation vigor and drought/heat stress related indicators**

- CWSI
- TCI
- TVDI
- SMADI
- NDVI
- SSM

- Start of Season
- Peak of Season
- End of Season

- Vegetative period
- Productive period
<table>
<thead>
<tr>
<th>CONTEXT</th>
<th>DATA</th>
<th>METHODS</th>
<th>BIOMASS &amp; CSTR</th>
<th>YIELD</th>
<th>CONCLUSION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SARRA-O CROP MODEL [BARON ET AL., 2005; CASTETS ET AL., IN PREP]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SARRA-O CROP MODEL [BARON ET AL., 2005; CASTETS ET AL., IN PREP]

- Sarra-H, a crop model for maize
- Daily time step
- Attainable biomass and yields under **climatic constraints**

- Implementation under the **Ocelet Modelling Plateform**
- **ECMWF** agrometeorological data
- **TAMSAT** rainfall data

- Validated for the Tuy province [Akakpo 2017]
### SARRA-o CROP MODEL [BARON ET AL., 2005; CASTETS ET AL., IN PREP]

<table>
<thead>
<tr>
<th>CONTEXT</th>
<th>DATA</th>
<th>METHODS</th>
<th>BIOMASS &amp; CSTR</th>
<th>YIELD</th>
<th>CONCLUSION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sarra-H, a crop model for maize</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily time step</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attainable biomass and yields under <strong>climatic constraints</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implementation under the <strong>Ocelet Modelling Plateform</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECMWF agrometeorological data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TAMSAT</strong> rainfall data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Validated for the Tuy province [Akakpo 2017]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Aboveground biomass at flowering**

**Crop Water stress from flowering to maturation**

**Attainable maize final yield**

---

2011-2015

- **AGB-F**
- **Cstr**
- Validated for the Tuy province [Akakpo 2017]
<table>
<thead>
<tr>
<th>STATISTICAL MODELS AND STRATEGY</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Above Ground Biomass at flowering (AGB-F) and water stress index (Cstr)</td>
</tr>
</tbody>
</table>

| AGB-F | Vegetation and drought indices – Vegetative period |
STATISTICAL MODELS AND STRATEGY

• *Above Ground Biomass at flowering (AGB-F) and water stress index (Cstr)*

<table>
<thead>
<tr>
<th>AGB-F</th>
<th>Vegetation and drought indices – Vegetative period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cstr phase 4-5</td>
<td>Vegetation and drought indices – Productive period</td>
</tr>
</tbody>
</table>
STATISTICAL MODELS AND STRATEGY

- Above Ground Biomass at flowering (AGB-F) and water stress index (Cstr)

<table>
<thead>
<tr>
<th>METHODS</th>
<th>BIOMASS &amp; CSTR</th>
<th>YIELD</th>
<th>CONCLUSION</th>
</tr>
</thead>
</table>

- AGB-F
  - Vegetation and drought indices – Vegetative period
- Cstr phase 4-5
  - Vegetation and drought indices – Productive period

- Multi Linear Regression Model (MLR)
- Random Forest Model (RF)
STATISTICAL MODELS AND STRATEGY

• Above Ground Biomass at flowering (AGB-F) and water stress index (Cstr)

EVALUATION

• 10-fold cross validation
• Cv-R², cv-RMSE ...
STATISTICAL MODELS AND STRATEGY

- Above Ground Biomass at flowering (AGB-F) and water stress index (Cstr)

AGB-F
Vegetation and drought indices – Vegetative period

Cstr phase 4-5
Vegetation and drought indices – Productive period

EVALUATION
- 10-fold cross validation
- Cv-R², cv-RMSE ...

IMPORTANCE VARIABLES
- LMG method
- Mean decrease in MSE
<table>
<thead>
<tr>
<th>Context</th>
<th>Data</th>
<th>Methods</th>
<th>Biomass &amp; Cstr</th>
<th>Yield</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statistical models and strategy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• <em>Final maize yield</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CONTEXT</td>
<td>DATA</td>
<td>METHODS</td>
<td>BIOMASS &amp; CSTR</td>
<td>YIELD</td>
<td>CONCLUSION</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
<td>---------</td>
<td>----------------</td>
<td>-------</td>
<td>------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**STATISTICAL MODELS AND STRATEGY**

- **Final maize yield**

**ESTIMATION**

Final maize yield → AGB-F x Cstr Phase 4-5
**STATISTICAL MODELS AND STRATEGY**

- **Final maize yield**

  **ESTIMATION**
  
  Final maize yield → AGB-F x Cstr Phase 4-5

  **FORECASTING**
  
  Final maize yield → Vegetation and drought indices – Vegetative period → MLR, RF
**Statistical Models and Strategy**

- **Final maize yield**
  
  **Estimation**
  
  Final maize yield → AGB-F x Cstr Phase 4-5
  
  **Forecasting**
  
  Final maize yield → Vegetation and drought indices – Vegetative period
  
  **Validation**
  
  - Yield from field survey
  - Village scale

**Methods**

- Biomass & Cstr
- Yield
• Overestimation of low values and underestimation of high values
• RF model significantly better than the MLR model (57% of the AGB-F variability)
• Overestimation of low values and underestimation of high values
• RF model significantly better than the MLR model (57% of the AGB-F variability)
• Importance variable for RF = TCI and NDVI (57%)
• Low but still significant predictive power
• But with cv-RRMSE < 2%? Low variability observed in SARRA-O Cstr!
• Low but still significant predictive power
• But with cv-RRMSE < 2%? Low variability observed in SARRA-O Cstr!
• **Temperature Condition Index** is the most important driver
• Impact of heat stress on maize: *grain number* [Eyshi Rezaei et al., 2015]
EVALUATION OF MAIZE YIELDS ESTIMATION AT THE END OF THE SEASON

- **Yield** = \( f(AGB - F, Cstr \text{ Phase } 4 - 5) \)
- Good potential for maize yield estimation (RMSE < 300 kg/ha)
- Good fitting of probability distribution curves:
  - Median SARRA-O: 3634 kg/ha
  - Median MLR: 3648 kg/ha
  - Median RF: 3659 kg/ha
Evaluation of Early Assessment of Maize Yields

- **Yield** = f(Remote Sensing Indices – Vegetative period)
- Good potential for maize yield forecasting (RMSE<300 kg/ha)
- ~50% of maize yield variability can be explained ~2 months before harvest
**Validation of Maize Yields with Ground Data**

*Independent dataset*

*2014, 2015 and 2016*
Validation of Maize Yields with Ground Data

*Independent dataset*
*2014, 2015 and 2016*

- RF outperforms MLR:
  - Estimation: $R^2=0.60$
  - Forecasting: $R^2=0.46$
- High overestimation in forecasting
- MLR: 2016 not accurately estimated
A good and effective potential of ‘uncalibrated approach’ to estimate maize yield in scarce data environment
A GOOD AND EFFECTIVE POTENTIAL OF ‘UNCALIBRATED APPROACH’ TO ESTIMATE MAIZE YIELD IN SCARCE DATA ENVIRONMENT

LOOK-BACK ON THE STUDY OBJECTIVES

- Linear vs NonLinear models:
  - Higher performance of RF models both for estimation and early assessment
  - Complex interaction among biophysical, ecological, physiological and management practices
A GOOD AND EFFECTIVE POTENTIAL OF ‘UNCALIBRATED APPROACH’ TO ESTIMATE MAIZE YIELD IN SCARCE DATA ENVIRONMENT

LOOK-BACK ON THE STUDY OBJECTIVES

- **Linear vs NonLinear models:**
  - Higher performance of RF models both for estimation and early assessment
  - Complex interaction among biophysical, ecological, physiological and management practices

- **Estimation vs Early assessment:**
  - Early assessment of maize yields ~ 2 months before harvesting (RF)
  - Complementary of approaches:
    - In-season forecasting: food aids strategies or market and trade information
    - After harvesting: agricultural statistics
    - Both + outputs of crop model: ‘convergence of evidence’ in EWS

GOOD AND EFFECTIVE POTENTIAL OF ‘UNCALIBRATED APPROACH’ TO ESTIMATE MAIZE YIELD IN SCARCE DATA ENVIRONMENT

LOOK-BACK ON THE STUDY OBJECTIVES

- **Linear vs NonLinear models:**
  - Higher performance of RF models both for estimation and early assessment
  - Complex interaction among biophysical, ecological, physiological and management practices

- **Estimation vs Early assessment:**
  - Early assessment of maize yields ~ 2 months before harvesting (RF)
  - Complementary of approaches:
    - In-season forecasting: food aids strategies or market and trade information
    - After harvesting: agricultural statistics
    - Both + outputs of crop model: ‘convergence of evidence’ in EWS

GOOD AND EFFECTIVE POTENTIAL OF ‘UNCALIBRATED APPROACH’ TO ESTIMATE MAIZE YIELD IN SCARCE DATA ENVIRONMENT

LOOK-BACK ON THE STUDY OBJECTIVES

- **Linear vs NonLinear models:**
  - Higher performance of RF models both for estimation and early assessment
  - Complex interaction among biophysical, ecological, physiological and management practices

- **Estimation vs Early assessment:**
  - Early assessment of maize yields ~ 2 months before harvesting (RF)
  - Complementary of approaches:
    - In-season forecasting: food aids strategies or market and trade information
    - After harvesting: agricultural statistics
    - Both + outputs of crop model: ‘convergence of evidence’ in EWS
THANK YOU FOR LISTENING

Leroux Louise – CIRAD UR AÏDA/CSE
louise.leroux@cirad.fr
www.louise.leroux.igeo.fr

MORE INFORMATION:
LEROUX ET AL., 2017, Maize grain yield estimating in a west African agricultural landscape at the crossroads between remote sensing, crop modelling and statistical methods: Case study in south-west of Burkina Faso, to be submitted to Agric.FoR.MeteoroL