K&C Phase 3 – Brief project essentials

Assessing woody structural properties of semi-arid African savannahs from multi-frequency SAR data

Renaud Mathieu, Laven Naidoo, Konrad Wessels, Greg Asner, Mikhail Urbazaev, Christiane Schmullius

Council for Scientific and Industrial Research, South Africa
Carnegie Institute for Science, USA
Friedrich-Schiller-University Jena, Germany
Context

- Savannahs (94% of forests): mixed grass / woody layer
- Woody component essential in region
  - Food / energy security → poor communities
  - Biodiversity
  - Carbon accounting (REDD+), 1/6 of land surface;
    3rd carbon pool after tropical and temperate forest
  - Climate modeling (grass vs. woody dynamic)
- Limited “validated” spatial data in Southern Africa
- Research programme aims to:
  - Assess a range of remote sensing techniques to quantify woody structure parameters
  - Develop national/regional calibrated & validated woody structure products
  - Develop the technology base (and human capacity) for developing remote sensing products for SA environments
Context

Policy on Woodlands

*National Forests Act (84 of 1998) caters for woodlands* and recognizes them explicitly as renewable energy source. The Act makes provision for research, monitoring, dissemination of information and reporting. *DAFF has legally to report every three years on the status of woodlands to the minister.*

- **Carbon accounting (REDD+), 1/6 of land surface;**

Policy on Woodlands

Medium Term Strategic Framework Outcome 10 led by DEA

“Net deforestation to be maintained at no more than 5% woodlands by 2020”

“Undertake provincial and national forest resource assessment programs”

Green economy

Nationally households use 4.5-6.7 million T / yr of wood for energy to a value of R3 billion. Savings on new electricity generation between USD14.6-77.4 million per yr.
Savannahs & woodlands in southern Africa

- Arid / semi-arid: 10-50% woody cover, < 60 woody T/ha ABG
- Mostly gradual changes: logging, encroachment
- Fine scale heterogeneity = remote sensing challenge
  - Woody plant size & cover (3-6 m, 10-40%)
  - Soil properties & water availability
  - Disturbance factors: fire, herbivore, human
- Woody plant: multi-stemmed clumps, high biomass in branches rather than in main stem
Project area: Greater Kruger National Park, South Africa

- Dominant landscape = woodlands and savannahs, with plantations and remnants of indigenous forest
- Generally flat or gently undulated
- Ca. 60000 km²
- Dominant land uses: commercial plantations, private & public conservation areas, private rangelands (beef production), communal rangelands (subsistence use)
- Issues: biodiversity conservation, energy security (woodland thinning), bush encroachment
Project area: Greater Kruger National Park, South Africa

- Dominant landscape = woodlands and savannahs, with plantations and remnants of indigenous forest
- Generally flat or gently undulated
- Ca. 60000 km²
- Dominant land uses: commercial plantations, private & public conservation areas, private rangelands (beef production), communal rangelands (subsistence use)
- Issues: biodiversity conservation, energy security (woodland thinning), bush encroachment
Dominant landscape = woodlands and savannahs, with plantations and remnants of indigenous forest

- Generally flat or gently undulated
- Ca. 60000 km²
- Dominant land uses: commercial plantations, private & public conservation areas, private rangelands (beef production), communal rangelands (subsistence use)

- Issues: biodiversity conservation, energy security (woodland thinning), bush encroachment

Project area: Greater Kruger National Park, South Africa
Data currently available

- **Ground**
  - Woody cover (N= 37) and biomass plots (N= 152)

- **Airborne**
  - LiDAR 2008, 10 & 12 (Carnegie Airborne Observatory) – end wet season

- **Satellite**
  - SAR: Radarsat-2 (C-band), ALOS-1 PALSAR (L-band), TerraSAR-X (X-band)

- High res SAR biomass / cover mapping
- Medium res SAR biomass / cover mapping
Data currently available

- **Ground**
  - Woody cover (N= 37) and biomass plots (N= 152)

- **Airborne**
  - LiDAR 2008, 10 & 12 (Carnegie Airborne Observatory) – end wet season

- **Satellite**
  - SAR: Radarsat-2 (C-band), ALOS-1 PALSAR (L-band), TerraSAR-X (X-band)

  - High res SAR biomass / cover mapping
  - Medium res SAR biomass / cover mapping
Project objectives and schedule

- **General objective:** assess and develop “affordable” methods to predict woody cover and biomass in southern African woodlands and savannahs using SAR imagery

- **Secondary objectives:**
  - Investigate the potential of combining multiple SAR frequencies (L-band ALOS PalSAR, C-band Radarsat-2, X-band TerraSAR-X)
  - Optical / SAR “fusion”
  - Investigate full polarimetric ALOS PalSAR / RADARSAT-2 imagery and polarimetric decompositions
Project objectives and schedule (cont)

- List the project milestones
  - Milestone 1 (11/2012): field & airborne LiDAR campaign
  - Milestone 2 (12/2012): LiDAR data processing
  - Milestone 3 (04/2013): SAR data acquisition, and SAR processing chains (including training and script development, i.e GAMMA)
  - Milestone 4 (09/2013): Assessment of multifrequency SAR for woody and biomass prediction
  - Milestone 5 (12/2013): Assessment of dual and full polarimetric ALOS PalSAR for woody and biomass prediction
  - Milestone 6 (02/2014): 2000 – 2010 change analysis over Kruger National Park
General methodology

Access best scale for data integration

Independent variables:
- Scenario
  - Polarization
  - Season
  - Frequency

Applied to Data mining, regression and machine learning algorithms (in WEKA 3.6.9 & R rattle software)

R2, RMSE

SAR woody structural parameter maps

Dependent variable:
- Cover
- Canopy volume
- AG biomass

TerraSAR-X + RADARSAT + ALOS PALSAR + LiDAR integrated variable dataset (at 105m cell size)

Cover
Canopy volume
AG biomass

Combined model:
- X-band model
- C-band model
- L-band model

LiDAR – SAR Error Assessment (stats + spatial)
Field data collection

- **Field data**
  - 37 sites with clustered plots (min 50m apart, autocorrelation determined via geostatistics on LiDAR)

- **Biomass**
  - DBH, height, species

- **Woody cover using transect and point intercept methods**

Biomass plots DBH >5cm, height > 1.5m

Biomass subplots DBH >3cm, height > 1.5m

Cover plots with transects
LiDAR-based structural variables

- Woody cover: area vertically projected on a horizontal plane (%)
- Canopy volume: approximated from integration of vertical profile of laser hits
- Biomass: Linear model between field AGB and LiDAR woody cover and height metrics

Methods after Colgan et al, 2012, Biogeosci. Disc
LiDAR-based cal/cal: height and cover

Field metric

LiDAR metric

Height

Woody cover

Field metric
LiDAR-based cal/cal: above ground biomass

Field biomass (25x25m plots)

Biomass = 9.8 \times \text{Mean Woody Height} \times \text{Canopy Cover} + 32

LiDAR models and field biomass derived from three sets of allometric equations
### SAR processing

<table>
<thead>
<tr>
<th>Dual Pol ALOS PALSAR, Level 1.1</th>
<th>Fine Quad Pol RADARSAT-2, SLC</th>
<th>Dual Pol TerraSAR-X, StripMap, Level 1b, MGD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi – looking</td>
<td>Multi – looking</td>
<td>Multi – looking</td>
</tr>
<tr>
<td>Range: 2, Azimuth: 8</td>
<td>Range: 1, Azimuth: 5</td>
<td>Range: 4, Azimuth: 4</td>
</tr>
<tr>
<td>Calibration: HH &amp; HV backscatter intensity</td>
<td>Calibration: HH &amp; HV backscatter intensity</td>
<td>Calibration: HH &amp; HV backscatter intensity</td>
</tr>
<tr>
<td>Geocoded (90m SRTM DEM)</td>
<td>Geocoded (90m SRTM DEM)</td>
<td>Geocoded (90m SRTM DEM)</td>
</tr>
<tr>
<td>final resolution: 12.5m</td>
<td>final resolution: 5m</td>
<td>final resolution: 3m</td>
</tr>
</tbody>
</table>

Final SAR products integrated using 105m sampling grid
Sampling blocks overlaid on the LiDAR canopy height model, blue to red, increasing woody canopy cover.

**Scale of SAR / LiDAR integration**

- Inherent SAR noise: speckle
- Savannah heterogeneity
- Geolocation issues
- Best trade-off mapping details, model performance around 100 m, average size on landscape units along catena

**Effect of size of sampling blocks on relationships between LiDAR metrics cover and Radarsat-2 C-HV bands**

**Inherent SAR noise:** speckle

**Savannah heterogeneity**

**Geolocation issues**

**Best trade-off mapping details,** model performance around 100 m, average size on landscape units along catena
Sampling blocks overlaid on the LiDAR canopy height model, blue to red, increasing woody canopy cover.

**Effect of size of sampling blocks on relationships between LiDAR metrics cover and Radarsat-2 C-HV bands**

- Inherent SAR noise: speckle
- Savannah heterogeneity
- Geolocation issues
- Best trade-off mapping details, model performance around 100 m, average size on landscape units along catena

**Scale of SAR / LiDAR integration**

- RMSE vs woody cover
- RMSE vs canopy volume
- R² vs woody cover
- R² vs canopy volume

- 105m
- 25m
- 50m
- 125m
- 200m
- 495m
- 105m

- Radarsat-2
- PALSAR
Season

- Strong change of vegetation condition – water balance with season and phenology
  - Middle Wet (Summer) – grass green and woody leaf-on
  - End of Wet (Autumn) – grass dry and woody leaf-on
  - Dry (Winter) – grass dry and woody leaf-off
  - Early Wet (Spring) – grass dry and woody leaf-on
- Best is dry season, than middle wet, and end of wet (early wet?)
  - Low moisture effect, higher penetration (high frequency)
- Similar pattern for L- and C-band

Mean $R^2$ between PALSAR HH backscatter intensity and LiDAR-based woody cover at four aggregation levels (DRY dry season; EWET end of wet season; MWET middle of wet season)
Polarization

- HV > HH >>> VV
- HV and HH similar with leaf off (dry season)
- Polarimetric decomposition: Freeman-Durden, Van Zyl (double bounce, single bounce, volume) Cloude-Pottier (entropy H, anisotropy A, alpha angle $\alpha$)
- Volume = or < to HV
- Similar pattern for L- and C-band

<table>
<thead>
<tr>
<th></th>
<th>25 m</th>
<th>50 m</th>
<th>125 m</th>
<th>200 m</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DRY</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (HH DRY)</td>
<td>0.32</td>
<td>0.63</td>
<td>0.69</td>
<td>0.73</td>
</tr>
<tr>
<td>Mean (HV DRY)</td>
<td>0.34</td>
<td>0.64</td>
<td>0.69</td>
<td>0.72</td>
</tr>
<tr>
<td><strong>MWET</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (HH MWET)</td>
<td>0.26</td>
<td>0.30</td>
<td>0.51</td>
<td>0.52</td>
</tr>
<tr>
<td><strong>EWET</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (HH EWET)</td>
<td>0.07</td>
<td>0.22</td>
<td>0.36</td>
<td>0.45</td>
</tr>
<tr>
<td>Mean (VV EWET)</td>
<td>0.04</td>
<td>0.14</td>
<td>0.27</td>
<td>0.35</td>
</tr>
<tr>
<td>Mean (HV EWET)</td>
<td>0.12</td>
<td>0.34</td>
<td>0.51</td>
<td>0.60</td>
</tr>
<tr>
<td>Mean (VH EWET)</td>
<td>0.10</td>
<td>0.30</td>
<td>0.47</td>
<td>0.57</td>
</tr>
<tr>
<td><strong>EWET</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (Freeman Vol)</td>
<td>0.21</td>
<td>0.42</td>
<td>0.52</td>
<td>0.61</td>
</tr>
<tr>
<td>Mean (VanZyl Vol)</td>
<td>0.21</td>
<td>0.42</td>
<td>0.52</td>
<td>0.61</td>
</tr>
</tbody>
</table>
## Model assessment

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Algorithm type</th>
<th>Algorithm Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Regression</td>
<td>Parametric &amp; linear</td>
<td>Utilises linear regression for prediction and is able to deal with weighted instances.</td>
</tr>
<tr>
<td>SMO Regression</td>
<td>Non-parametric &amp; non-linear</td>
<td>Implements a support vector machine for regression.</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>Non-parametric &amp; non-linear</td>
<td>Uses a recursive partitioning approach to split data into 'branches' of common parameter attributes</td>
</tr>
<tr>
<td>Artificial Neural Network</td>
<td>Non-parametric &amp; non-linear</td>
<td>Utilises multiple layers of neurons connected to each other which feed the data through the network</td>
</tr>
<tr>
<td>Random Forest</td>
<td>Non-parametric &amp; non-linear</td>
<td>Utilises an ensemble of un-pruned decision trees which votes on the best decision tree design based on plurality</td>
</tr>
</tbody>
</table>

Tested on C-, and L-band and combination

RF generally the best, but not a huge improvement

### C-, L- and C+L Band TCC (%) Modelling Results [35% Training; 65% Validation]

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Algorithm type</th>
<th>C-band only</th>
<th>L-band only</th>
<th>C+L band</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>R²</td>
<td>RMSE</td>
<td>R²</td>
</tr>
<tr>
<td>Linear Regression</td>
<td>Parametric &amp; linear</td>
<td>0.72</td>
<td>12.42</td>
<td>0.81</td>
</tr>
<tr>
<td>SMO Regression</td>
<td>Non-parametric &amp; non-linear</td>
<td>0.72</td>
<td>12.61</td>
<td>0.82</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>Non-parametric &amp; non-linear</td>
<td>0.73</td>
<td>12.09</td>
<td>0.82</td>
</tr>
<tr>
<td>ANN</td>
<td>Non-parametric &amp; non-linear</td>
<td><strong>0.76</strong></td>
<td><strong>11.55</strong></td>
<td><strong>0.83</strong></td>
</tr>
<tr>
<td>Random Forest</td>
<td>Non-parametric &amp; non-linear</td>
<td>0.75</td>
<td>11.82</td>
<td>0.83</td>
</tr>
</tbody>
</table>

Tested on C-, and L-band and combination

RF generally the best, but not a huge improvement
Multi-frequency: hypothesis long vs. shorter SAR wavelengths

- C-band saturation occurs at higher biomass level in savannas (Lucas et al. 2006)
- Savannas woody structure (low density) favours increased wave penetration, better for C-band?
- Expected free C-band access with Sentinel-1
Multi-frequency assessment

- Assess various scenario with winter HH/HV polarization, RF model
  - Single frequency (e.g. C band)
  - Dual frequency (e.g. C and L bands)
  - Tri frequency (X, C, and L bands)
- Best single L>>C>>>X, X gave poor results
- X, C, L combination gave systematically best result
- Improvement needs to consider costs of multiple datasets
- Structural metric canopy volume>biomass>woody cover

### CC Model Validation Results [split into 35% Training & 65% Validation]; Units = %

<table>
<thead>
<tr>
<th></th>
<th>X-band only</th>
<th>C-band only</th>
<th>L-band only</th>
</tr>
</thead>
<tbody>
<tr>
<td>R²</td>
<td>RMSE (SEP)</td>
<td>R²</td>
<td>RMSE (SEP)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.34</td>
<td>18.12 (50.87%)</td>
<td>0.61</td>
<td>13.20 (38.50%)</td>
</tr>
<tr>
<td>X+C band</td>
<td></td>
<td>X+L band</td>
<td>C+L band</td>
</tr>
<tr>
<td>R²</td>
<td>RMSE (SEP)</td>
<td>R²</td>
<td>RMSE (SEP)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.69</td>
<td>11.71 (33.94%)</td>
<td>0.80</td>
<td>9.90 (27.78%)</td>
</tr>
<tr>
<td>X+C+L band</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>RMSE (SEP)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.83</td>
<td>8.76 (25.40%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TCV Model Validation Results [split into 35% Training & 65% Validation]; Units = unitless per hectare

<table>
<thead>
<tr>
<th></th>
<th>X-band only</th>
<th>C-band only</th>
<th>L-band only</th>
</tr>
</thead>
<tbody>
<tr>
<td>R²</td>
<td>RMSE (SEP)</td>
<td>R²</td>
<td>RMSE (SEP)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.35</td>
<td>35534.50 (33.79%)</td>
<td>0.66</td>
<td>24731.06 (24.07%)</td>
</tr>
<tr>
<td>X+C band</td>
<td></td>
<td>X+L band</td>
<td>C+L band</td>
</tr>
<tr>
<td>R²</td>
<td>RMSE (SEP)</td>
<td>R²</td>
<td>RMSE (SEP)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.72</td>
<td>22243.64 (21.59%)</td>
<td>0.82</td>
<td>18609.04 (17.70%)</td>
</tr>
<tr>
<td>X+C+L band</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>RMSE (SEP)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.85</td>
<td>16443.57 (15.96%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### AGB Model Validation Results [split into 35% Training & 65% Validation]; Units = tonnes/ha

<table>
<thead>
<tr>
<th></th>
<th>X-band only</th>
<th>C-band only</th>
<th>L-band only</th>
</tr>
</thead>
<tbody>
<tr>
<td>R²</td>
<td>RMSE (SEP)</td>
<td>R²</td>
<td>RMSE (SEP)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.32</td>
<td>10.88 (59.82%)</td>
<td>0.60</td>
<td>7.81 (43.66%)</td>
</tr>
<tr>
<td>X+C band</td>
<td></td>
<td>X+L band</td>
<td>C+L band</td>
</tr>
<tr>
<td>R²</td>
<td>RMSE (SEP)</td>
<td>R²</td>
<td>RMSE (SEP)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.67</td>
<td>7.19 (40.33%)</td>
<td>0.81</td>
<td>5.70 (31.35%)</td>
</tr>
<tr>
<td>X+C+L band</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>RMSE (SEP)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.83</td>
<td>5.20 (29.18%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Validation of woody cover model

From short to longest wavelength and addition of bands tend to decrease over estimation and underestimation at low and high density
Validation of canopy volume model
Woody cover mapping

Canopy volume

$R^2 = 0.71$, RMSE = 8.1% (38%)

Multi-season C-band prediction

LiDAR observation

Mathieu et al., RSE, 2013

$R^2 = 0.75$, RMSE = 17200 (34%)

Multi-season C-band prediction

LiDAR observation
Provincial up scaling

Maps are produced using random forest models and dual pol (HH/HV), 14 scene mosaic.

Total canopy cover maps from ALOS Palsar L-band in the South African Loweld.

Need more extended cal/val
Final comments

- **ALOS PALSAR:** high potential for mapping woody structure
  - C-band also potential → free SAR upcoming satellite (Sentinel-1)

- **Future steps**
  - “Fusion” with freely available Landsat product (L8)
  - Compare 2008-2010 PALSAR to earlier map (2000, JERS-1 / Landsat), and PALSAR-2
  - Assess multi-temporal (backscatter / coherence) and fully polarimetric ALOS-PALSAR-2

- **Coupled LiDAR - SAR**
  - Good trade off between area covered (large track, landscape scale) and details (patchy environment) between point sampling and wall-to-wall mapping
  - Understanding long term dynamics and effects of drivers (human, fire, herbivory)
  - Wall-to-wall up-scaling, calibration/validation of satellite imagery (e.g. SAR models with LiDAR), error propagation needs too be documented (biomass)

- **Establish network of regional long-term pilot sites** (ground, airborne, and satellite data): e.g. structure (3-5 yr, LiDAR), species (> 5 yr, WV-2)

- Opportunity to develop within region monitoring tools and EO products (GMES) answering to local needs and suited to local ecosystems
Final comments

- ALOS PALSAR: high potential for mapping woody structure
  - Scheduled (next 1-2yrs) for a SAR mission on satellite (Sentinel-1)

- Future steps
  - "Fusion" with freely available Landsat product (L8)
  - Compare 2008-2010 PALSAR to earlier map (2000, JERS-1 / Landsat), and PALSAR-2
  - Assess multi-temporal (backscatter / coherence) and fully polarimetric ALOS-PALSAR-2
  - Coupled LiDAR - SAR
  - Good trade off between area covered (large track, landscape scale) and details (patchy environment) between point sampling and wall-to-wall mapping
  - Understanding long term dynamics and effects of drivers (human, fire, herbivory)
  - Wall-to-wall up-scaling, calibration/validation of satellite imagery (e.g. SAR models with LiDAR), error propagation needs too be documented (biomass)
  - Establish network of regional long-term pilot sites (ground, airborne, and satellite data):
    - e.g. structure (3-5 yr, LiDAR), species (> 5 yr, WV-2)
  - Opportunity to develop within region monitoring tools and EO products (GMES) answering to local needs and suited to local ecosystems
Thank you!

Acknowledgement:
Carnegie Institute of Science, JAXA, DLR
EU FP7 AGRICAB, SA Department of Science and Technology