Fusing LiDAR with RTK GPS Using Random Forest Regression Shows Promise for DEMs of Mangrove and Sawgrass Soil Heights in Florida’s Coastal Everglades

GEER 2017: Coral Springs, FL
Hannah Cooper & Dr Caiyun Zhang, Department of Geosciences, Florida Atlantic University
April 18, 2017
Significance of Digital Elevation Models (DEM) in the Everglades

- Restoration requires accurate DEMs to monitor & simulate water levels, water depths & hydroperiods (e.g. Jones et al., 2012).
- Accurate & fine horizontal resolution DEMs are necessary for regional sea-level rise assessments (e.g. Zhang, 2011).
The Everglades hydrologic community agreed upon vertical elevation error threshold of $\pm 15$ cm (Desmond, 2003; Jones et al., 2012).

- High Accuracy Elevation Data (HAED) collected by USGS 1995-2007
  - Pros: Meets the strict error requirements
  - Cons: Coarse horizontal resolution (400 m)

- Light Detection and Ranging (LiDAR)
  - Pros: Fine horizontal resolution (e.g. 2 m)
  - Cons: Does not meet the strict error requirements
LiDAR is known to overestimate coastal marsh elevation, so previous studies make improvements by:

- **Minimum Bin (MB) technique** (e.g. Schmid et al., 2011)
  - Assigns minimum LiDAR ground return value within a grid cell to that grid cell
- **Bias correction technique** (e.g. Montane and Torres, 2006)
  - Calculates correction factors based on GPS & LiDAR DEM for each vegetation community

Few efforts were made to apply contemporary modeling to correct LiDAR for DEM improvement
Develop models to correct LiDAR for an improved DEM product that meets Everglades’ vertical error by:

1) Using RTK GPS and Random Forest Regression (RFR)
2) Comparing RFR with Minimum Bin (MB) and biased-correction techniques for best results
Figure 1 Map of study area located just south of Nine Mile Pond.
Data

- **FDEM LiDAR ground returns**
  - Collected Feb 2008
  - Point density 2 pts/m$^2$

- **SFWMD Land cover data**
  - Photointerpreted from Feb-March 2004 imagery

- **Real Time Kinematic (RTK) GPS**
  - Collected for this study Feb 2016
  - Vertical positions <0.03 m (1 $\sigma$)
  - 256 total (162 mangrove, 72 sawgrass, 22 road)
To assess **accuracy of LiDAR ground returns** and **test RFR model**, the relationship between RTK GPS and LiDAR is determined by **proximity and statistical analysis**:

**Methods: Proximity Analysis**

<table>
<thead>
<tr>
<th>GPS ID</th>
<th>LiDAR min</th>
<th>LiDAR max</th>
<th>LiDAR μ</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.02</td>
<td>0.04</td>
<td>0.01</td>
<td>3</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>256</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>1</td>
</tr>
</tbody>
</table>
Methods: RFR

RTK GPS and respective LiDAR descriptive statistics

- Randomly split data into two datasets
  - Calibration dataset
    - Explanatory variables: LiDAR descriptive statistics
    - Response variable: RTK GPS
      - 10-fold cross-validation to fit Random Forest Regression model
        - Predict calibration RTK GPS, report RMSE
  - Validation dataset
    - Predict validation RTK GPS, report RMSE

Generate Random Forest Regression DEM
Methods

Minimum Bin (MB) technique

- LiDAR elevation (m)
- Distance (m)
- 2 m resolution grid cell
- 0 m elevation

Bias-correction technique

- Calculate mean bias ($\Delta \mu$) between RTK GPS calibration dataset and MB for each land cover
- Adjust MB by $\Delta \mu$ per respective land cover
- MB calibrated DEM where accuracy & error assessed by validation dataset
Results: Proximity Analysis

Table 1 Descriptive statistics for surveyed GPS ($Z_{GPS}$) and LiDAR ground returns ($Z_{LiDAR}$) within a 1 m radius of each GPS. Where $\mu_{\Delta_{\text{min}}}$ = mean difference between minimum LiDAR values and GPS, all in meters.

<table>
<thead>
<tr>
<th>Land cover</th>
<th>n</th>
<th>$Z_{GPS}$ min</th>
<th>$Z_{GPS}$ max</th>
<th>$Z_{GPS}$ $\mu$</th>
<th>$Z_{GPS}$ $\sigma$</th>
<th>$Z_{LiDAR}$ min</th>
<th>$Z_{LiDAR}$ max</th>
<th>$Z_{LiDAR}$ $\mu$</th>
<th>$Z_{LiDAR}$ $\sigma$</th>
<th>$\mu_{\Delta_{\text{min}}}$</th>
<th>$\mu_{\Delta_{\text{max}}}$</th>
<th>$\mu_{\Delta_{\mu}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consolidated</td>
<td>256</td>
<td>-0.55</td>
<td>1.1</td>
<td>-0.05</td>
<td>0.33</td>
<td>-0.37</td>
<td>1.21</td>
<td>0.17</td>
<td>0.34</td>
<td>0.17</td>
<td>0.26</td>
<td>0.22</td>
</tr>
<tr>
<td>Mangrove</td>
<td>162</td>
<td>-0.55</td>
<td>0.26</td>
<td>-0.14</td>
<td>0.11</td>
<td>-0.37</td>
<td>0.96</td>
<td>0.11</td>
<td>0.24</td>
<td>0.19</td>
<td>0.30</td>
<td>0.25</td>
</tr>
<tr>
<td>Sawgrass</td>
<td>72</td>
<td>-0.26</td>
<td>0.06</td>
<td>-0.16</td>
<td>0.06</td>
<td>-0.20</td>
<td>0.84</td>
<td>0.03</td>
<td>0.18</td>
<td>0.15</td>
<td>0.24</td>
<td>0.19</td>
</tr>
<tr>
<td>Road</td>
<td>22</td>
<td>0.51</td>
<td>1.10</td>
<td>0.98</td>
<td>0.13</td>
<td>0.60</td>
<td>1.21</td>
<td>1.06</td>
<td>0.12</td>
<td>0.04</td>
<td>0.12</td>
<td>0.08</td>
</tr>
</tbody>
</table>

For our study area:
- LiDAR continuously overestimates ground
- It is more difficult for LiDAR to penetrate surface water with dark peat soils below low-density mangroves compared to high density sawgrass
Figure 2 Relationships between actual and predicted elevation of all land covers for calibration and validation data.
# Results: Corrected LiDAR DEMs

Table 2 Descriptive statistics for each experiment. Where RFR = Random Forest Regression, MB = Minimum Binning, $MB_{\text{calibrated}} = MB$ calibrated by the respective mean bias, $\Delta \mu = \text{mean difference between experiment values and GPS}$, $n = \text{number of GPS}$.

<table>
<thead>
<tr>
<th>Land cover</th>
<th>Experiment</th>
<th>RMSE$_Z$</th>
<th>$\Delta \mu$</th>
<th>Median</th>
<th>Skew</th>
<th>$\sigma$</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consolidated</td>
<td>RFR</td>
<td>0.09</td>
<td>0.01</td>
<td>0.00</td>
<td>1.54</td>
<td>0.09</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>$MB_{\text{calibrated}}$</td>
<td>0.20</td>
<td>0.05</td>
<td>0.00</td>
<td>1.22</td>
<td>0.19</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>MB</td>
<td>0.26</td>
<td>0.17</td>
<td>0.10</td>
<td>1.23</td>
<td>0.20</td>
<td>185</td>
</tr>
<tr>
<td>Mangrove</td>
<td>RFR</td>
<td>0.11</td>
<td>0.01</td>
<td>0.00</td>
<td>1.55</td>
<td>0.11</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>$MB_{\text{calibrated}}$</td>
<td>0.22</td>
<td>0.08</td>
<td>0.10</td>
<td>0.43</td>
<td>0.21</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>MB</td>
<td>0.23</td>
<td>0.15</td>
<td>0.12</td>
<td>0.53</td>
<td>0.18</td>
<td>132</td>
</tr>
<tr>
<td>Sawgrass</td>
<td>RFR</td>
<td>0.09</td>
<td>0.01</td>
<td>0.01</td>
<td>0.87</td>
<td>0.09</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>$MB_{\text{calibrated}}$</td>
<td>0.21</td>
<td>0.03</td>
<td>-0.03</td>
<td>1.64</td>
<td>0.21</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>MB</td>
<td>0.15</td>
<td>0.12</td>
<td>0.10</td>
<td>1.11</td>
<td>0.09</td>
<td>42</td>
</tr>
</tbody>
</table>
Primary purpose of this study: develop models to correct LiDAR for improved DEM product that meets Everglades’ elevation error requirements (±15 cm).

- Our results demonstrate the RFR-based DEM exceeds these requirements:
  - Mangrove = σ and RMSE = 11 cm; Δμ = 1 cm
  - Sawgrass = σ and RMSE = 9 cm; Δμ = 1 cm

This work has potential to better assist restoration projects with more accurate, precise, and high resolution DEMs.
Acknowledgements

This study was supported by the 2016 Mel Marcus Fund for Physical Geography Grant.

Many thanks to Everglades National Park staff!

Coastal Everglades American Crocodile
References


Thank you!
Nature of LiDAR Elevation Data

1. Light Detection and Ranging (LiDAR) elevation data collection

2. Process LiDAR where vegetation points are removed and 'ground' points remain

3. LiDAR 'ground' points never truly reach the ground, but survey-grade GPS will
Results: RFR DEM
Future Research

- Test this approach to predict e.g. HAED for other vegetation communities within the entire coastal Everglades.

- Incorporate accretion rates into DEMs for modeling water depth change due to regionally adjusted sea-level rise projections.