Validation and Assimilation of OSM for the Global Roads Open Access Data Set

Alex de Sherbinin & Paola Kim-Blanco

CIESIN, Columbia University
CODATA Global Roads Data Development Working Group

Summary

- OSM data are increasingly viewed as authoritative by development / humanitarian agencies to carry out relief and development work.
- However, a question that often arises is, **how complete are the data in OSM for country X?**
- To help answer this question, we are collaborating with OSM to **assess road length, spatial coverage, attributes, and the geometry of OSM data**.
- Goal is to **assimilate** OSM data into Global Roads Open Access Data Set (gROADS), version 2.
Global Roads Open Access Data Set (gROADS)

Goal: To develop a global roads open access data set (gROADS) that is:

1. fit for modeling / research purposes (not navigation)
2. Based on a globally consistent model (UNSDI-T)
3. spatially accurate (~50m positional accuracy)
4. topologically integrated
5. focused on roads between settlements (not streets)
6. well documented
7. freely distributed (on attribution only basis)
Prior OSM Validation work

West Africa Analysis

- Downloaded data in **May 2014** for a USAID project
- Downloaded data from Geofabrik in **May 2015**
- A very low income (OSM lagging) region
- Several countries (Guinea, Liberia and Sierra Leone) experienced a major outbreak of Ebola in 2014-15
Liberia: example random sampling procedure

Accuracy testing
Positional accuracy of OSM is generally very good

Results from accuracy testing

<table>
<thead>
<tr>
<th></th>
<th>Country</th>
<th>Ebola affected</th>
<th>RMSE (meters)</th>
<th>NSSDA accuracy (meters)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ivory Coast</td>
<td>No</td>
<td>49</td>
<td>84</td>
</tr>
<tr>
<td>2</td>
<td>Cameroon</td>
<td>No</td>
<td>31</td>
<td>54</td>
</tr>
<tr>
<td>3</td>
<td>Guinea-Bissau</td>
<td>No</td>
<td>18</td>
<td>30</td>
</tr>
<tr>
<td>4</td>
<td>Liberia</td>
<td>Yes</td>
<td>22</td>
<td>38</td>
</tr>
<tr>
<td>5</td>
<td>Sierra Leone</td>
<td>Yes</td>
<td>13</td>
<td>23</td>
</tr>
<tr>
<td>6</td>
<td>Guinea</td>
<td>Yes</td>
<td>21</td>
<td>36</td>
</tr>
</tbody>
</table>

Some seemingly systematic spatial displacement in certain regions
Road length in ebola affected countries more than tripled, compared to a 50% increase for the other countries

<table>
<thead>
<tr>
<th>Country</th>
<th>May 2014 length (km)</th>
<th>May 2015 length (km)</th>
<th>Ebola affected</th>
<th>% change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ivory Coast</td>
<td>32,769</td>
<td>59,081</td>
<td>0</td>
<td>80%</td>
</tr>
<tr>
<td>Cameroon</td>
<td>60,124</td>
<td>84,386</td>
<td>0</td>
<td>40%</td>
</tr>
<tr>
<td>Guinea Bissau</td>
<td>3,109</td>
<td>4,209</td>
<td>0</td>
<td>35%</td>
</tr>
<tr>
<td>Liberia</td>
<td>8,715</td>
<td>32,037</td>
<td>1</td>
<td>268%</td>
</tr>
<tr>
<td>Guinea</td>
<td>33,742</td>
<td>99,111</td>
<td>1</td>
<td>194%</td>
</tr>
<tr>
<td>Sierra Leone</td>
<td>11,459</td>
<td>50,401</td>
<td>1</td>
<td>340%</td>
</tr>
<tr>
<td><strong>Non-Ebola countries</strong></td>
<td><strong>96,003</strong></td>
<td><strong>147,677</strong></td>
<td><strong>0</strong></td>
<td><strong>54%</strong></td>
</tr>
<tr>
<td><strong>Ebola countries</strong></td>
<td><strong>53,917</strong></td>
<td><strong>181,551</strong></td>
<td><strong>1</strong></td>
<td><strong>237%</strong></td>
</tr>
</tbody>
</table>
Liberia: number of edits per road segment
Liberia: length increase by type

268% increase in one year
Validation and assimilation challenges

- OSM is rapidly evolving → a moving target
  - Validation at any point in time will soon be obsolete
  - Especially in countries that later experience disasters

- Absence of independent “objective” data in low income countries
  - Few authoritative commercial providers
  - High resolution imagery patchy in many countries

- Levels of detail and road types vary by country and may require simplifying and recoding for consistency

Ideas are welcome! [www.groads.org](http://www.groads.org)