High Resolution Mapping of Chimpanzee Habitat Suitability

Society for Conservation GIS Annual Conference

Pacific Grove, CA

7-27-2015

Sam Jantz – University of Maryland
Dr. Lilian Pintea – Jane Goodall Institute
Dr. Janet Nackoney – University of Maryland
Dr. Matthew C. Hansen – University of Maryland
Dr. Peter Potapov – University of Maryland
Presented research is in support of the project:
“Monitoring and Forecasting Chimpanzee Habitat Health in Africa to Inform Conservation Actions, Strategies and Measure Success.”

Project Objective:
Develop a practical and operational Decision Support System that could be used by the Jane Goodall Institute and partners to annually monitor and forecast chimpanzee habitat suitability.

Funded by:
NASA Ecological Forecasting for Conservation and Natural Resource Management Program.
Collective range size is approximately 2,342,000 km², close to 1/3 of the continental U.S. land area.

Four identified sub-species of *Pan troglodytes* occupy a variety of forest types, from dry savanna woodlands and woodland-forest mosaics to humid closed canopy and montane forests (Goodall, 1986; Teleki, 1989; Jones et al, 1996; Nowak, 1999).
Population Distribution and Status

- Cumulative chimpanzee population is estimated to have declined by more than 60% over last 40 years (Kormos et al. 2003).
- Between 172,000-300,000 individuals in the wild (Caldecott & Miles. 2005).
- Listed as vulnerable in 1986, status updated to endangered in 1996 (IUCN Red List).
- Populations continue to decline (Oates et al. 2008).
Threats

- Primary threats are habitat destruction and degradation along with bushmeat hunting, illegal pet trade and disease.
- Approximately 70% of chimpanzee tropical forest habitat is threatened by infrastructure development and land use change (Nellemen & Newton 2002).
Relatively few studies exist on chimpanzee habitat suitability mapping and have been at the local scale, coarse in resolution or both.

Only two studies have attempted to address habitat change through time (Torres et al. 2010 & Junker et al. 2012).

Junker et al. (2012) used chimpanzee presence locations from the IUCN A.P.E.S. database and environmental variables that include information on climate, vegetation and human impact as inputs into a suitability model for each sub-species.
Background

• MAXENT and logistic regression were used to relate presence locations to environmental variables and generate habitat suitability models and maps at 5km resolution.

• Multiple chimpanzee researchers were involved in the development of the Junker et al. (2012) model and journal manuscript therefore the suitability data represent the most currently accepted and published state of chimpanzee habitat suitability across the field.
**Objective:**
- Generate a temporally dynamic map of chimpanzee habitat suitability at a spatial resolution relevant to conservation planning and decision making.

**Method:**
- Downscale coarse resolution habitat suitability map using annually updated remotely sensed data sets.

**Definitions:**
- Junker et al (2012) define suitability as the probability of chimpanzee occurrence given certain environmental conditions.
- Down-scaling method does not map probability of occurrence at 30m pixel scale, but rather provides information on areas that are more or less conducive for chimpanzee occupancy or use.
15 Landsat ETM+ based “dynamic” variables derived from Hansen et al. (2013) data.

<table>
<thead>
<tr>
<th>Spectral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Band 3 median reflectance 2000-2005</td>
</tr>
<tr>
<td>Band 4 median reflectance 2000-2005</td>
</tr>
<tr>
<td>Band 5 median reflectance 2000-2005</td>
</tr>
<tr>
<td>Band 7 median reflectance 2000-2005</td>
</tr>
<tr>
<td>Normalized difference (band4/band3)</td>
</tr>
<tr>
<td>Normalized difference (band4/band5)</td>
</tr>
<tr>
<td>Normalized difference (band4/band7)</td>
</tr>
<tr>
<td>band3/band5</td>
</tr>
<tr>
<td>band3/band7</td>
</tr>
<tr>
<td>band5/band7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Forest structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent canopy cover</td>
</tr>
<tr>
<td>Canopy height</td>
</tr>
<tr>
<td>Percent bare ground</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Disturbance and fragmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proximity to forest loss (minimum 0.5 ha)</td>
</tr>
<tr>
<td>Proximity of interior forest to forest edge (minimum 1 ha patch size)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Topographic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elevation</td>
</tr>
<tr>
<td>Slope</td>
</tr>
<tr>
<td>Proximity to steep slopes (&gt;15°)</td>
</tr>
<tr>
<td>Proximity to rivers</td>
</tr>
</tbody>
</table>

5 Shuttle Radar Topography Mission based “static” variables.
Model Flow

Junker et al. (2012) model 5km resolution

Mosaic → Aggregation to 5km

Regional suitability model for western Tanzania 90m resolution

5km habitat suitability map

Co-registration → Data-split 90% training 10% testing 5-fold

Random Forest Regressor

Model diagnostics

30m habitat suitability map 2000-2005 → Model assessment using crowd sourced data

Dynamic variables 2000-2005

Independent Variables

Static variables
Results
Model Diagnostics

<table>
<thead>
<tr>
<th>Sub-species</th>
<th>RMSE %</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>P.t. verus</em></td>
<td>2.90 (2.91-2.89)</td>
</tr>
<tr>
<td><em>P.t. ellioti</em></td>
<td>4.86 (4.88-4.85)</td>
</tr>
<tr>
<td><em>P.t. troglodytes</em></td>
<td>6.04 (6.04-6.03)</td>
</tr>
<tr>
<td><em>P.t. Schweinfurthii</em></td>
<td>3.75 (3.78-3.72)</td>
</tr>
<tr>
<td>Combined</td>
<td>4.77 (4.78-4.77)</td>
</tr>
</tbody>
</table>

Pseudo $R^2 = 0.94 +/- 0.0001$

---

### Model Diagnostics

<table>
<thead>
<tr>
<th>Percent contribution</th>
<th>Elevation</th>
<th>Band5 reflectance</th>
<th>Percent canopy cover</th>
<th>Distance to rivers</th>
<th>Band7 reflectance</th>
<th>Band3 reflectance</th>
<th>Band4 reflectance</th>
<th>Canopy height</th>
<th>Band5/Band7</th>
<th>Normalized difference Band4/Band4</th>
<th>Distance to forest edge</th>
<th>Normalized difference Band4/Band5</th>
<th>Band3/Band5</th>
<th>Band3/Band7</th>
<th>Percent bare ground</th>
</tr>
</thead>
</table>
# Model Diagnostics

Pseudo $R^2 = 0.94 \pm 0.0001$

<table>
<thead>
<tr>
<th>Sub-species</th>
<th>RMSE %</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P.t.\text{verus}$</td>
<td>2.90 (2.91-2.89)</td>
</tr>
<tr>
<td>$P.t.\text{elliotti}$</td>
<td>4.86 (4.88-4.85)</td>
</tr>
<tr>
<td>$P.t.\text{trogloidytes}$</td>
<td>6.04 (6.04-6.03)</td>
</tr>
<tr>
<td>$P.t.\text{schweinfurthii}$</td>
<td>3.75 (3.78-3.72)</td>
</tr>
<tr>
<td>Combined</td>
<td>4.77 (4.78-4.77)</td>
</tr>
</tbody>
</table>

![Model Diagnostics Diagram](image-url)
5km resolution

30m resolution

**Suitability**
- High: 1
- Low: 0

**Sub-species**
- Pt. elliotti
- Pt. schweinfurthii
- Pt. troglodytes
- Pt. verus
5km resolution

30m resolution

Suitability

High : 1

Low : 0
5km resolution

30m resolution

Suitability

High: 1
Low: 0

0 5 10 Kilometers

0 5 10 Kilometers
30m resolution 2000 - 2005

30m resolution 2005-2010

Suitability

High: 1

Low: 0
Validation with crowd-sourced location data

- Chimpanzee location data used for evaluation were collected from a variety of sources including research surveys, protected area rangers and local community monitoring efforts.
- Android Nexus tablets with mapping applications, such as Open Data Kit (ODK) were utilized by local communities and rangers to collect geo-referenced location data.
- 9,657 unique locations were extracted from point database.
- Spatial sampling bias was addressed by randomly sampling one point per 1km grid-cell resulting in 2,105 points used for map evaluation.
$F_i = \frac{P_i}{E_i}$
Conclusions and Further Work

• The method outlined here demonstrates a promising effort to map suitability, as well as change in suitability, at a fine scale.

• Further analysis is needed to determine if the change in suitability is biologically relevant.

• Method assumes the model is spatial scale invariant, a strong and possibly incorrect assumption.

• However, several input variables have been demonstrated to be insensitive to spatial scale, including canopy cover (Sexton et al. 2013) and surface reflectance (Gao et al. 2006; Hilker et al. 2009).

• Additional location data will be needed for a more robust map validation.
Questions?
References

Method Overview

• Combine Junker et al. (2012) suitability map with suitability map created specifically for western Tanzania.

• Use variables derived from remotely sensed data sets and coarse scale suitability map as inputs to train a Random Forest regression model and then predict suitability for entire range at the 30m pixel scale.

• Random Forests (Breiman et al. 2001) are non-parametric, non-linear models that average over a large collection of decision/regression trees, each trained on a separate bootstrap sample of the input data and sub-set of the features.

• 250 trees were grown to full size under the constraint that a minimum of 5 samples are present in each internal node.

• All models were generated using the Scikit Learn module (Pedregosa et al. 2011) in the python programming language.

• A total of 182,731 grid-cells were extracted from the coarse scale map.

• Use 5-fold cross-validation with 90% of data used for training and 10% used to assess model fit.

• Use independently collected location data to provide a preliminary validation of the generated map.
Validation Procedure

• Map accuracy was assessed for a portion of the *P.t.schweinfurthii* (Eastern) range using the procedure outlined in Hirzel et al. (2006) developed for situations where presence only data are available.

• Method entails partitioning habitat suitability range into *b* classes. For each class *i*, calculate *P*<sub>i</sub>, the predicted frequency of evaluation points given by:

\[
P_i = \frac{p_i}{\sum_{j=1}^{b} p_j} \quad \text{(Eq.1)}
\]

• Where *p*<sub>i</sub> is the number of evaluation points predicted by the model to fall in suitability class *i* and \(\sum p_j\) is the total number of evaluation points.

• Next, the frequency expected from a random distribution across the study area is calculated:

\[
E_i = \frac{a_i}{\sum_{j=1}^{b} a_j} \quad \text{(Eq.2)}
\]

• Where *a*<sub>i</sub> is the number of random points predicted by the model to fall in suitability class *i* and \(\sum a_j\) is the total number of random points.

• Finally, the predicted to expected ratio for class *i* is given by:

\[
F_i = \frac{P_i}{E_i} \quad \text{(Eq.3)}
\]

• *Fi* was calculated for 20 classes over the range 0 – 1 using an increment of 0.05.

• Five sets of 10,000 random points were used to calculate *E*<sub>i</sub>.