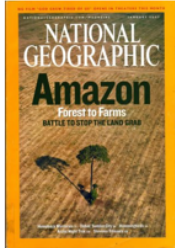# ForestEyes Project: Can Citizen Scientists Help Rainforests?

Fernanda B. J. R. Dallaqua, Álvaro L. Fazenda, and Fabio A. Faria

Instituto de Ciência e Tecnologia
Universidade Federal de São Paulo, São José dos Campos

25/09/2019

ForestEyes project main goal

Use volunteer contributions to detect deforestation's areas in a tropical rain forest, joined to a future semi-automatic classifier based on Machine Learning.

- ▶ Developed in 1988
- ▶ Gives annual deforestation surveys in Brazilian Legal Amazon
- ▶ Uses Landsat imagery
- ▶ From 2003 started to use a computer-assisted interpretation process
- ▶ Bands red, near-infrared and shortwave infrared are used to generate fraction images of the components soil, vegetation and shade
- ▶ The soil and shade fraction images segmentation and classification are performed next
- ▶ An expert analyzes the thematic polygons, agreeing or correcting the automatic classification

Background and Related Works

PRODES

- ▶ Provides the deforestation stats and classified mosaics
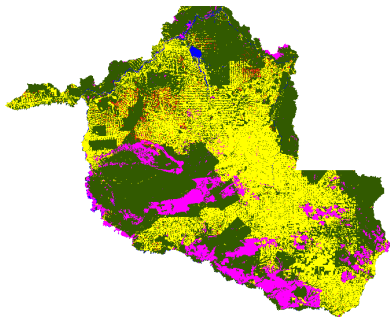- ▶ The mosaics, until 2016, had 60m resolution

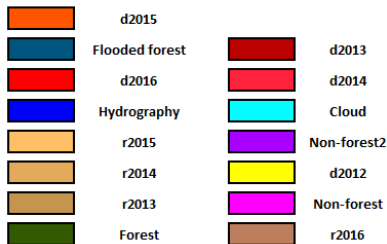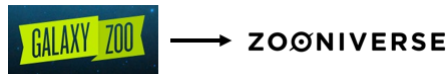Figure 1: Rondônia state classified by PRODES (2016).

Figure 2: Color code for PRODES 2016 image.

- Christmas Bird Count: First and oldest Citizen Science Project (1900)
- High volume of processed data and with low cost
- Information and Communication Technology: Citizen Cyberscience
  - Volunteered Computing
  - Volunteered Thinking
  - Participatory Sensing

- ▶ Volunteers' motivation
 - ▶ Altruism
  - ▶ Contribution for research
  - ▶ Interest in science
  - ▶ Online communities
  - ▶ Competitiveness
- ▶ Data quality: efficient as specialists
- ▶ But some validation mechanisms are needed
  - ▶ Send redundant tasks to multiple users
  - ▶ Calibration tasks
  - ▶ Comparison with volunteers' consensus
  - ▶ Assign weights to individual users according to their skill

Background and Related Works
ForestWatchers

- Developed in 2012
- Citizen Science to track rainforests' deforestation
- Used MODIS sensor's imagery (250m resolution)
- Had 3 applications: Best-Tile, Deforestation and Correct Classification
- Two areas inspected in Correct Classification: Rondônia 2011 and Awá-Guajá 2014

ForestEyes

- Inspired by ForestWatchers' Correct Classification
- Ally Citizen Science with Machine Learning
  - Volunteers classify remote sensing areas into Forest, Non-forest or Undefined
  - Volunteers' classifications will be used to train an automatic classifier
- To classify the remote sensing areas, volunteers need to analyze:
  - If the area have 70% or more pixels of one class → Classify the area of this class
  - If it isn't → Classify the area as Undefined

ForestEyes

- Hosted by Zooniverse.org
- Beta Review: same tasks as ForestWatchers' Correct Classification plus 6 more tasks from Awá-Guajá 2014
  - But without showing area classified by Artificial Neural Network
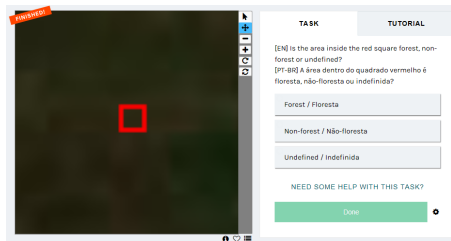
- ► Complaints about:
  - ► Image's resolution
  - ► Image too dark
  - ► Display of the tasks
  - ► Tutorial
- ► Proposed solutions
  - ► Remote sensing images from Landsat-8 resampled to 60m resolution, according to PRODES
  - ► Use of a different color composition besides RGB
  - ► Segments instead of fixed squares: SLIC technique
  - ► Improvement of the tutorial
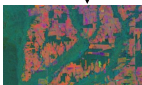
Resampling to 60m and crop the area of interest
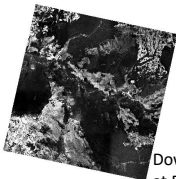
Download of Landsat-8 scene
at EarthExplorer portal (7 bands, 30m resolution)
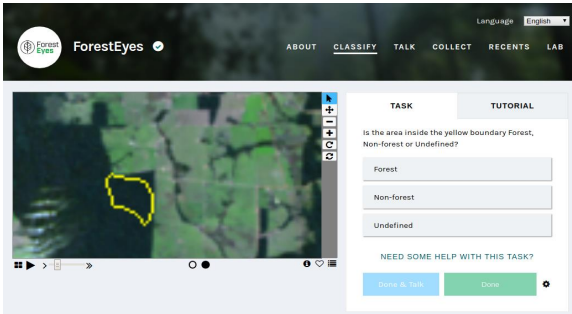
Apply PCA technique to reduce from 7 to 3 components

SLIC to segment image
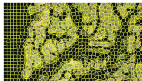
Each segment becomes a task

► Image from an area of Rondônia in the year of 2016 with 1022 tasks
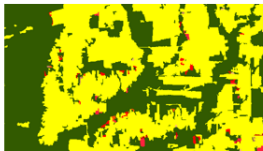
(a) RGB composition   (b) PRODES classification   (c) Binarized PRODES
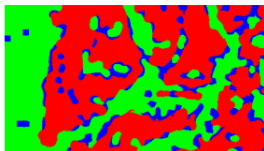
(d) Groundtruth with undefined class
(blue pixels)
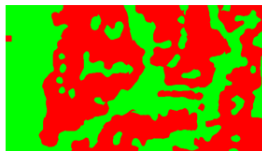
(e) Groundtruth with majority

- One week after the official launch all the tasks were completed
- A new set of tasks was built. This time for the same area of Rondônia but now from 2013, with 1027 tasks
    - Purpose of seeing if the changes between 2013 and 2016 could be noticed by the volunteers
    - Same building steps as Landsat-8 segments 2016
- With one week all the tasks for 2013 were completed

▶ Registered volunteers answered more tasks than anonymous → Some registered answered A LOT of tasks (for Landsat-8)

TABLE I
CS WORKFLOWS' GENERAL INFORMATION.

	ForestWatchers' CC		ForestEyes		
	Rondônia 2011	Awá-Guajá 2014	ForestWatchers tasks	Landsat-8 segments 2016	Landsat-8 segments 2013
Tasks	72	26	104	1022	1028
Answers per task	15	30	15	15	15
Expected number of answers	1,080	780	1,560	15,330	15,420
Total answers	1,080	780	2,275	19,803	15,610
Registered volunteers	15	17	77	151	137
Anonymous volunteers	31	36	125	76	41
Number of answers (registered)	522	348	1,604	17,239	14,758
Number of answers (anonymous)	558	432	671	2,564	852

Convergence Evaluation of answers
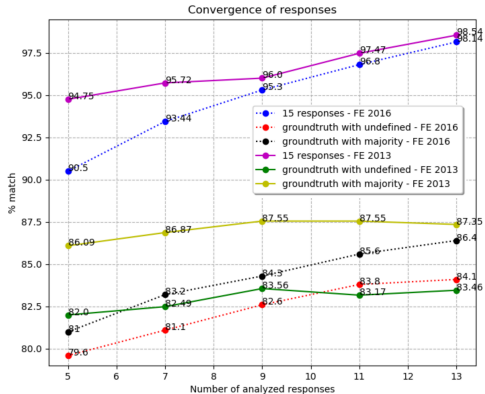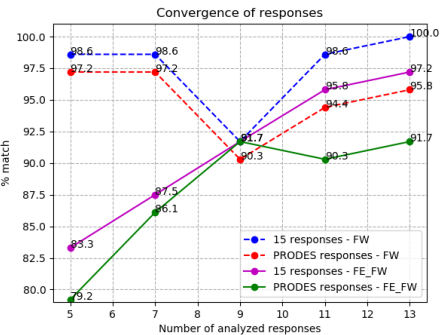
► Decision of using the first 15 answers for ForestEyes' workflows

Citizen Science Accuracy

- For Rondônia 2011, comparing to PRODES:
  - ForestWatchers' Correct Classification: 95.8%
  - ForestEyes: 88.9%
- Volunteers achieved better performance using *groundtruth with majority*
  - Volunteers could be labeling the segment according to the majority class instead of analyzing if there are 70% or more pixels of one class

TABLE II
CS ACCURACY FOR LANDSAT-8 SEGMENT TASKS

	Landsat-8 segments	
	2016	2013
PRODES	85.42%	83.89%
Groundtruth with majority	86.5%	87.55%
Groundtruth with undefined	84.25%	83.66%

Volunteer hit rate's behavior

▶ Volunteers improve their ranking as more tasks are answered.

Volunteers Ranking
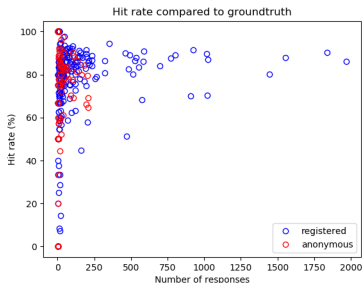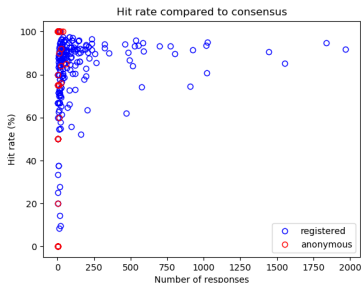
▶ The volunteers' Hit Rare (HR) and scores (VS) is calculated through:

$$HR = \frac{hits}{total\_answers} \times 100 \qquad (1)$$

$$VS = (0.3 \times number\_answers) + (0.7 \times hits) \qquad (2)$$

TABLE III
REGISTERED VOLUNTEER SCORES FOR LANDSAT-8 SEGMENTS TASKS IN FORESTEYES.

User_id	Number answers	Consensus hits	Consensus hit rate (%)	Score	Hit rate groundtruth with undefined (%)	Hit rate groundtruth with majority (%)
1805790	1971	1809	91.78	1857.6	88.33	86.0
1061480	1840	1744	94.78	1772.8	86.25	90.16
1893875	1553	1323	85.19	1392	88.15	87.96
1851320	1446	1308	90.46	1349.4	79.94	80.29
...	...	...	...	...	...	...

Task Difficulty Level

- The difficulty of each task can be calculated by Shannon's Entropy

$$H = - \sum_{i=1}^{n} p_i \times \log_2 p_i \qquad (3)$$

Where p_i is the probability of the class i be chosen, calculated by the ratio between the number of votes given to class i and the total of votes for the task, and n is the number of possible classes in the task.

TABLE VI
TASK ENTROPY'S CLASSIFICATION FOR ALL THE CS WORKFLOWS.

Entropy	Classification	ForestWatchers Rondônia 2011	ForestWatchers Awá-Guajá 2014	ForestEyes Rondônia 2011	ForestEyes Awá-Guajá 2014	ForestEyes Landsat-8 segments
		Frequency				
Up to 0.33	Easy	10 (13.89%)	24 (92.31%)	4 (5.56%)	18 (56.25%)	1329 (64.83%)
Between 0.33 and 0.66	Medium	24 (33.33%)	2 (7.69%)	26 (36.11%)	12 (37.5%)	438 (21.36%)
Above 0.66	Hard	38 (52.78%)	0 (0%)	42 (58.33%)	2 (6.25%)	283 (13.8%)

Evaluation of Volunteer Variability

▶ The volunteers' variability can be calculated with Shannon's entropy by replacing p_i with a normalized weight w_j calculated with the volunteers' scores s_j

$$w_j = \frac{s_j}{\sum\limits_{i=1}^{V} s_i} \tag{4}$$

Where V is the number of volunteers, and s_i is the score of the ith volunteer.

TABLE VII
VOLUNTEERS VARIABILITY.

	Registered (H_R)	Anonymous (H_A)
ForestWatchers' Rondônia 2011	2.16	2.38
ForestEyes' ForestWatchers	3.94	2.27
ForestEyes' Landsat-8 segments	5.78	0.80

Comparison between Landsat-8 Segments Workflows

- ▶ Was taken the difference between Landsat-8 segments 2016 and Landsat-8 segments 2013
- ▶ Difference between PRODES 2016 and PRODES 2013 - 2184 new deforested pixels
- ▶ From these 2184 pixels with new deforestation, 1163 also appeared in the difference of Landsat-8 segments
  - ▶ 570 pixels correctly classified as non-forest
  - ▶ 302 were labeled as undefined
  - ▶ 176 occurred ties
  - ▶ 115 were wrongfully classified as forest
- ▶ More investigation is needed to explain why differences over time weren't fully noticed
  - ▶ Error in segmentation
  - ▶ Error in tasks display
  - ▶ Satellite variability
  - ▶ Error in volunteers' classification

Conclusion

- ► ForestEyes is a Citizen Science project with the goal of tracking rainforests' deforestation
- ► It was inspired in the late ForestWatchers' Correct Classification
- ► Volunteers classify remote sensing segments into Forest, Non-forest or Undefined
- ► Volunteers had accuracy higher than 83%
- ► 2049 tasks were completed in 2.5 weeks
- ► MODIS images appear to be more difficult to classify $\rightarrow$ worst resolution
- ► Citizen Science: powerful tool that can complement data from official monitoring programs

Future Work

- New ForestEyes' campaigns (you can help at `https://www.zooniverse.org/projects/dallaqua/foresteyes`)
- Use volunteers' classification in an Active Learning procedure to train an automatic classifier
- Improve resolution and segmentation method
- Assign weights to individual volunteers according to their ranking

Acknowledgment