Objective:

Determine the annual Hansen tree cover loss and emissions for 2001-2017 using the same criteria that Brazil uses for its official Amazon emissions statistics in order to compare Brazil’s official national record with GFW’s independent data

Data sources:

* PRODES primary forest annual boundaries from the end of 2006/beginning of 2007 to 2017: <http://www.dpi.inpe.br/prodesdigital/dadosn/mosaicos/2017/PDigital2000_2017_AMZ_gtif.zip>
  + This raster only goes back to deforestation in 2007, so it can only be used to reconstruct primary forest from 2007 to 2017
  + This raster did not align with Hansen pixels. Got this to align with Hansen pixels using gdal\_warp: C:\GIS\GFW\_Climate\Brazil\_emis\_comparison\PRODES>gdalwarp -t\_srs EPSG:4326 -tr 0.00025 0.00025 -tap PDigital2017\_AMZ\_30m.tif PDigital2017\_AMZ\_30m\_warp.tif
  + Verified that this raster aligned with Hansen pixels.
* PRODES primary forest annual boundaries from the end of 2000/beginning of 2001 to 2014: <http://www.dpi.inpe.br/prodesdigital/dadosn/mosaicos/2014/PDigital2000_2014_AMZ_gtif.zip>
  + This raster can be used to reconstruct primary forest from the start of 2001 to the start of 2014
  + This raster did not align with Hansen pixels. Its resolution is also coarser than the 2007-2016 PRODES raster. Got this to align with Hansen pixels using gdal\_warp: C:\GIS\GFW\_Climate\Brazil\_emis\_comparison\PRODES>gdalwarp -t\_srs EPSG:4326 -tr 0.00025 0.00025 -tap Prodes2014\_AMZ\_60m.tif Prodes2014\_AMZ\_60m\_warp.tif
  + Verified that this raster aligned with Hansen pixels.
* Hansen annual tree cover loss: s3://gfw2-data/forest\_change/hansen\_2017/
* Annual fire tree cover loss from MODIS: <ftp://fuoco.geog.umd.edu/MCD64A1/C6>
* Legal Amazon boundary: same as Liz and Mikaela used in their blog post. Zipped shapefile sent via Slack on 8/21/18. Currently at C:\GIS\Multi\_project\Brazil\_legal\_Amazon\.
  + This shapefile is clearly based on a raster but the raster it is from was not aligned with Hansen loss pixels. This caused shifting of the resulting layers during the analysis.
  + Converted this to a raster, aligning it with Hansen loss pixels [following these directions](https://support.esri.com/en/technical-article/000013221) (using the Processing Extent, Output Coordinates, and Raster Analysis environments) in Polygon to Raster conversion.
  + Verified that this raster aligned with Hansen pixels.

Processing methods:

* Clipped Hansen loss 2001-2017 to Brazil’s legal Amazon boundary (from Liz Goldman) so that all results would be constrained by that.
* Generated map of which Hansen loss pixels from 2001 to 2017 had burning that year or the preceding year. This method was developed by Sam Gibbes for the global forest carbon flux model.

1. Downloaded raw hdf files from ftp site to s3
   1. Ran carbon-budget/burn\_date/multi\_thread\_download\_upload\_raw\_hdf.py
   2. Saved raw hdf to s3:// gfw2-data/climate/carbon\_model/other\_emissions\_inputs/burn\_year/raw\_hdf/
2. Stacked raw hdf files by year. This produced a raster for each year (e.g., 2015) for each hdf tile (e.g., h06\_v11) that says the last day in each year that a pixel was burned
   1. Ran carbon-budget/burn\_date/multi\_thread\_stack\_ba\_hv.py
   2. Saved hdf stacked by year to s3://gfw2-data/climate/carbon\_model/other\_emissions\_inputs/burn\_year/burn\_year/
3. Clipped files to Hansen tile size, resolution, alignment, etc. This produced a Hansen-sized tile (e.g., 10N\_100E) for each year (e.g., 2012) with the last date each pixel was burned in that year.
   1. Ran carbon-budget/burn\_date/multi\_thread\_clip\_year\_tiles.py
   2. Saved s3:// gfw2-data/climate/carbon\_model/other\_emissions\_inputs/burn\_year/burn\_year\_10x10\_clip/
4. Flattened burn year tile to get most recent burn year:
   1. Ran carbon-budget/burn\_date/hansen\_burnyear.py
   2. Saved burn year tiles to s3://gfw-files/sam/carbon\_budget/burn\_loss\_year/
5. Initially, I tried putting the tiles into a mosaic dataset in a geodatabase. However, the script kept failing on the Con function for reasons I couldn’t figure out. So I merged all 12 burn year tiles in the study area into a single raster (ugggh!) and that still didn’t work. But when I created a raster attribute table for the merged burn year raster, the script did work on it. So, I used the merged burn year raster (burnyear\_merge\_20181013.tif) for the analysis.

* Created annual primary forest rasters from PRODES rasters for 2007 to 2017 using the PRODES 2017 raster:
  + Legend for PRODES rasters of different years at: <http://www.dpi.inpe.br/prodesdigital/dadosn/mosaicos/class_rgb.txt>.
  + Needed to create a separate primary forest raster for each year to include in the loss/emissions analysis for each year.
  + For Liz Goldman and Mikaela Weiss’s analysis of 2017 loss comparison using the 2017 PRODES raster, primary forests were codes 1, 16, and 24 (FLORESTA, d2017, r2017).
  + This is on the principal that to get primary forest at the start of 2017, you remove deforestation from all previous years, but not from 2017, which hadn’t occurred yet. I applied the same principle to get primary forest at the start of each year from 2007 to 2017. I confirmed this approach with Marcelo Matsumoto (forest GIS analyst in Brazil) during a phone call and subsequent e-mails.
  + More explicitly: to get primary forest for each year from 2007 to 2017, I included code 1 (FLORESTA) and the d and r codes from the year of interest to 2017. For example, primary forest in 2009 was codes 1 and 8-24 (FLORESTA, d2009-d2017, r2010-r2017) and primary forest in 2014 was codes 1, 13-16, and 21-24 (FLORESTA, d2014-d2017, r2014-r2017). I did not include cloud (code 5, nuvem) pixels in primary forest because there was no way of knowing what was there.
  + I implemented this in a function in Brazil\_emis\_comparison/utilities.py, which used the ArcPy Reclassify geoprocessing command to reclassify the 2017 PRODES raster into individual annual primary forest rasters. Maybe I could’ve written the overall analysis to not need to create and save the annual PRODES rasters but this was easier and I wanted to have the annual rasters for my records and potential additional analyses. I used two different reclassification commands (2007-2009 and 2010-2017) because of the addition of the residual deforestation classes for 2010-2017.
  + QC: Checked that 2007, 2012, and 2017 were producing the expected outputs. Made sure to include rasters from before and after 2010 because the PRODES reclassification command was different for those two categories.
* Created annual primary forest rasters from PRODES rasters for 2000 to 2014 using the PRODES 2014 raster:
  + Legend for PRODES rasters at: <http://www.dpi.inpe.br/prodesdigital/dadosn/mosaicos/class_rgb.txt>
  + Needed to create a separate primary forest raster for each year to include in the loss/emissions analysis for each year. Each primary forest map was at the beginning of that year, i.e. the deforestation of that year had not occurred yet.
  + From e-mailing with Marcelo Matsumoto, I learned that codes 21 (desflorestamento) and 17 (residuo) were deforestation in the most recent year (2014). Thus, they I treated them as if they were code d2014, which meant they remained as primary forest for each of the years I was creating (even 2014, because that is primary forest at the beginning of 2014, so d2014 hadn’t occurred yet). I did not include cloud (code 6, nuvem) pixels in primary forest because there was no way of knowing what was there.
  + Because the legend for the raster is all out of order (deforestation from consecutive years don’t have consecutive codes), there wasn’t an easy way to automate the creation of the annual primary forest rasters from this file, unlike for the 2007-2017 rasters. So I created each primary forest raster manually in ArcMap using the Reclassify tool. The Python snippets are below. In the tool environments, I set the output raster resolution to be the same as the Hansen raster (0.00025 degrees) (Raster Analysis menu).
  + arcpy.gp.Reclassify\_sa("Prodes2014\_AMZ\_60m\_warp\_low\_res.tif", "Value", "0 NODATA;1 NODATA;2 NODATA;3 NODATA;4 NODATA;5 1;6 NODATA;7 NODATA;8 NODATA;9 NODATA;10 NODATA;11 NODATA;12 NODATA;13 NODATA;14 NODATA;15 NODATA;16 NODATA;17 1;18 NODATA;19 NODATA;20 NODATA;21 1;22 NODATA;23 NODATA;24 NODATA", "C:/GIS/GFW\_Climate/Brazil\_emis\_comparison/PRODES/PRODES\_primary\_forest\_2014\_early\_raster\_NoData\_v4.tif", "DATA")
  + arcpy.gp.Reclassify\_sa("Prodes2014\_AMZ\_60m\_warp\_low\_res.tif", "Value", "0 NODATA;1 NODATA;2 NODATA;3 NODATA;4 NODATA;5 1;6 NODATA;7 NODATA;8 NODATA;9 NODATA;10 NODATA;11 NODATA;12 NODATA;13 1;14 NODATA;15 NODATA;16 NODATA;17 1;18 NODATA;19 NODATA;20 NODATA;21 1;22 NODATA;23 NODATA;24 NODATA", "C:/GIS/GFW\_Climate/Brazil\_emis\_comparison/PRODES/PRODES\_primary\_forest\_2013\_early\_raster.tif", "DATA")
  + arcpy.gp.Reclassify\_sa("Prodes2014\_AMZ\_60m\_warp\_low\_res.tif", "Value", "0 NODATA;1 NODATA;2 NODATA;3 NODATA;4 NODATA;5 1;6 NODATA;7 NODATA;8 NODATA;9 NODATA;10 NODATA;11 NODATA;12 NODATA;13 1;14 NODATA;15 1;16 NODATA;17 1;18 NODATA;19 NODATA;20 NODATA;21 1;22 NODATA;23 NODATA;24 NODATA", "C:/GIS/GFW\_Climate/Brazil\_emis\_comparison/PRODES/PRODES\_primary\_forest\_2012\_early\_raster.tif", "DATA")
  + arcpy.gp.Reclassify\_sa("Prodes2014\_AMZ\_60m\_warp\_low\_res.tif", "Value", "0 NODATA;1 NODATA;2 NODATA;3 NODATA;4 NODATA;5 1;6 NODATA;7 NODATA;8 NODATA;9 NODATA;10 NODATA;11 NODATA;12 NODATA;13 1;14 NODATA;15 1;16 NODATA;17 1;18 NODATA;19 NODATA;20 NODATA;21 1;22 1;23 NODATA;24 NODATA", "C:/GIS/GFW\_Climate/Brazil\_emis\_comparison/PRODES/PRODES\_primary\_forest\_2011\_early\_raster.tif", "DATA")
  + arcpy.gp.Reclassify\_sa("Prodes2014\_AMZ\_60m\_warp\_low\_res.tif", "Value", "0 NODATA;1 NODATA;2 NODATA;3 NODATA;4 NODATA;5 1;6 NODATA;7 NODATA;8 NODATA;9 NODATA;10 NODATA;11 NODATA;12 NODATA;13 1;14 NODATA;15 1;16 1;17 1;18 NODATA;19 NODATA;20 NODATA;21 1;22 1;23 NODATA;24 NODATA", "C:/GIS/GFW\_Climate/Brazil\_emis\_comparison/PRODES/PRODES\_primary\_forest\_2010\_early\_raster.tif", "DATA")
  + arcpy.gp.Reclassify\_sa("Prodes2014\_AMZ\_60m\_warp\_low\_res.tif", "Value", "0 NODATA;1 NODATA;2 NODATA;3 NODATA;4 NODATA;5 1;6 NODATA;7 NODATA;8 1;9 NODATA;10 NODATA;11 NODATA;12 NODATA;13 1;14 NODATA;15 1;16 1;17 1;18 NODATA;19 NODATA;20 NODATA;21 1;22 1;23 NODATA;24 NODATA", "C:/GIS/GFW\_Climate/Brazil\_emis\_comparison/PRODES/PRODES\_primary\_forest\_2009\_early\_raster.tif", "DATA")
  + arcpy.gp.Reclassify\_sa("Prodes2014\_AMZ\_60m\_warp\_low\_res.tif", "Value", "0 NODATA;1 NODATA;2 NODATA;3 1;4 NODATA;5 1;6 NODATA;7 NODATA;8 1;9 NODATA;10 NODATA;11 NODATA;12 NODATA;13 1;14 NODATA;15 1;16 1;17 1;18 NODATA;19 NODATA;20 NODATA;21 1;22 1;23 NODATA;24 NODATA", "C:/GIS/GFW\_Climate/Brazil\_emis\_comparison/PRODES/PRODES\_primary\_forest\_2008\_early\_raster.tif", "DATA")
  + arcpy.gp.Reclassify\_sa("Prodes2014\_AMZ\_60m\_warp\_low\_res.tif", "Value", "0 NODATA;1 1;2 NODATA;3 1;4 NODATA;5 1;6 NODATA;7 NODATA;8 1;9 NODATA;10 NODATA;11 NODATA;12 NODATA;13 1;14 NODATA;15 1;16 1;17 1;18 NODATA;19 NODATA;20 NODATA;21 1;22 1;23 NODATA;24 NODATA", "C:/GIS/GFW\_Climate/Brazil\_emis\_comparison/PRODES/PRODES\_primary\_forest\_2007\_early\_raster.tif", "DATA")
  + arcpy.gp.Reclassify\_sa("Prodes2014\_AMZ\_60m\_warp\_low\_res.tif", "Value", "0 NODATA;1 1;2 NODATA;3 1;4 1;5 1;6 NODATA;7 NODATA;8 1;9 NODATA;10 NODATA;11 NODATA;12 NODATA;13 1;14 NODATA;15 1;16 1;17 1;18 NODATA;19 NODATA;20 NODATA;21 1;22 1;23 NODATA;24 NODATA", "C:/GIS/GFW\_Climate/Brazil\_emis\_comparison/PRODES/PRODES\_primary\_forest\_2006\_early\_raster.tif", "DATA")
  + arcpy.gp.Reclassify\_sa("Prodes2014\_AMZ\_60m\_warp\_low\_res.tif", "Value", "0 NODATA;1 1;2 NODATA;3 1;4 1;5 1;6 NODATA;7 NODATA;8 1;9 NODATA;10 NODATA;11 NODATA;12 NODATA;13 1;14 NODATA;15 1;16 1;17 1;18 NODATA;19 NODATA;20 NODATA;21 1;22 1;23 1;24 NODATA", "C:/GIS/GFW\_Climate/Brazil\_emis\_comparison/PRODES/PRODES\_primary\_forest\_2005\_early\_raster.tif", "DATA")
  + arcpy.gp.Reclassify\_sa("Prodes2014\_AMZ\_60m\_warp\_low\_res.tif", "Value", "0 NODATA;1 1;2 NODATA;3 1;4 1;5 1;6 NODATA;7 NODATA;8 1;9 NODATA;10 NODATA;11 NODATA;12 NODATA;13 1;14 NODATA;15 1;16 1;17 1;18 NODATA;19 1;20 NODATA;21 1;22 1;23 1;24 NODATA", "C:/GIS/GFW\_Climate/Brazil\_emis\_comparison/PRODES/PRODES\_primary\_forest\_2004\_early\_raster.tif", "DATA")
  + arcpy.gp.Reclassify\_sa("Prodes2014\_AMZ\_60m\_warp\_low\_res.tif", "Value", "0 NODATA;1 1;2 NODATA;3 1;4 1;5 1;6 NODATA;7 NODATA;8 1;9 NODATA;10 NODATA;11 NODATA;12 NODATA;13 1;14 1;15 1;16 1;17 1;18 NODATA;19 1;20 NODATA;21 1;22 1;23 1;24 NODATA", "C:/GIS/GFW\_Climate/Brazil\_emis\_comparison/PRODES/PRODES\_primary\_forest\_2003\_early\_raster.tif", "DATA")
  + arcpy.gp.Reclassify\_sa("Prodes2014\_AMZ\_60m\_warp\_low\_res.tif", "Value", "0 NODATA;1 1;2 NODATA;3 1;4 1;5 1;6 NODATA;7 NODATA;8 1;9 NODATA;10 NODATA;11 NODATA;12 1;13 1;14 1;15 1;16 1;17 1;18 NODATA;19 1;20 NODATA;21 1;22 1;23 1;24 NODATA", "C:/GIS/GFW\_Climate/Brazil\_emis\_comparison/PRODES/PRODES\_primary\_forest\_2002\_early\_raster.tif", "DATA")
  + arcpy.gp.Reclassify\_sa("Prodes2014\_AMZ\_60m\_warp\_low\_res.tif", "Value", "0 NODATA;1 1;2 NODATA;3 1;4 1;5 1;6 NODATA;7 NODATA;8 1;9 NODATA;10 NODATA;11 1;12 1;13 1;14 1;15 1;16 1;17 1;18 NODATA;19 1;20 NODATA;21 1;22 1;23 1;24 NODATA", "C:/GIS/GFW\_Climate/Brazil\_emis\_comparison/PRODES/PRODES\_primary\_forest\_2001\_early\_raster.tif", "DATA")
  + arcpy.gp.Reclassify\_sa("Prodes2014\_AMZ\_60m\_warp\_low\_res.tif", "Value", "0 NODATA;1 1;2 NODATA;3 1;4 1;5 1;6 NODATA;7 NODATA;8 1;9 NODATA;10 1;11 1;12 1;13 1;14 1;15 1;16 1;17 1;18 NODATA;19 1;20 NODATA;21 1;22 1;23 1;24 NODATA", "C:/GIS/GFW\_Climate/Brazil\_emis\_comparison/PRODES/PRODES\_primary\_forest\_2000\_early\_raster.tif", "DATA")
* Wrote code that applies the four PRODES emissions exclusion criteria that Liz and Mikaela used for each year to annual Hansen loss: 1) within the legal Amazon, 2) within PRODES primary forest for that year, 4) no fire that year or the preceding year, and 4) loss cluster was greater than 6.25 ha. The first was implemented by clipping Hansen loss to the legal Amazon before doing anything else. The latter three were implemented in the analysis script.
  + Ran this with Brazil\_emis\_comparison/Brazil\_GFW\_emissions\_comparison.py on my laptop.
  + The final output for this is a shapefile for each year with the Hansen loss that met these four criteria. This shapefile can then have its emissions calculated for comparison with Brazil’s official emissions statistics.
  + The order in which I applied the exclusions to Hansen loss was somewhat arbitrary, though I started with loss only in the legal Amazon so that all the following statistics were about the legal Amazon.
  + Note on the “no fire” exclusion: Hansen loss pixels were excluded if they had fire that year or the preceding year (except for loss 2001 pixels, which were only excluded if they had loss in 2001 since there is no record for fire areas in 2000). The reason for excluding loss with fires the year before (not just that year) is that the loss could have occurred the year before it was recorded and therefore been the same year as the fire; that is, fire from the year before could have caused the loss observed the following year. I developed this rule upon advice from Mikaela Weiss. When examining the output rasters, I noted that almost none of the raw loss pixels for a given year were on burn pixels from any year besides that year or the preceding year. In other words, among the loss pixels for a given year that were on burn pixels, they were almost always on burn pixels from that year or the preceding year.
  + The script creates an intermediate shapefile for each exclusion step so that emissions can be calculated after each exclusion is applied. This will allow me to see how much each exclusion affected the overall emissions (e.g., did excluding fire areas contribute a lot relative to the >6.25 ha exclusion?)
  + The way I wrote the script, it needs to be run separately for the two PRODES inputs (early: 2001-2014, and late: 2007-2017). To switch between these, I commented out different blocks of code and changed the years of the for loop. It’s ungainly but I didn’t feel like coming up with anything else. This takes a few hours to run for each year, so for the 25 years (between the two PRODES inputs) it’s a lot of run time.
  + QC: Checked the processing of each intermediate tif for loss years 2001, 2007, and 2012. Checked that the outputs for 2007 based on the two PRODES inputs were different, as expected (since they were using different PRODES inputs).
* Ran prep\_for\_tsv\_creation.py on the shapefiles to create a new field that had the file name in it. That will be used for Hadoop.
* Copied all the shapefiles and tifs to s3.
* Created a m4.16xlarge spot machine and ran 1b\_Summary-AOIs-to-TSV/convery-AOI-to-tsv.py from <https://github.com/wri/gfw-annual-loss-processing>. This converts all the shapefiles into tsvs for input into Hadoop without intersecting them with administrative boundaries (GADM) since that’s not relevant for this project. This took just a few minutes.
* Created a large spark Hadoop cluster and ran Hadoop on the tsvs following the instructions on <https://github.com/wri/gfw-annual-loss-processing>: python annual\_update.py --analysis-type loss --points-folder s3://gfw2-data/alerts-tsv/loss\_2017/ --output-folder s3://gfw-files/dgibbs/GFW\_Climate/Brazil\_emis\_comparison/full\_model\_201810/Hadoop\_output\_20190103/raw/ --polygons-folder s3://gfw-files/dgibbs/GFW\_Climate/Brazil\_emis\_comparison/full\_model\_201810/tsvs\_for\_Hadoop\_20190103/ --iterate-by None This took a few hours to run.
* Post-processed the Hadoop output using <https://github.com/wri/gfw-annual-loss-processing/blob/master/2_Cumulate-Results-and-Create-API-Datasets/cumsum_hadoop_output.py>.