Analysing the link between soybean production and deforestation

Firza Riany

04/02/2021

Description

Using the datasets of export-import of soy production, I was asked to identify:
1. The largest soy producing state in Brazil
2. The soy exporting companies from this state that are exposed to deforestation

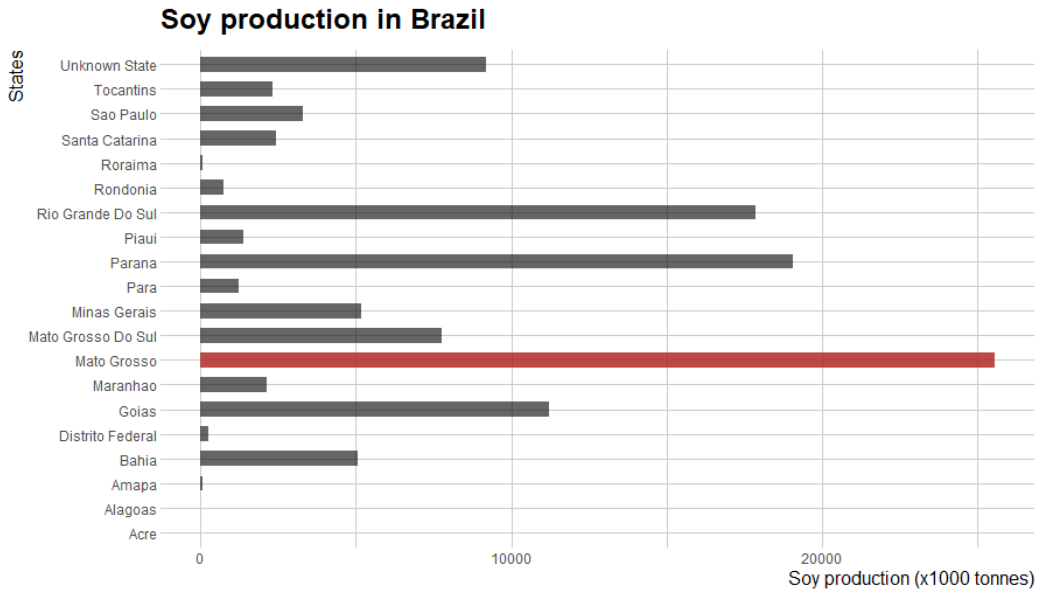
```
library(ggplot2) # for visualization
library(hrbrthemes)  # for ggplot theme
library(RColorBrewer)  # to generate a color palette
library(gghighlight)  # highlighting the bars in a plot
library(stringr)  # string manipulation
library(dplyr)  # subsetting the dataset

soy_df = read.csv("Brazil-soy - Brazil-soy.csv", header = TRUE, na.strings =
" ")
```

The largest soy producing state in Brazil

Procedures:
1. Calculating the aggregate sum of soy production per state
2. Histogram graphic

```
soy_aggr = aggregate(soy_df$SOY_EQUIVALENT_TONNES,
                     by = list(STATE = soy_df$STATE),
                     FUN = sum)

# Sort rows in descending order
soy_aggr_sort = soy_aggr[order(- soy_aggr$x),]

# Scaling it for easy visualization
soy_aggr_sort$scale = soy_aggr_sort$x/1000

# default = options("scipen"=0, "digits"=7)
# deactivate scientific notation = options("scipen" = 100, "digits" = 4)

# Tidy the names of the states
soy_aggr_sort$STATE = str_to_title(soy_aggr_sort$STATE)
```

```r
# Histogram
ggplot(data = soy_aggr_sort, aes(x = STATE, y = scale)) +
  geom_bar(stat = "identity", fill = "#F53434", alpha = 0.6, width = 0.6) +
  gghighlight(scale > 20000, label_key = STATE,
              unhighlighted_params = list(fill = "black")) +
  xlab("States") +
  ylab("Soy production (x1000 tonnes)") +
  labs(title = "Soy production in Brazil") +
  coord_flip() +
  theme_ipsum(base_size = 9,
              axis_title_size = 12,
              base_family = "Calibri")
```



Which companies from Moto Grosso that is exposed to deforestation?

Procedures:
1. I will select the exporting companies present in Moto Grosso
2. With the assumption *each company has the same probability to cause deforestation* I will calculate the deforestation for producing 1 tonne of soy bean
3. I will define the threshold to say whether the companies are posing a high deforestation risk

I found out that Moto Grosso is the state in Brazil with the largest soy production, namely 25 million tonnes of soy. So, firstly I will answer the question how many exporting companies are there in Moto Grosso?

```r
mg_companies = soy_df %>%
  filter(STATE == "MATO GROSSO")
```

```
# Calculating the deforestation caused by the production of 1 tonne of soy
mg_companies$def_per_tonne =
  mg_companies$SOY_DEFORESTATION_5_YEAR_ANNUAL_RISK_HECTARES/mg_companies$SOY
_EQUIVALENT_TONNES
```

I calculated the forest areas that are being cleared to produce 1 tonne of soy (def_per_tonne) because I wanted to compare the companies in the same production scale. To understand which companies that pose a high risk to deforestation, I have to determine a cutoff/threshold value. Companies are considered to pose deforestation risk when their def_per_tonne is larger than the cutoff value.

I determined the cutoff value based on:
1. The deforestation target in 2020 = 400 000 ha
2. Brazilian forest areas in 2020 = 49 662 000 ha 3. The yield of soybean in 2020 = 3.41 tonnes / ha

```
target = 400000
forest_areas = 49662000
soy_yield = 3.41

allowed = target / forest_areas

cutoff = allowed / soy_yield

cutoff

## [1] 0.002362008
```

The logic behind the calculation:
The deforestation target in 2020 is 400 000 ha. So, out of 49 million ha of forest, the allowed forest to be cleared is 400 000 ha / 49 million ha = 0.008 of Brazilian forests. The optimal soy production is 3.41 tonnes / ha, so 0.008 / 3.41 yields the amount of forest cleared to optimally produce 1 tonne of soy. That is then the cutoff value I will use in my calculation.

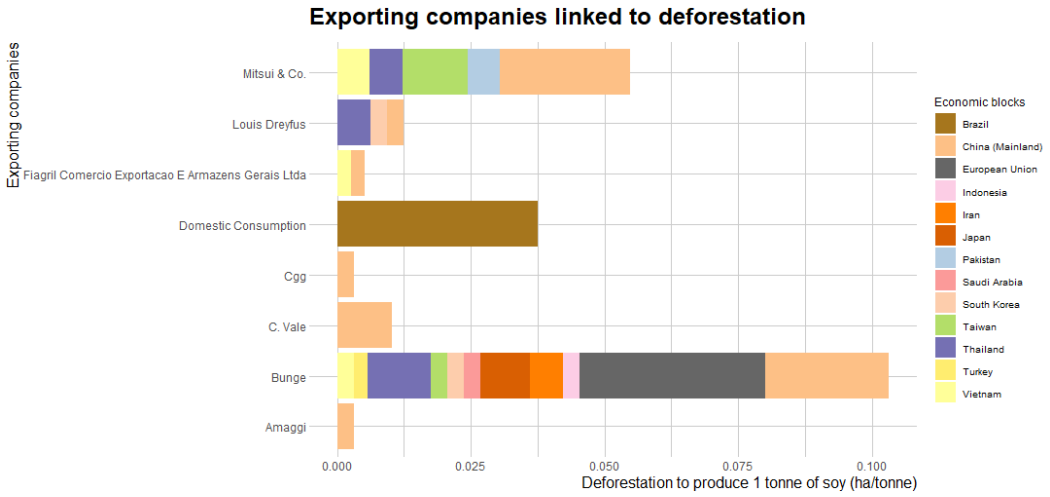
So, I will filter the companies whose def_per_tonne is larger than cutoff (0.002362)

```
mg_deforest = mg_companies %>%
  filter(def_per_tonne > cutoff)
```

There are 68 companies that potentially pose a risk to deforestation from Moto Grosso state. So, now I want to link these companies to the economic block. The aim is to see which blocks that directly support/drive deforestation.

```
# Tidying the strings in the dataset
soy_aggr_sort$STATE = str_to_title(soy_aggr_sort$STATE)

mg_deforest$EXPORTER.GROUP = str_to_title(mg_deforest$EXPORTER.GROUP)

mg_deforest$COUNTRY.OF.IMPORT = str_to_title(mg_deforest$COUNTRY.OF.IMPORT)
```

```
mg_deforest$ECONOMIC.BLOC = str_to_title(mg_deforest$ECONOMIC.BLOC)

# Plotting
## Generating the color palette using RColorBrewer
qual_col_pals = brewer.pal.info[brewer.pal.info$category == 'qual',]
col_vector = unlist(mapply(brewer.pal, qual_col_pals$maxcolors, rownames(qual
_col_pals)))

col_palette = sample(col_vector, 13)

ggplot(data = mg_deforest, aes(fill = ECONOMIC.BLOC, y = def_per_tonne, x = E
XPORTER.GROUP)) +
  geom_bar(position = "stack", stat = "identity") +
  scale_fill_manual(values = col_palette) +
  xlab("Exporting companies") +
  ylab("Deforestation to produce 1 tonne of soy (ha/tonne)") +
  labs(title = "Exporting companies linked to deforestation", fill = "Economi
c blocks") +
  coord_flip() +
  theme_ipsum(base_size = 9, axis_title_size = 12, base_family = "Calibri")
```

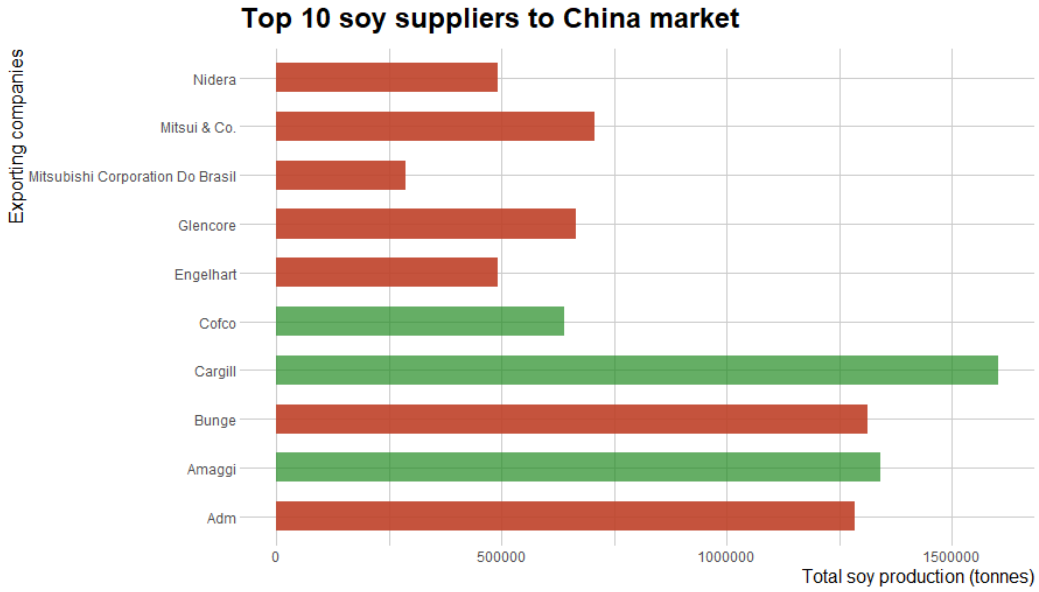


The exporting companies Bunge and Mitsui & Co. are the top 2 companies that cause the highest deforestation to produce 1 tonne of soy. These companies export soy to China that makes China the biggest economic block that support deforestation for soybean plantation in Moto Grosso. Beside China, the European Union also sources its soybean from Bunge which makes the EU the driver for deforestation in Moto Grosso.

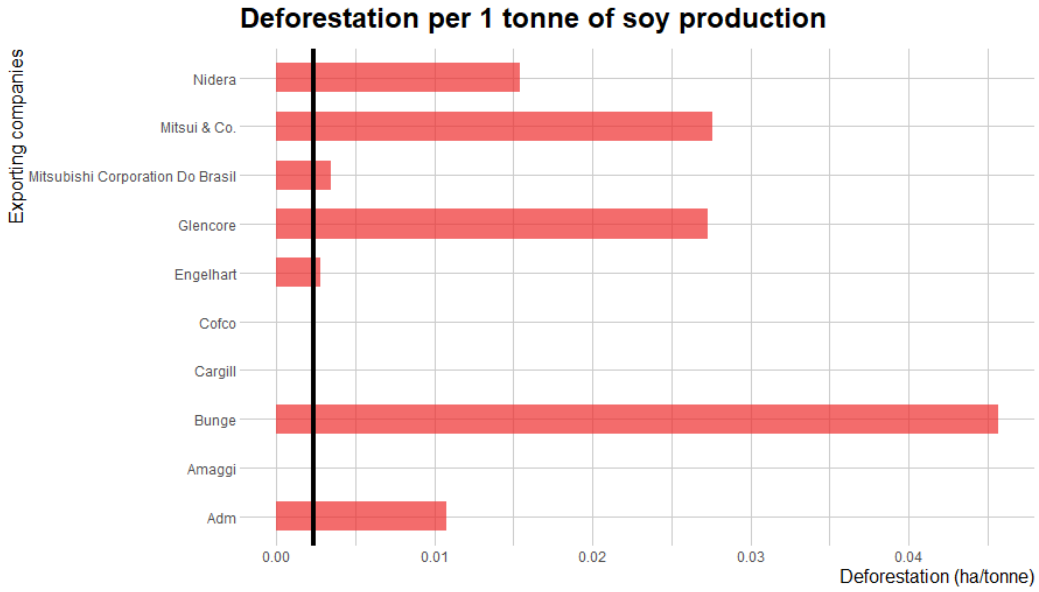
If I have to communicate the result to a sustaianability officer from China, I would also show him/her the Brazilian exporting companies that China is partnering with.

```
# Only considering the companies in Moto Grosso
mg_china = mg_companies %>%
```

```r
  filter(COUNTRY.OF.IMPORT == "CHINA (MAINLAND)")

# Exporting companies that supply the soybean to China
mg_china_supply = aggregate(mg_china$SOY_EQUIVALENT_TONNES,
                            by = list(EXPORTER.GROUP = mg_china$EXPORTER.GROU
P),
                            FUN = sum)

mg_china_supply = mg_china_supply[order(mg_china_supply$x), ]

names(mg_china_supply)[names(mg_china_supply) == "x"] = "total_soy_production
"


# The top 10 suppliers
top_suppliers = tail(mg_china_supply, n = 10)

# The deforestation per 1 tonne soy production, grouped by companies
mg_china_def = aggregate(mg_china$def_per_tonne,
                         by = list(EXPORTER.GROUP = mg_china$EXPORTER.GROUP),
                         FUN = sum)

names(mg_china_def)[names(mg_china_def) == "x"] = "def_per_tonne"

# Merge top suppliers with their deforestation per 1 tonne of soy production
top_suppliers_2 = merge(top_suppliers, mg_china_def, by = "EXPORTER.GROUP")

# Replace NA with 0
top_suppliers_2[is.na(top_suppliers_2)] = 0

top_suppliers_2$EXPORTER.GROUP = str_to_title(top_suppliers_2$EXPORTER.GROUP)

# Plotting the suppliers with no deforestation risks
suppliers = ggplot(data = top_suppliers_2, aes(x = EXPORTER.GROUP,
                                               y = total_soy_production)) +
  geom_bar(stat = "identity", fill = "firebrick2", alpha = 0.7, width = 0.6)
+
  gghighlight(def_per_tonne > cutoff, label_key = EXPORTER.GROUP,
              unhighlighted_params = list(fill = "forestgreen")) +
  xlab("Exporting companies") +
  ylab("Total soy production (tonnes)") +
  labs(title = "Top 10 soy suppliers to China market") +
  theme_ipsum(base_size = 9, axis_title_size = 12, base_family = "Calibri") +
  coord_flip()

suppliers
```




```
deforest = ggplot(data = top_suppliers_2, aes(x = EXPORTER.GROUP,
                                              y = def_per_tonne)) +
  geom_bar(stat = "identity", fill = "firebrick2", alpha = 0.7, width = 0.6)
+
  geom_hline(aes(yintercept = cutoff, color = "Cutoff"), size = 1.5) +
  scale_color_manual(name = "Cutoff", values = "black") +
  xlab("Exporting companies") +
  ylab("Deforestation (ha/tonne)") +
  labs(title = "Deforestation per 1 tonne of soy production") +
  theme_ipsum(base_size = 9, axis_title_size = 12, base_family = "Calibri") +
  theme(legend.position = "none") +
  coord_flip()

deforest
```

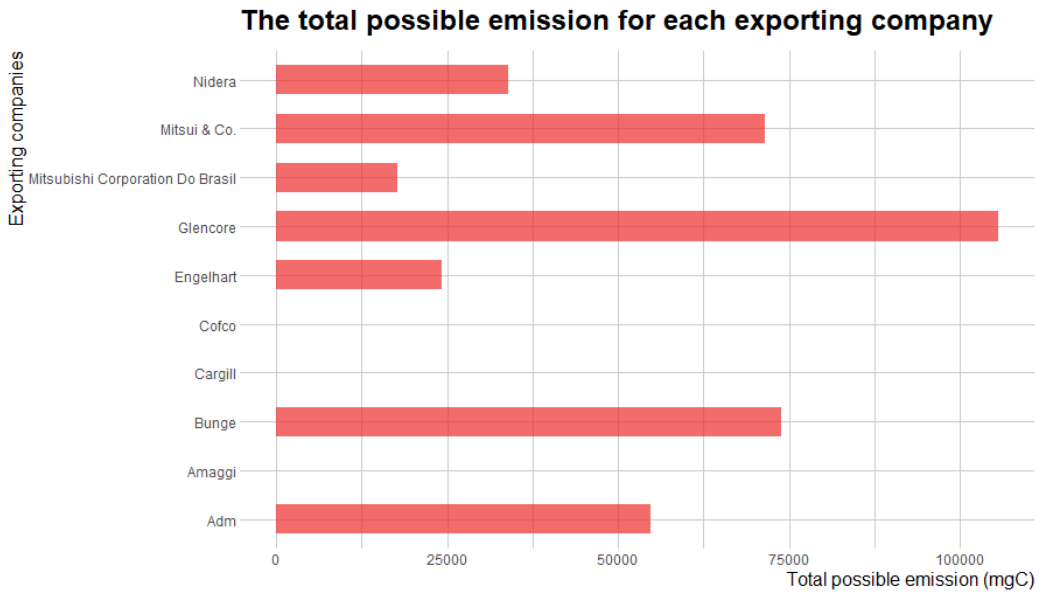


I identified the top 10 soy exporting companies that are partnering with China I also analyzed their deforestation risks per 1 tonne of soy production. With the cutoff value = 0.00020, I identified 7 exporting companies that are potentially causing deforestation. There are only 3 companies that do not cause deforestation. Cargill can be the best choice for a Chinese companies to source Brazilian soybean. This is because Cargill exports up to 1.5 million tonnes without posing deforestation risk.

I have an idea to estimate the possible carbon emission from deforestation given the exporting company's total soy production. I found an equation that can return the equivalent CO2 emission from deforestation using the amount of area changed (ha) and the potential carbon stock in the area (C/ha) (Winrock International 2014). I sourced the carbon stock for Moto Grosso from Nogueira et al. 2009, and it was 159.7 mgC/ha. So I will try to estimate the possible carbon emission caused by these top 10 soy exporting companies.

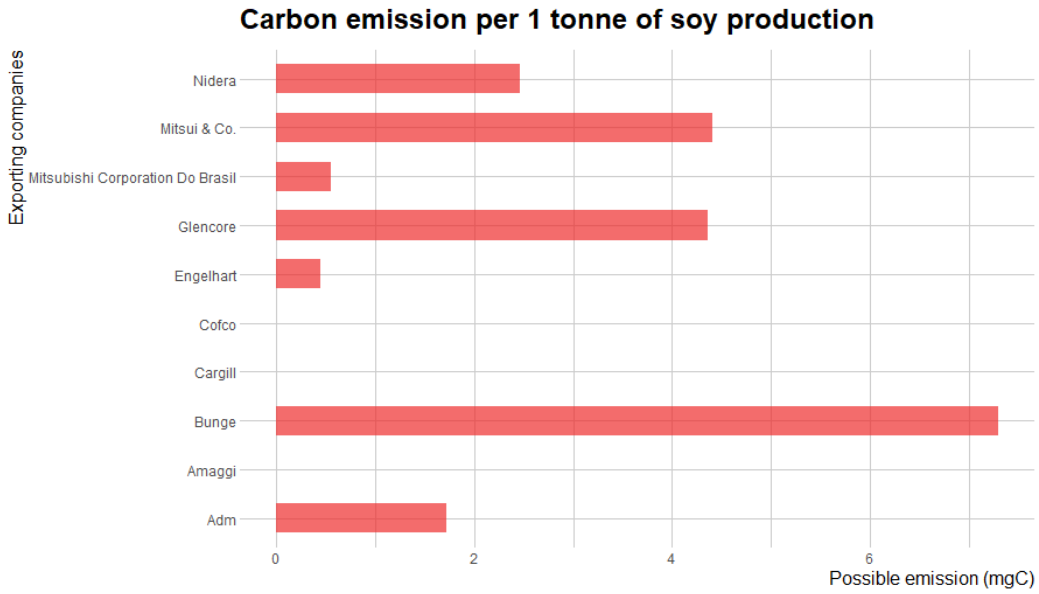
```
c_stock = 159.7

# Calculating the total deforestation risk per company that exports to China
# Assumption: companies that are present in Moto Grosso
mg_china_total = aggregate(mg_china$SOY_DEFORESTATION_5_YEAR_ANNUAL_RISK_HECT
ARES,
                           by = list(EXPORTER.GROUP = mg_china$EXPORTER.GROUP
),
                           FUN = sum)

names(mg_china_total)[names(mg_china_total) == "x"] = "deforestation_risk_ha"

mg_china_total[is.na(mg_china_total)] = 0
```

```r
mg_china_total$EXPORTER.GROUP = str_to_title(mg_china_total$EXPORTER.GROUP)

# Merge top suppliers with their emission
top_suppliers_3 = merge(top_suppliers_2, mg_china_total, by = "EXPORTER.GROUP
")

top_suppliers_3$emission_mgC = top_suppliers_3$deforestation_risk_ha * c_stoc
k

# Plotting
emission = ggplot(data = top_suppliers_3, aes(x = EXPORTER.GROUP,
                                              y = emission_mgC)) +
  geom_bar(stat = "identity", fill = "firebrick2", alpha = 0.7, width = 0.6)
+
  xlab("Exporting companies") +
  ylab("Total possible emission (mgC)") +
  labs(title = "The total possible emission for each exporting company") +
  theme_ipsum(base_size = 9, axis_title_size = 12) +
  coord_flip()

emission
```




```r
# The emission per 1 tonne of soy production
top_suppliers_3$emission_per_ton_mgC =
  top_suppliers_3$def_per_tonne * c_stock

# Plotting
emission_per_ton = ggplot(data = top_suppliers_3, aes(x = EXPORTER.GROUP,
                                                      y = emission_per_ton_mg
```

```
C)) +
  geom_bar(stat = "identity", fill = "firebrick2", alpha = 0.7, width = 0.6)
+
  xlab("Exporting companies") +
  ylab("Possible emission (mgC)") +
  labs(title = "Carbon emission per 1 tonne of soy production") +
  theme_ipsum(base_size = 9, axis_title_size = 12) +
  coord_flip()

emission_per_ton
```



I calculated the total possible emission for each company. In this calculation, I summed the deforestation risk for each company that are present in Moto Grosso. Then I multiply the total deforestation risk with the carbon stock in the area (159.7 mgC/ha). The result indicates that Glencore emits the most. However, it's the total emission and it is difficult to compare Glencore's emission with the other companies because they produce different amount of soy. To make the emission comparable, I wanted to know how many carbon emitted from the production of 1 tonne of soy. It shows that Bunge emits the most. Glencore also emits a fairly great amount of carbon.

Bottomline

I would recommend the Chinese sustainability officer to partner with Cargill.

References

AFOLU Carbon Calculator Series. The Forest Protection Tool: Underlying Data and Methods. Winrock International, 2014.

Anand, Asim. 'Brazil 2020-21 Soybean Output Forecast at All-Time High despite Delayed Start: Conab', 10 November 2020. https://www.spglobal.com/platts/en/market-insights/latest-news/agriculture/111020-brazil-2020-21-soybean-output-forecast-at-all-time-high-despite-delayed-start-conab.

Global Forest Resources Assessment 2020. FAO, 2020. https://doi.org/10.4060/ca9825en.

Morton, Douglas C, Marcio H Sales, Carlos M Souza, and Bronson Griscom. 'Historic Emissions from Deforestation and Forest Degradation in Mato Grosso, Brazil: 1) Source Data Uncertainties'. *Carbon Balance and Management* 6, no. 1 (December 2011): 18. https://doi.org/10.1186/1750-0680-6-18.

Nogueira, Euler Melo, Philip Martin Fearnside, Bruce Walker Nelson, Reinaldo Imbrozio Barbosa, and Edwin Willem Hermanus Keizer. 'Estimates of Forest Biomass in the Brazilian Amazon: New Allometric Equations and Adjustments to Biomass from Wood-Volume Inventories'. *Forest Ecology and Management* 256, no. 11 (November 2008): 1853–67. https://doi.org/10.1016/j.foreco.2008.07.022.

Silva Junior, Celso H. L., Ana C. M. Pessôa, Nathália S. Carvalho, João B. C. Reis, Liana O. Anderson, and Luiz E. O. C. Aragão. 'The Brazilian Amazon Deforestation Rate in 2020 Is the Greatest of the Decade'. *Nature Ecology & Evolution* 5, no. 2 (February 2021): 144–45. https://doi.org/10.1038/s41559-020-01368-x.