**Forecasting the Volume and Effects**

**of Cattle Farming**

Gregory Buckmaster, Gary Chen, Nick Granozio, Katie Pierce

Syracuse University iSchool, IST-718

**Executive Summary**

There is cause for concern and immediate action needed to prevent the necessary increase in agriculture production from causing unnecessary environmental impact if the world is going to safely support continued population growth through 2031.

**Findings:**

* There is correlation between population growth and agriculture production
* There is correlation between agriculture production and GHG Emissions
* There are positive trends in the forecasted livestock production for India
* There are positive trends in the forecasted livestock production for China
* North America leads the Top 3 agriculture producing regions in efficiency
* GHG Emissions reduction possible by best-practice sharing
  + 1.12x10^12 kg (net) = 1.12 billion metric tons
  + 6-8% reduction from Top 3 commodities (Cattle, Pigs, Chickens)

If increasing agriculture production efficiency is not feasible in the regions of focus due to other limiting factors, the next best recommendation is highlighting the importance of local growers and farm-to-table subsistence so that the global supply chain isn’t burdened with transportation and waste associated with cross-continent logistics.

**Specification**

**Problem Statement:**

Feeding the world’s growing population comes at a cost. Not only the costs of natural and manufactured resources, but also via the long-term environmental impacts.

As described below, agriculture contributes approximately one-third of the total greenhouse gas (GHG) emissions released worldwide.

The world faces a significant challenge with respect to managing population growth, preserving natural resources and protecting the environment. Left unchecked, agriculture, as a necessity, could lead to an environmental disaster that creates a compound threat to the world’s population.

This is a complex problem which needs to be supported by advanced data science techniques.

**Hypothesis:**

Regional best-practices in agriculture can be shared to increase efficiency within large populations centers that have correlation to the most significant GHG emissions. With increases in agricultural production having an assumed correlation to population growth, the reasoning is that GHG emissions will increase with a relationship to both.

Null Hypothesis – There is no difference in emissions intensity between regions and sharing best-practices will not lead to more efficient agricultural models. Thus, the world will see an increase in GHG emissions as a measure of agricultural production necessary to meet the needs of a growing population, regardless of controls.

**Data Descriptions:**

From various web sources we learned how agriculture accounts for 25-35% of greenhouse gas emissions. (Appendix 1) These calculations are done in Carbon Dioxide Equivalents (CO2e) and they range from 13.6 to 17.9 billion metric tons, depending on the method used. There are two primary models, one by Poore & Nemecek (2018) and another by Crippa et. al. (2021). The main difference between these two models is how post-retail waste, cooking and deforestation play a role in the output. The model by Crippa has more emphasis on these sources than in Poore & Nemecek.

**Datasets:**

For this analysis, we utilized three separate Kaggle datasets pertaining to livestock production and global emissions. While parts of this analysis used information from individual datasets, the sets were also joined together to gain a larger perspective.

The largest dataset obtained, entitled Food Bank of the World, contains over 100,000 rows of livestock and agriculture values categorized by country and year. The production (as well as the unit used) of each commodity is reported by country within each year between 1961 and 2020. For analysis, this time series data was then trimmed to only focus on livestock.

We then found a dataset called Green House Gas Historical Emissions Data which was also formatted as time series data. Although found on Kaggle, this information was originally published by the World Resources Institute. Similarly, to our first dataset, this data spanned over time (1990-2018) and was categorized by country. Whereas the first dataset gave values for livestock production, this set details the amount of greenhouse gas (GHG) emissions produced.

Finally, the third resource used in this analysis is the GLEAM (Global Livestock Environmental Assessment Model) Livestock Emissions dataset. This information differs from the first two sets as this is not data formatted as a time series, but instead as a snapshot of emissions and protein (in kilograms) produced in 2010. Originally obtained by United Nations Food & Agriculture Organization, this data details the emissions produced by each world region, and further categorized by animal species and production system. Although this dataset provides emissions broken down into specific gasses produced, we focused on the aggregated column “Total GHG Emissions”. Although the lack of time series data made this set harder to relate to the others, the added detail of having emissions linked to units of livestock production (kgs of protein) made this information essential to determining the extent of emissions produced by the livestock industry of each region.

**Scope:**

With respect to time available, once initial observations were made, the scope of this analysis was scaled to focus on the three main commodity types (Cattle, Pigs, Chickens) from the three largest producers / emission sources (China, India, United States).

What questions are we asking -

Q#1) What country / region is producing livestock in the most sustainable way?

Q#2) What type of collaboration between countries / regions can be recommended?

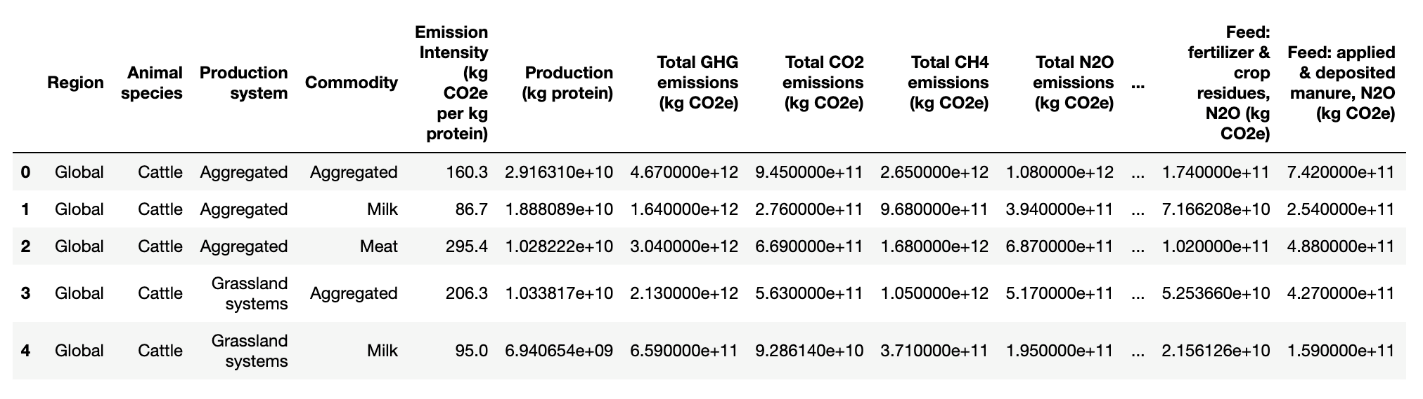
Q#3) What is the potential saved emissions - point in time or forecast?

**Observation**

Around the world, livestock farming has significant environmental impacts due to factors such as greenhouse gas emissions. The effects of this growing concern can be difficult to predict, so this analysis will forecast the emissions created from current rates of livestock farming. The goal of this analysis is to better predict the environmental impacts of this practice and, hopefully, inspire strategies to alleviate them.

**Data Exploration:**

As we explored the GLEAM Livestock Emissions dataset, we noticed that the emissions for various livestock were reported in terms of production system and commodity. For example, emissions from cattle raising were reported for those raised on grassland systems or feedlots and whether the product being cultivated was meat, milk, or eggs (in the case of chickens). Since this level of detail wasn’t available in the other datasets, we decided to only look at the aggregated production system and commodity values in our analysis.



The GLEAM dataset also grouped countries by region whereas the GHG and Food Bank of the World datasets had metrics on individual countries. For certain parts of our analysis, we looked at the countries with the largest populations and attributed the bulk of the emissions in that region to them. For example, in North America 88% of the population resides in the United States, in South Asia 69% of the population is in India, and 85% of the population in East Asia is in China.

GHG Dataset had 193 reporting countries, many of whom did not make a significant contribution to emissions. Therefore, after being transposed, the dataset was trimmed down to the top 12 countries. The final dataframe is below. (Fig.1)

Table

Description automatically generated

Figure 1: GHG Dataframe (Top 12 Countries, by Year)

To understand variation over time across the top 12 countries, a box plot was used for visualization. What this revealed was that China and the United States were observed to be outliers as compared to the other 10 countries. (Fig.2) Overall increase in GHG Emissions can be attributed to a few specific countries and confirms China, India and the United States as appropriate focus in this analysis.

Chart

Description automatically generated

Figure 2: Boxplot, Top 12 GHG Emissions Countries, by Year

This is more clearly illustrated in a standard line chart. (Fig.3)

Chart, line chart

Description automatically generated

Figure 3: Line chart, Top 12 GHG Emissions Countries, by Year

At the scale we are visualizing the top 12 countries, trends outside of the top 3 are hard to detect. Therefore, in relative terms, their trend is stable by comparison. We detect the most significant trend from China, with a sharp increase throughout the early 2000’s. While the United States started out higher than China in the 1990’s, they have remained relatively stable and/or declined slightly in the past 10-15 years. Around the same time as the increase noticed in China there was a similar increasing trend seen in India. However, despite being part of the top 3, India remains the lowest contributor of that set. These trends are all aligned with agriculture production within those countries.

At this point we consider limiting our focus to those top 3 countries for further investigation. Similarly, we see that the common top 3 commodities across those specific regions are Cattle, Pigs and Chickens. While there is some evidence that the emissions generated from raising Buffalo, Sheep and Goats outside of the United States warrant further investigation, for consistency across all regions we focused on these three.

**Analysis**

To begin discussion on our data analysis, we start by displaying the final dataframe for aligning current production values of kilograms protein per commodity against the total GHG emissions by region. (Fig.4) This dataframe has the top 3 regions (countries) and commodities (animal species) aligned with their emissions intensity and production value to result in total GHG emissions per category. Imported back into the timeseries datasets, this creates a way to display changes in production and emissions over time.

Table

Description automatically generated

Figure 4: Final Dataframe, Top 3 Countries & Commodities with Production Values in kg Protein & GHG Emissions Total

Using the GLEAM dataset, we filtered the data down to the regions containing the biggest emissions-producing countries (USA, China, India). Using aggregated values for production system and commodity, we further focused the data on our animals of interest – cattle, pigs, and chickens. We narrowed the scope of our analysis to these animals based on their contributions to overall emissions.

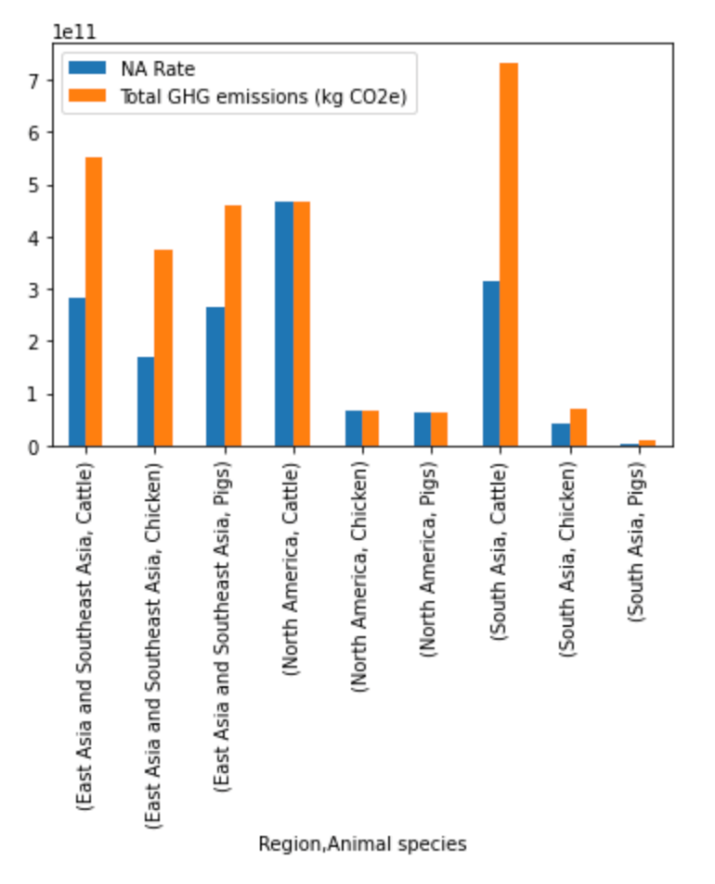
Chart

Description automatically generated

Figure 5: Comparison across regions for production and emissions of specific commodities

Looking at emission intensity, we determined that North America shows the most efficient livestock production (in other words, producing the least amount of GHG emissions per kilogram of livestock protein). This suggests that improvements in global livestock sustainability can be achieved if the other major regions (East Asia and South Asia) were to adopt some of the methods and/or infrastructures of the North American livestock industry. This collaboration between regions would likely result in decreased emissions caused by livestock cultivation and production.

In order to demonstrate this, we created a calculated value of total GHG emissions for East Asia and South Asia if these regions achieved the same emissions intensity for each animal as North America. We then graphed these findings on the bar chart below.

Although this information is taken from 2010, this demonstrates the potential benefit of this cross-collaboration between countries. With China and India being some of the world’s largest livestock producers, this improvement in emissions per kilogram of protein would lead to an immense reduction in GHG emissions. Specifically, this model shows a combined saved GHG emissions value of 665 billion kilograms for East Asia and 451 billion kilograms for South Asia.

Since these are one-time net savings predictions, we now want to look forward with a forecast on how these savings might stack up over time for increased benefit.

Both in India and in China we see where cattle production stands out among the rest in providing substantial impact.

When we look at this forecasted 10 years into the future we further validate that because North America cattle production is on a decline that focus should be placed outside of the US for any potential reduction in emissions.

Below is the forecast for North America cattle production, equitable with GHG emissions from the same process. The trend is clear.

Chart, line chart

Description automatically generated

Figure 6: Example Forecast – North America (US) GHG Emissions, resulting from Cattle Production

When we compare the decline in North America with the upward trend in India and China, the choice is clear where there are regions of concern and actions needed.

Chart, line chart

Description automatically generated

Figure 7: Example Forecast - South Asia (India) GHG Emissions, resulting from Cattle Production

Note: For both India and China that the scale is one order of magnitude larger than that of North America. (1e12 vs 1e11) Therefore, even while seemingly flat in its forecast, China still adds a significant amount of livestock production and GHG emissions into the future.

Chart, line chart

Description automatically generated

Figure 8: Example Forecast - East Asia (China) GHG Emissions, resulting from Cattle Production

Note: See additional forecast for other commodities (animals) in Appendix 3.

Additional forecasts for other commodities across these regions may be found in the appendix, but their emissions are negligible when compared to those of cattle.

**Recommendation**

After conducting our analysis of livestock production and forecasts of emissions, we determined that since production was decreasing in North America, but increasing in East Asia and South Asia, one method for lowering overall emissions would be for countries with high emission intensities to adopt more efficient production and transportation methods. We estimated from the GLEAM dataset that by making improvements to reduce their agricultural emission intensities, these countries could have a net savings of 1.12x10^12 kg of CO2 emissions from these 3 commodities alone. This seemed to be a more reasonable solution than attempting to limit livestock production altogether given the trends in population growth in these regions.

**References**

<https://www.kaggle.com/datasets/selfvivek/environment-impact-of-food-production>

<https://www.kaggle.com/datasets/amandaroseknudsen/gleamlivestockemissions>

<https://www.kaggle.com/datasets/saurabhshahane/green-house-gas-historical-emission-data>

<https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature-projections>

<https://ourworldindata.org/greenhouse-gas-emissions-food#:~:text=The%20specific%20number%20that%20answers,we%20include%20all%20agricultural%20products>.

<https://www.worldwildlife.org/industries/sustainable-agriculture#:~:text=Agriculture%20is%20the%20leading%20source,in%20the%20environment%20for%20generations>.

<https://www.worldometers.info/world-population/eastern-asia-population/>

<https://www.worldometers.info/world-population/southern-asia-population/>

<https://www.worldometers.info/world-population/northern-america-population/>

<https://ourworldindata.org/co2-and-greenhouse-gas-emissions>

**Appendices**

**Appendix 1:** Figure on the calculation of total GHG emissions sourced from agricultural processes and byproducts.

**Chart, bar chart

Description automatically generated**

**Appendix 2: Raw Data Dimensions**

GLEAM livestock emissions

* Kaggle datasource
* 11 rows (global average, livestock category)
* 22 attributes (emissions intensity, production volume)
* Mix of categorial and continuous data
* Static results – Not timeseries

Food Bank of the World

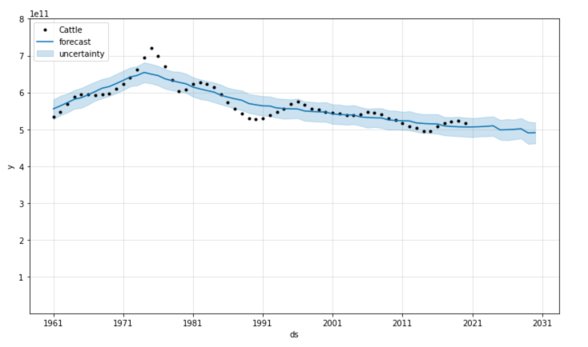
* Kaggle datasource
* >100k rows (country per year over 6 decades)
* 6 attributes (production value)
* Mix of categorial and continuous data
* Timeseries data

GHG Emissions datasource

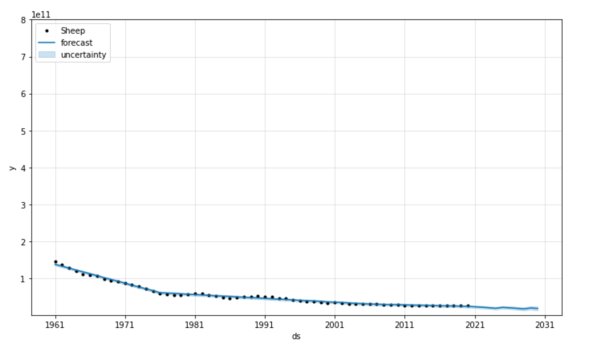
* CO2 emissions by country for 1990 – 2018
* 193 rows (countries)
* 28 attributes (for years 1990 – 2018)
* Not Categorical / All continuous
* Timeseries data

**Appendix 3: Forecasts of Emissions**

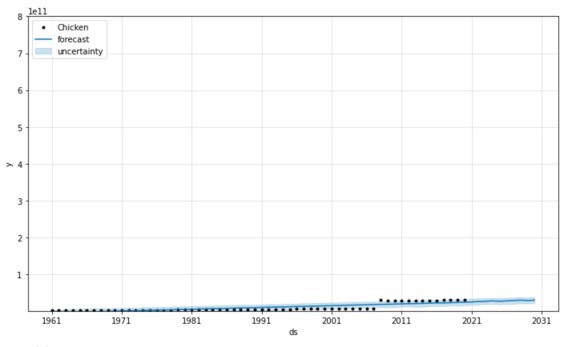
North America (United States):



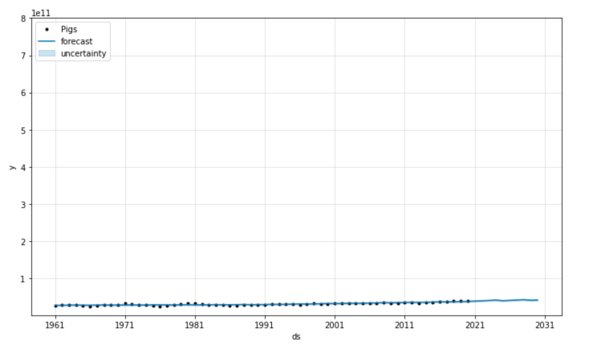
North America GHG Emissions, resulting from Cattle Production



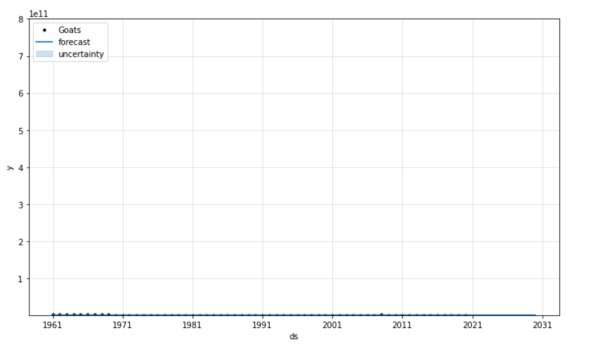
North America GHG Emissions, resulting from Sheep Production



North America GHG Emissions, resulting from Chicken Production

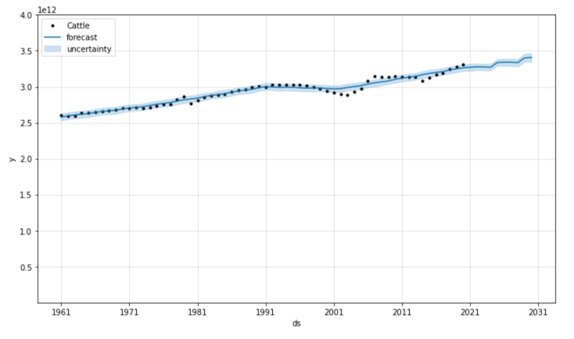


North America GHG Emissions, resulting from Pig Production

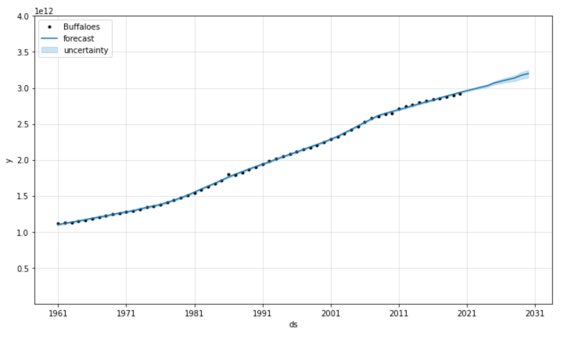


North America GHG Emissions, resulting from Goat Production

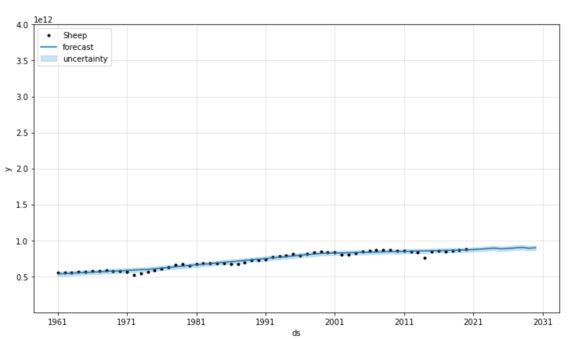
South Asia (India):



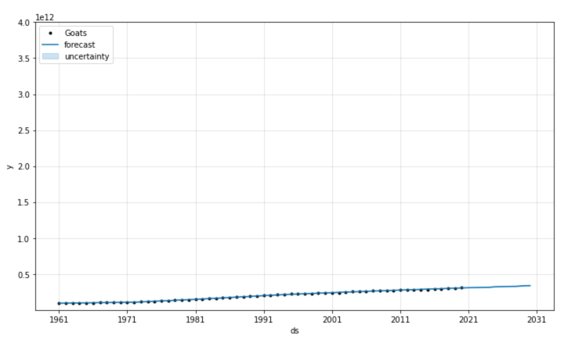
South Asia GHG Emissions, resulting from Cattle Production



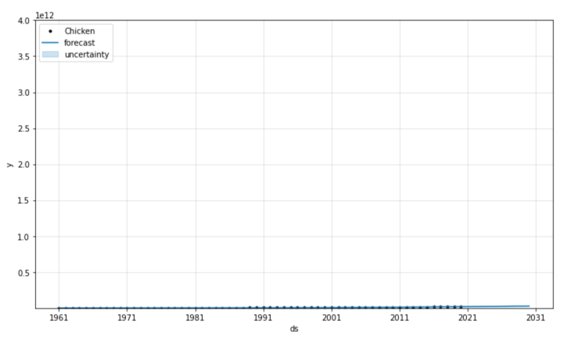
South Asia GHG Emissions, resulting from Buffalo Production



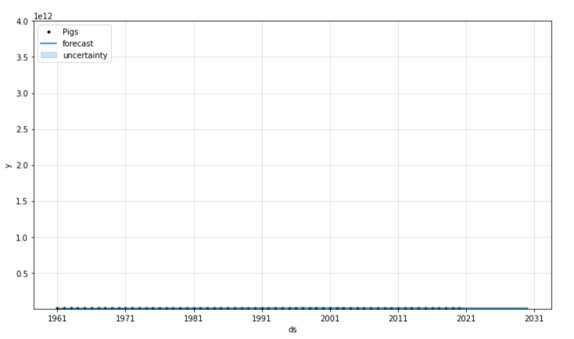
South Asia GHG Emissions, resulting from Sheep Production



South Asia GHG Emissions, resulting from Goat Production



South Asia GHG Emissions, resulting from Chicken Production



South Asia GHG Emissions, resulting from Pig Production

East Asia (China):

Chart, line chart

Description automatically generated

East Asia GHG Emissions, resulting from Cattle Production

Chart

Description automatically generated

East Asia GHG Emissions, resulting from Buffalo Production

Chart

Description automatically generated

East Asia GHG Emissions, resulting from Cattle Production

Graphical user interface, application, table, Excel

Description automatically generated

East Asia GHG Emissions, resulting from Goat Production

Graphical user interface, application, table, Excel

Description automatically generated

East Asia GHG Emissions, resulting from Chicken Production

Graphical user interface

Description automatically generated

East Asia GHG Emissions, resulting from Pig Production