Methods and Data

Estimates of GHG were computed for 42 crops and five species of livestock (Table X) at the district level. A modified CCAFS-MOT model was used with input data primarily from Government of India (GoI)sources. The model was run with input parameters at the tehsil administrative level with results aggregated by to district level to account of the differences in district over the different time periods of the input datasets.

* 1. Data

The input data for the CCAFS-MOT model utilized information from the GoI Input Census 2016-17 and 2011-12, GoI Livestock Census 20 from 2019, Soil Health Cards, Area Production and Yield statistical data (APY), SeedNet, India Meteorological Department (IMD) rainfall grids, NASA FIRMS data, FAO Domestic Animal Diversity Information System (DAD-IS) database, and Spatial data from the Water Resource Information System (WRIS). Cropping data was limited to the 42 crops present in both the APY and Input Census data. Livestock data was limited to the six species reported in the “Breeds xxx “ report (ref).

* + 1. GoI Agriculture Input Census

The GoI Input Census data was downloaded from the website XX during February of 2022 and compiled into a database. The database was used to develop input data for the CCAFS-MOT model for organic and inorganic fertilizer use. Tables 4, 4a, 5E and 5L were used for this purpose. The records in each table were linked to the tehsil boundaries. The tehsils were organized into the districts present in 2016. Table 4 was used as the source of total fertilizer area for the 42 crops and was used to estimate. The data from Table 4a contained information on the types and amount of fertilizer used on 42 crops along with information on the nitrogen percentage within each fertilizer. Table 5E contained information on the amount of farmyard manure applied to each of the 42 crops while 5L contained information on the area utilized for green manure production.

The Input Census data had several data anomalies that needed to be addressed before the data would be input to the model. First the entire data for the State of Punjab was reported under a single district and was proportionally distributed between all districts based on the crop areas within each district. Second fertilizer and FYM amounts where unrealist for small areas and it was determined that units were most likely mis-reported. Data cleaning was carried out to adjust the scale of the reported amount as follows. The median of the fertilizer amount was computed for each fertilizer type and crop. The reported amount was then divided by 1, 10, 100, and 1000 with the resulting amount compared to the median. The value with the least difference was kept as the final amount of fertilizer or FYM used. The records effect by this data cleaning are reported in Table X with the script used available at URL. The resulting cleaning table where then used to develop the input datasets for the model.

* + 1. Area, Production, Yield Statistical Data

The Area, Production, Yield Statistical Data was downloaded from URL in February of 2022 with data extracted for the 2016-17 growing season. The data was linked to tehsil boundaries in a similar fashion as the Input Census data. The data was used as to create the input variable for crop yield and total crop area. The data was also used to fill-in missing crop and crop area data for each district detailed in the Model Input section. The source code used to download and process the data can be found at URL

* + 1. Cost of Cultivation Survey Data

To assess the uncertainty around the rates of nitrogen application the Cost of Cultivation surveys from 2014 to 2000 were downloaded from URL and compiled into a database.

* + 1. SeedNet

The SeedNet data on crop morphology was scrapped from the online resource at URL. Data on rice varieties growing length was extracted along with the India states the variety was release in. This data was used to develop the average growing duration of rice per state. The resulting dataset was used in the construction of the input datasets for the model discussed under the Model Input section. A summary of the rice growing duration per state is summarized in Table X. The github repository for the source code used to extract and process the data can be found at URL.

* + 1. Soil Health Cards

Soil Health Cards data was extracted for the URL. The records with then linked to tehsil boundaries. The individual test results were then aggregated into soil median properties for each tehsil and used to construct soil parameters the input dataset for the model. The Soil Health Cards data was also used to develop a dataset of crop specific fertilizer requirements. for The source repository are the code used to download the data and process it can be found here URL.

* + 1. India Meteorological Department Rainfall Grids

The IMD daily gridded rainfall data were used to develop maps of the proportion of time an area was classified as drought-prone for rice. The IMD daily rainfall grids for 1901 to 2022 where download from IMD website. The daily grids were summed to create annual rainfall grids. The annual rainfall grids where then converted to a bit-masks for drought-prone rice if the annual rainfall was below XX mm of rainfall. The annual bit-masks were then summed and divided by the total of number of annual bit-masks to produce a grid of the proportion of time the cell was drought-prone for rice. The resulting grid was then overlayed with the tehsil boundaries and the cells within each tehsil per averaged to create the proportion of time a tehsils was drought-prone for rice. This data was then used to compute and map rice methane as detailed in the Mapping section below. The source code used for this process can be found in this repository URL.

* + 1. Water Resource Information System

The WRIS contains a number of vector based GIS layers that where used to map the results and develop the spatial components of the input datasets. Tehsil boundaries were downloaded and used as indicated above. The layer “Agro-Ecological Sub-Regions” was used to classify tehsils into the IRRI rice climate typology used in the CCAFS-MOT model.

* + 1. NASA FIRMS

NASA FIRMS data was downloaded from 2012 to 2022 for India. The hotspot point where then field for cropland using the ESRI Landcover 2020 dataset (ref) and linked to the district they occurred in. The average number of annual hotspots was then computed for each district and used to estimate burned area per district as described in section XXX. The source code used to create the average annual burned area per district from the FIRMS data can be found here URL.

* + 1. Livestock Census

Several datasets and reports were used to estimate livestock emissions at the district level. The excel sheet on village level total livestock numbers for cattle, buffalo, sheep, goats and pigs was aggregated to develop district level total number of each livestock breed. The state level tables from the XXX report that contain information on breeds by age and sex was used in combination with the data from the DAD-IS system (detailed below) to estimate state level average livestock species body weight. The scripts and code used to develop the database for the Livestock Census 20 can be found here URL.

* + 1. FAO Domestic Animal Diversity Information System

The DAD-IS database contains information for 8800 breeds of 38 different livestock species. Data on breed body weight for both males and females was extracted and linked to the breeds reported in the Livestock Census report “XXXX” and used to develop the average livestock species body weight per state. Breed that would The code and scripts used for this process can be found here URL.

* + 1. Sdfdsf
    2. sdfsdf