Estimation of Total Nitrogen Inputs

Contents

[Introduction 1](#_Toc164371646)

[Input Survey 1](#_Toc164371647)

[Estimating Total Nitrogen Input 2](#_Toc164371648)

[Cleaning The Input Tables 3](#_Toc164371649)

[Inorganic Inputs 3](#_Toc164371650)

[Organic Inputs 4](#_Toc164371651)

[Organic Input Conversion so Nitrogen 5](#_Toc164371652)

[Estimating Total Nitrogen Rates 10](#_Toc164371653)

# Introduction

Estimating the total nitrogen inputs to Indian crops used the Input Survey data for the cropping season of 2016-2017 (the most recent available). The data used five different tables to compute the total amount of nitrogen added to a hectare within a district. The raw data had approximately 15% of values that where not plausible according to realistic values of nitrogen application when compared to the broad distribution of recommendations in the Soil Health Cards data and crop requirements reported in the SeedNet data. To better represent the total nitrogen inputs to the cropping system, five different methods for cleaning the data are presented.

# Input Survey

The Input Survey is a national survey carried out every five years. The survey is conducted at the district level and collects data on the agricultural crops grown and their inputs. It is described as:

"The main objective of the survey is to generate data on consumption of various agricultural inputs, according to major size-groups of operational holdings, viz.,      marginal (below 1 ha .), small (1- 1.99 ha.), semi-medium (2- 3.99 ha.), medium  (4- 9.99 ha.) and large (10 ha. and above), for getting an insight into the consumption pattern of inputs by various categories of farmers. This information is vital for planning the production, imports, and distribution of fertilizers. The inputs covered in the survey include chemical fertilizers, HYV seeds, pesticides, farmyard manures/compost, bio-fertilizers, agricultural implements and machinery, livestock, and agricultural credit, besides data on input use, including use of certified/notified seeds, high-yielding variety seeds, pest control measurements adopted by cultivators, educational qualification, age and size of households of operational holders are captured through Input Survey."

The sample design, according to the final report, was:

"A two-stage stratified sampling was adopted for the Input Survey 2016-17. Tehsils/CD Blocks constitute the strata, villages within a stratum form the first-stage units and Operational Holdings' in the selected villages as second-stage units. The sample size of first stage units will be 7 per cent of the total number of villages from each stratum. Survey was conducted in all States/UTs. However, in case of Punjab, the data of 13 districts were destroyed in Flood as such the estimate at State level has been generated by applying appropriate statistical technique based on data of 9 districts."

With the sample size and methodology as "Seven percent villages in each State were covered for Input Survey 2016-17. These 7 percent villages were selected randomly out of 20 percent villages already selected for Phase-II of Agriculture Census 2015-16 for purpose of preparation of sampling frame. In selected villages, operational holdings were grouped into the following five size-groups of operational holdings:- Sl.No. Operated area Size-group of holding i Below 1 ha. Marginal ii 1 ha. and above but below 2 ha. Small iii 2 ha. and above but below 4 ha. Semi-medium iv 4 ha. and above but below 10 ha. and medium v 10 ha. and above. Large 2.6.2 Four operational holdings were selected from each of the above-mentioned five size groups of operational holdings. The selection was made separately from each of these size groups following Sample Random Sampling method. If in a particular size group, the total number of operational holdings were less than 4, all the holdings of that size group were covered. The data for Input Survey was collected through field inquiries from these selected operational holders of sampled villages."

More information on the Input Survey of 2016-2017 can be found at [All India Report (da.gov.in)](https://agcensus.da.gov.in/document/is2016/air_is_16-17_210121-final_220221.pdf).

# Estimating Total Nitrogen Input

The estimate for total nitrogen used five tables that included data on inorganic fertilizer (Table 4), farmyard manure (Table 5E), oil cake (Table 5F), other manure (Table 5G), and green manure (Table 5LA). Nitrogen content was directly reported for inorganic fertilizer, while for farmyard manure, oil cake and green manure, nitrogen values were assembled from the literature. Monte Carlo simulation was used to build distributions of the nitrogen content for farmyard manure, oil cake, other manure and green manure. In the Input Survey only the area where green manure is grown is reported and not the quantity produced as in the case of the other fertilizer. Monte Carlo simulation was used to build a distribution of green manure nitrogen additions using values from the literature.

## Cleaning The Input Tables

## Inorganic Inputs

The input table contained invalid data up to approximately 15% of the records. Five candidate methods were developed to correct the input data. All five methods assumed the unrealistically high values of nitrogen per hectare were caused by their values being reported in units other than the stated tonnes. The five methods for adjusting the very high values are all based on the same basic approach. First, the median nitrogen per hectare for each crop was computed. Next, each record's value for nitrogen per hectare was divided by 1 (no change), 10 (data reported in quintiles), 100, and 1000 (data reported in kilograms). For each data point in the input data, the absolute difference between the value with changed units and the median for the crop was computed. The correction with the minimum difference was assumed to be the correct unit conversion for the input data. Five candidate methods for correcting the data are:

1) The base method is referred to as the **corrected** method on the graphs in Appendix A.

2) The second method assumed the data was reported in either tonnes, quintiles, or kilograms. This eliminates the possibility of dividing by 100 and assumes that the mistake in units would occur for either quintiles or kilograms which are common units using in Government of India surveys and reports. The same method was used to find the unit conversion for the input data as in the corrected method. The second method is referred to as **unit corrected** on the graphs in appendix A.

3)The third method only corrected values that were potentially mis-reported in kilograms using the base method. The third method is referred to as **kg corrected** on the graphs in appendix A.

The last two methods use the Soil Health Card (SHC) recommendation data to qualify if the values reported needed unit correction. The SHC recommendations vary based on crop, variety, soil type, irrigation, and crop duration.

4) The fourth method uses the maximum recommended value from the SHC and only applies unit conversion to input values above it. It is referred to as **unit corrected > soil health cards maximum** on the graphs in Appendix A**.**

5) The fifth method used the 95th percentile of the recommended nitrogen amounts in the SHC and then only applied unit conversions to the input values that are greater. It is referred to as **unit corrected > soil health cards 95th percentile** on the graphs in Appendix A.

Appendix A contains two graphs for each crop (44). The first graph zooms the y-axis to better display the bulk of the data, while the second graph shows the full range of input values on a log10 scale. Each graph contains nine distribution plots, either a violin (zoomed plot) or a boxplot (full range). The first of the nine distribution plots is the raw nitrogen per hectare from the Input Survey, and the next five distribution plots show the corrected data distributions. The values at the bottom of the five distribution plots that represent the corrected data contain both the number and percent of records affected.

The last three distribution plots are for reference to compare the data corrected using the above methtods. The first is the data from the Cost of Cultivation (CoC) survey for the years 2000 to 2022. The CoC data represents individual farm field-level application of nitrogen compared to the Input Survey, which represents district mean nitrogen application. The following distribution plot shows the SHC recommendations for the crop. The last distribution plot is the required nitrogen distribution reported for different varieties in the SeedNet database. Not all crops contain nutrient requirements in the SeedNet database, so it may not be present. Both the SHC and SeedNet distributions can be used to determine which of the correction methods might be most realistic. Given the number of plots the Table of Contents can be used to navigate to specific crops quickly.

## Organic Inputs

Organic inputs include farmyard manure (FYM), green manure, other manures, and oil cake. The data of organic inputs come from four tables in the Input Survey:

1. Table 5E: Usage Of FYM/Compost for Different Crops
2. Table 5F: Usage Of Oil Cakes for Different Crops
3. Table 5G: Usage Of Other Organic Manures for Different Crops
4. Table 5LA: Usage Of Green Manure for Different Crops

The table contains crop-specific areas of application and quantities applied in tones, except for Table 5LA, which only reports the area where green manure is grown.

The input data Tables 5E, 5F, and 5G were subjected to the same methods for adjusting invalid values as for inorganic fertilizer, with a few differences. The **unit corrected > soil health cards maximum** and **unit corrected > soil health cards 95th percentile** methods used recommendations for FYM reported on individual farm Soil Health Cards and were used to correct data only for Tables 5E and 5G. Crop-specific data was limited for some crops. Two additional methods were developed for the manure data (Tables 5E and 5G) that aggregated the SHC FYM recommendations across all crops. Data from the tables greater than the maximum or 95 percentile of all crop values from the SHC data was then unit adjusted and referred to as **unit corrected greater than soil health card maximum all crops** and **unit corrected greater than soil health card 95 percentile all crops**, respectively, on the graphs in Appendix B and D.

There was no SHC data for recommendations for the use of oil cake, resulting in only the first three cleaning methods used for data from Table 5F.

No adjustments were made to the green manure data in Table 5LA, as only the area was reported.

Crop-specific graphs showing the resulting distribution for the different methods used to adjust invalid data for organic inputs are presented in Appendix B for FYM, Appendix C for oil cake, and Appendix D for other manures. There are two graphs per crop: a violin plot with the y-axis zoomed to show better the distribution of the bulk of the data and a boxplot with a Log10 y-axis that is full scale to show the range of the outliers. For FYM and other manure data, the cost of cultivation manure distribution is shown for comparison. Similarly, the SHC recommendation for FYM application is plotted for comparison purposes.

## Organic Input Conversion so Nitrogen

The organic input quantities per hectare were converted to nitrogen using values from published sources. Distributions for possible values were constructed and used to sample from to conduct Monte Carlo (MC) simulations to estimate the uncertainty of the nitrogen inputs. The MC method varied depending on the organic input.

|  |
| --- |
| Table 1: Data from Arunachalam et al., 2015 used to convert FYM amounts to nitrogen application rates. |
| A table with numbers and symbols  Description automatically generated |

To estimate nitrogen inputs for FYM, data from Arunachalam et al., 2015 was used to build a sampling distribution for MC simulation (Table 1). The data included nitrogen content (%) and standard deviation from four different composting methods, measured every 15 days starting at 0 up to 120. The MC simulation procedure first sampled a random day between 0 and 120 using a continuous uniform distribution. This was followed by estimating the nitrogen content and standard deviation using linear interpolation of the data presented in Table 1. Two methods were then employed to aggregate across the different composting methods. The first method used simple averaging, while the second weighted the improved, vermi-compost, and enriched method at 5% and the conventional method at 85%. For each method, the aggregated standard deviation was also computed. Samples of nitrogen content (%) were then drawn from normal distribution parameterized with the aggregate means and standard deviations for the two methods, Figure 1. The two methods equally weight the days of composting, representing the farmer's practice of piling the FYM over the year and then applying it during field preparation, resulting in an FYM composting pile made up of a range of FYM ages. The resulting density distribution of FYM nitrogen content produced by the two MC simulation methods can be seen in Figure 1. The resulting nitrogen content estimates were then multiplied by the corrected FYM data to estimate the nitrogen application rates.

|  |
| --- |
| A graph showing different colored lines  Description automatically generated |
| Figure 1: Monte Carlo Sampling to estimate nitrogen content of FYM. |

|  |
| --- |
| A graph of a graph showing a number of different colored lines  Description automatically generated with medium confidence |
| Figure 2: Density plot of the two MC methods used to estimate FYM nitrogen content. |

**NOTE: These methods are not finalized**

|  |
| --- |
| Table 2: Data used to convert oil cake amounts to nitrogen application rates. |
| A table with text and numbers  Description automatically generated |

To convert oil cake to nitrogen data from Rana(2011) was used that contain nitrogen content for the common oil cakes available in India (Table 2). A gamma distribution was fit to the data. MC simulations draw samples from the gamma distribution to estimate the uncertainty of oil cake nitrogen content, which was then multiplied by the amount applied to estimate oil cake nitrogen application rates.

|  |
| --- |
| A chart of different types of meat  Description automatically generated |
| Figure 3: Data used to convert Other Manure to nitrogen application rates. |

Other manure used data from The University of Minnesota (UoM) extension service to develop an MC sampling distribution. The UoM data represented large sample sizes of manure nitrogen content for different livestock species (Figure 3). The Animal Census, 19 of India, was used to weigh the different animal species' manure nitrogen content to develop a district-specific sampling distribution that draws samples from a normal distribution. The resulting estimates of manure nitrogen content were then multiplied by the corrected other manure amounts from Table 5G to estimate the nitrogen application rate.

Estimating nitrogen additions from green manure used data presented in Das et al. 2020 (Table 3). The data represents the common green manure crops used in India and the kilogram of nitrogen added per hectare. A gamma distribution was estimated from this data and used to draw samples during MC simulations. The MC estimates of green manure nitrogen additions per hectare are then used to estimate total nitrogen application rates.

|  |
| --- |
| Table 3: Data used to convert green manure into nitrogen addition. |

### Estimating Total Nitrogen Rates

Non-linear models used to estimate N2O emissions are sensitive to the nitrogen application rate. The Input Survey does not contain any information about the combination of the different nitrogen inputs applied to a given area. MC simulations were used to estimate the possible combinations of the five different nitrogen inputs. The process used was as follows:

1. Randomly select one of the nitrogen inputs to start with
2. Randomly select one of the remaining nitrogen inputs.
3. Compute the required overlap between the two inputs based on the total area where the crop was grown.
4. Randomly select a number from 0 to 1 from a uniform distribution that represents the proportion of the small area input that overlaps with the other input.
5. Randomly select one of the remaining inputs.
6. Compute the required overlap with the current area covered by one or more inputs.
7. Compute the area for the distinct set of overlapping areas.
8. For any area less than the required overlap with the currently selected input
9. Randomly order the distinct set of areas currently allocated and loop through them to estimate how much of the required overlap for the current input is allocated to the area.
10. Repeat step 9 for the remaining area of the current input to determine the unrequired overlap proportions.
11. Repeat 5 through 10 for all the remaining unallocated inputs.
12. Once all input areas have been allocated, compute their area and the total nitrogen rate applied by summing the nitrogen rates of the individual inputs that overlap within the area.
13. Repeat the process 1000 times to build the distribution of possible total nitrogen rates.

The resulting dataset of possible total nitrogen rates can then be used to model N2O emissions per hectare weighted by the area they represent. The above method can yield high nitrogen rates when multiple inputs overlap, but this is compensated for by the smaller total areas that have nitrogen inputs.