Increasing accessibility and usability of open-source data through a web map for better decision-making in India's cold chain of fresh produce

Kanaha Shoji[a], Roberta Evangelista[b], Joaquin Gajardo[a], Thomas Motmans[b], Thijs Defraeye[a]

[a] *Empa, Swiss Federal Laboratories for Materials Science and Technology, Laboratory for Biomimetic Membranes and Textiles, Lerchenfeldstrasse 5, CH-9014 St. Gallen, Switzerland*

[b] *BASE, Basel Agency for Sustainable Energy, Elisabethenstrasse 22, CH-4051 Basel, Switzerland*

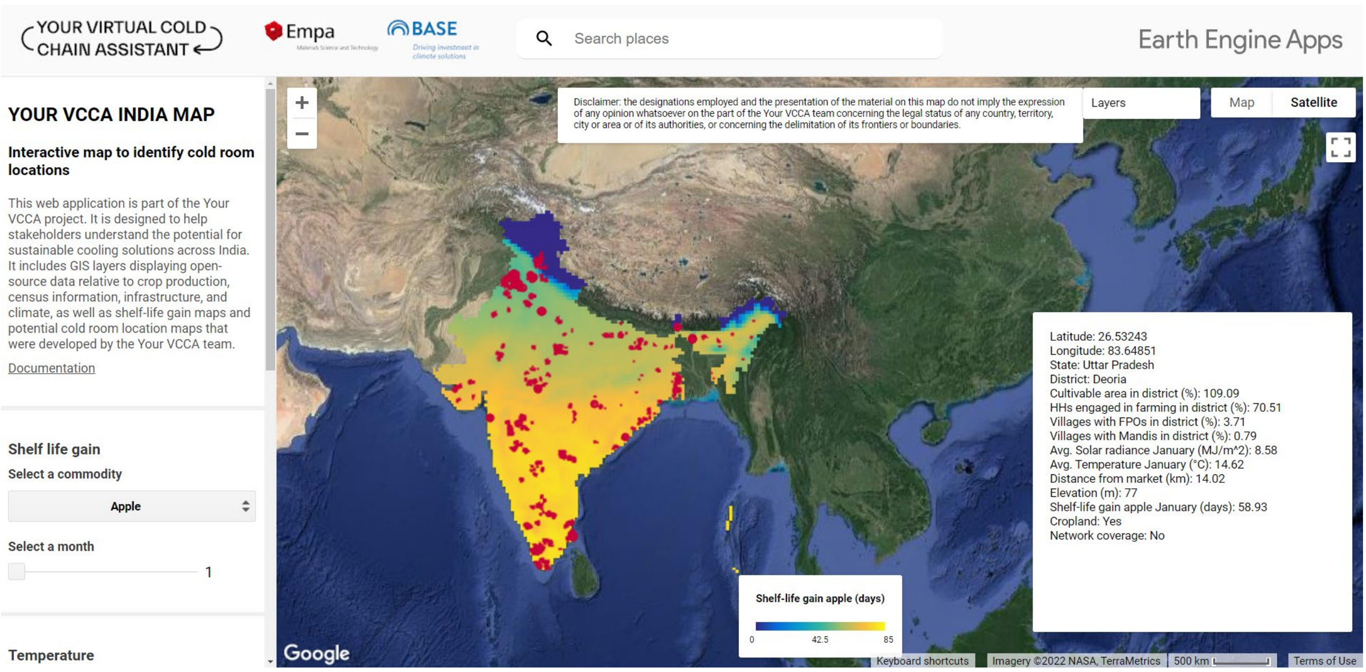
Keywords: data visualization, Google Earth Engine, knowledge exchange, data-driven decision-making, data literacy gap, smallholders, postharvest storage, social innovation, Geographical Information System, site location, cooling rooms, web application

Abstract

Access to actionable data on fresh food supply chains bring key advantages for various stakeholders from the public and private sectors for better decision-making. For India, many agriculture-related data are publicly available as open-source data. However, many available data are not easy to access or visualize and may require a certain degree of processing before they can be turned into interpretable information for decision-making. For that reason, these data are often not used to their full potential. Here we present an interactive web map[1] that collates and visualizes relevant open-source data for fresh food supply chains in India. The available data layers include cropland, elevation, electrical grid, crop production, cold storage locations, air temperature, solar radiation, road network, market locations, mobile coverage, and census. This visual representation of the selected data layers serves as a decision-making tool for stakeholders in the fresh produce supply chain, including social enterprises like farmer producer organizations, and policymakers, as well as financial institutions, cooling providers, NGOs, and government bodies. A practical example of this is identifying where cooling could provide the highest shelf life gains and where to site future cold storage facilities. We anticipate that this web map can lead to finding relevant insights that could help stakeholders optimize their supply chain of fresh produce. This web map could be a starting point for understanding correlations between various parameters and for narrowing the data literacy gap in agricultural sectors in India.

[1]The web map and this document were prepared as part of Your Virtual Cold Chain Assistant (YourVCCA) project. YourVCCA team: maps@yourvcca.org

Link to the web map: https://yourvcca-maps.users.earthengine.app/view/yourvcca-map-india
Link to the source code of the web map: https://code.earthengine.google.com/7acfad8c67f08e070acfbc4228c0709b



Screenshot of the web map

Table of contents

1 Introduction

Fresh food supply chains are complex and sensitive systems depending on a variety of factors. For example, access to roads and the electrical grid is often critical because of the need for a power supply for cooling and rapid transportation to avoid spoilage. Access to this information is key to successful decision-making by supply chain stakeholders, such as cold storage providers, farmers, retailers, or policymakers. Fortunately, there is a growing trend for open-source data sharing, and a considerable amount of relevant data for food supply chains is already publicly available [1].

In India, data from rural areas are especially essential as approximately 2/3 of the population live in the rural areas [2]. Agriculture provides employment to around 70% of India's rural population [3]. Many of these people are smallholder farmers who produce a significant share of the country’s food. Thus, agriculture-related data in rural areas play key roles in decision-making. There have been several initiatives to combat the lack of data in rural areas in India [4]–[6]. Mission Antyodaya, for instance, is a governmental initiative, following up on the census in 2011 in India [7], [8]. Mission Antyodaya was launched in 2017 and covers various schemes at the Gram Panchayat level, which is a basic village governing institute in India [9], [10]. The annual survey data are captured digitally and available on their website [9]. Remote sensing data can also provide relevant agriculture-related information for food supply chain stakeholders in rural areas [11], [12]. For example, global elevation [13], climate [14], and cropland data layers generated with satellite observations are publicly available.

However, many of these open-source data are challenging to find. Furthermore, these data need considerable processing and analysis to be effectively visualized and used for decision-making. For example, for cooling service providers to decide where to place their future cooling storage, they need to understand where the smallholders, who need cooling storage, are located at different scales of spatial granularity. They may also need to consider environmental information such as solar radiation levels, ambient temperature, and topography in order to evaluate how viable a site is for sustainable cold storage. In addition, the availability of infrastructure, such as mobile broadband coverage, in addition to the distance to roads, electrical grid, and markets, could be decisive factors for choosing prospective sites. All these factors, and more, need to be taken into account to make a decision. The parameters mentioned above are currently available open-source. Yet, grouping all these data and extracting actionable information from them requires processing, analysis, visualization, and tailoring to the stakeholders of the fresh produce supply chain. Small and medium-sized enterprises, who play a key role in giving smallholder farmers better marketability, often do not have the capacity to collect all these data or analyze it, despite these data being available.

In order to make these open-source data readily available and useful for stakeholders, we created an open-access interactive web map as a Google Earth Engine application[1]. The map contains different data layers related to the supply chain of fresh produce in India and allows for rapid visualization of relevant spatially distributed data.

This web map is part of the Your Virtual Cold Chain Assistant (Your VCCA) project in India[2]. Through this project, BASE (Basel Agency for Sustainable Energy) and Empa (Swiss Federal Laboratories for Materials Science and Technology) are aiming to reduce food loss and increase smallholder farmers’ incomes by providing smallholders with access to cold storage through a pay-per-use servitization model as well as equipping them with postharvest expertise and market intelligence via a data science-driven mobile application. The project is funded by the data.org Inclusive Growth and Recovery Challenge, which was launched in partnership with the Rockefeller Foundation and the Mastercard Center for Inclusive Growth.

[2] Link to Your Virtual Cold Chain Assistant web page: https://yourvcca.org/india/

To meet the project's purpose of improving food security, increasing smallholders' income, and reducing food loss while mitigating the impact on the environment, we designed our web map to help stakeholders understand the potential for sustainable cooling solutions across India. Our web map includes open-source GIS data layers, including cropland mask, elevation, predicted electricity network, crop production, cold storage locations, the climate, available road infrastructure, agricultural survey information, market locations, and mobile broadband coverage. We also calculated shelf life gain for specific crops and potential cold room locations that we selected based on our set thresholds for selected layers. In addition, we added a map of predicted cropland locations for three districts in Himachal Pradesh in 2020, as illustrated in [15].

The primary goal of this document is to provide details on what layers are shown in our web map. Together with this document, we aim that our web map delivers a sense of the capability and potential of visualizing open-source data. Below, we describe the data layers included on our web map and the details of how these data were obtained.

Disclaimer

The layers shown on the map were cropped for India, using the open-source shapefile of the country boundary of India (see Section 2). Thus, what is presented in our web map is 'as-is' with no express or implied political interest. We did not take district-level reorganizations beyond 2015 into account, as the latest open-source district boundaries we found were from the year 2015 (see Section 2). While we have taken utmost care in selecting and collecting the dataset from reliable sources, we cannot guarantee the validity and accuracy of the information. We include the license of the data shown on each layer as of January 2022. Yet, we recommend the users to check by themselves for further usage. Please refer to the original source for additional information on data collection and accuracy. Any comments or suggestions about this application should be addressed at maps@yourvcca.org.

2 Preprocessing steps

The open-source data we found had various spatial coverage and data formats. The following preprocessing steps were conducted to take these differences into account. Such uniformization made it possible to show them on the web map.

2.1 Global data preprocessing

The cropland mask (Section 3.1), elevation data (Section 3.2), predicted electricity network lines (Section 3.3), climate data (Section 3.5), market locations (Section 3.8), and mobile broadband coverage data (Section 3.9) contained global data. These layers were clipped to India using a shapefile with state boundaries from 2015 [15] (Fig. 1a). Note that we used these boundaries as they were the latest boundaries we could find (see Disclaimer).

Different software was used for this processing step. The elevation data was sourced from the Google Earth Engine data catalog [16] and was directly clipped for India in there. The predicted electricity network lines, market locations data, and mobile broadband coverage data were processed in QGIS (version 3.18.2) [18]. The climate data were processed in R (version 4.1.0) [19].

2.2 Tabular data preprocessing

Most of the collected open-source data consist of tabular data presenting metrics relative to a given district (or block, for some states). This is the case for the agricultural data sourced from the ICRISAT (Section 3.4), harvesting season data (Section 3.5) and the agricultural survey data sourced from the Mission Antyodaya (Section 3.7). For visualization purposes, the harvesting season data were merged with the state-level boundaries [16] (Fig. 1a). The ICRISAT data were merged with the administrative layer at the district level (as of 2015) [20] (Fig. 1b). To present the Mission Antyodaya on a map, we merged the information with the district boundaries [20] (Fig. 1b). Additionally, we joined the Mission Antyodaya information with the block boundaries for the states of Odisha [21] (Fig. 2a) and Bihar [22] (Fig. 2b) (as of 2021 and 2019, respectively). Ideally, all information from the Mission Antyodaya could be displayed at the block (and even Gram Panchayat [GP]) level. However, the limiting factor is the lack of open-source blocks and GPs boundaries for most Indian states.

The merging process was done in R (version 4.1.0) [19] for the harvesting season data. For these data, some of the states had old names[3]. These names were changed to the corresponding current names manually in order to conduct this merging process successfully. Additionally, when a state in a certain month had two different harvesting seasons indicated, the season with the lower yield was used for visualization. For example, the harvesting season of apples in Arunachal Pradesh in September is shown as 'Lean season' in the map, while the original image indicates a half Peak season and a half Lean season. For other tabular data, the merging was performed in Python (version 3.7.7) using the 'fuzzywuzzy' package [23] to match states, districts, and block names that might be spelled differently in the different data sources. Note that we are presenting the Ladakh region as a separate territory from Jammu and Kashmir to reflect the Jammu and Kashmir Reorganisation Act, 2019.

License of [20]: freely available for academic use and other non-commercial use

License of [21]: See https://github.com/justinelliotmeyers/Odisha_2021_Official_Boundaries/blob/main/Odisha_metadata_03232021.pdf (data sourced from the Odisha State and District Level Geospatial Web Portal for Planning and Development, or SAMPAD, is in public domain)

License of [22]: See https://github.com/justinelliotmeyers/official_bihar_india_village_boundary_shapefile/blob/master/source_and_terms_of_use.txt (data sourced from http://gis.bih.nic.in/DEPTS/Default.aspx)

[3] The states names of Orissa, Uttranchal, and Pondicherry, were replaced by Odisha, Uttarakhand, and Puducherry, respectively. The harvest season of Andhra Pradesh was applied to Telangana, as the original data did not contain Telangana. These processes were done solely for visualization purposes.

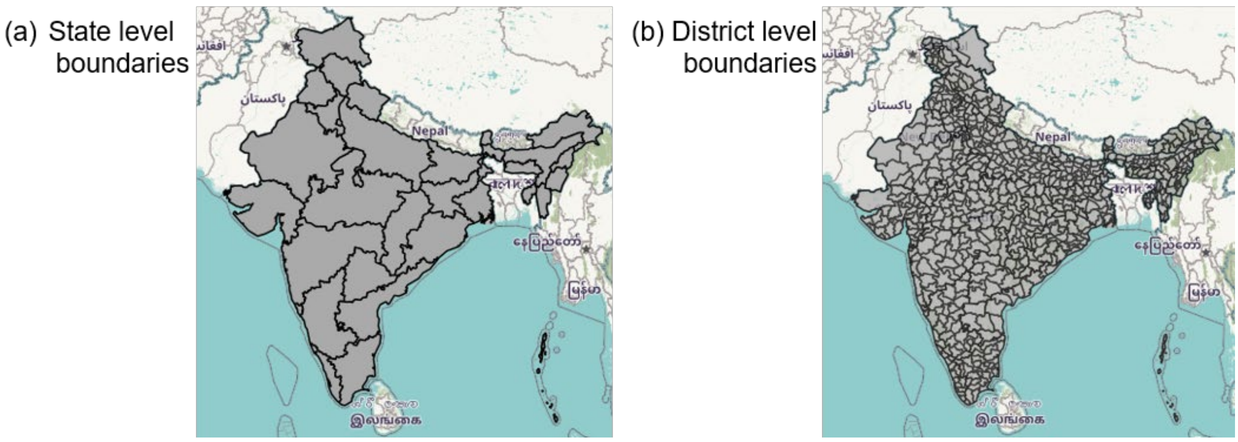


Figure 1: Open-sourced administrative boundary of 2015 for (a) state level [16], (b) district level [20].

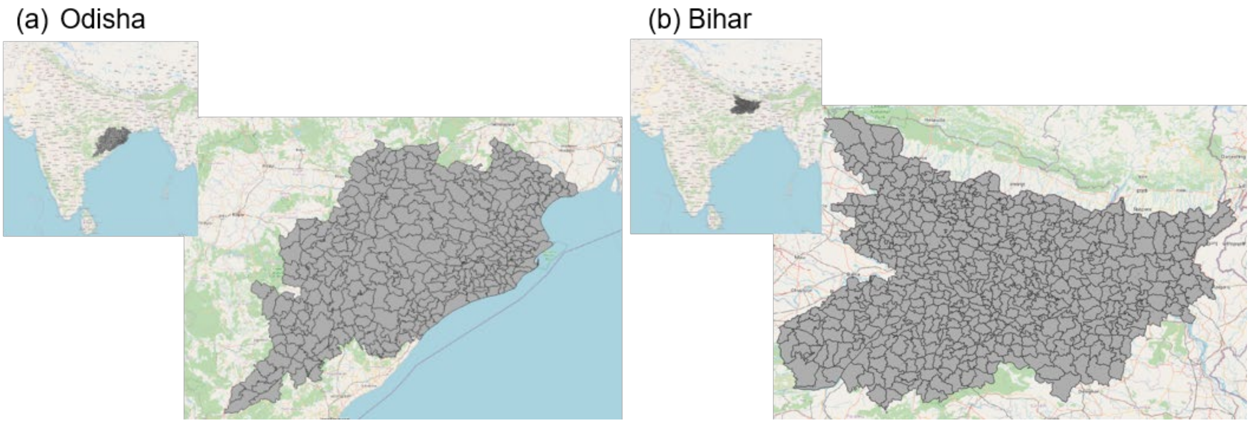


Figure 2: Open-sourced block boundary for the state of (a) Odisha in 2021 [21] and (b) Bihar in 2019 [22].

3 GIS Layers

In this section, we explain the main features of what is shown in each data layer. As the preprocessing steps of the data are discussed in the previous section (Section 2), the details of the preprocessing are omitted below. Unless otherwise stated, the layers are 'as-is' from the original source.

3.1 Cropland mask

This map presents the cropland vs. non-cropland and water extent as calculated in the Global Food Security-support Analysis Data (GFSAD) with data from 1990 to 2017, with a 30 m resolution (Fig. 3). Map services and data are made available from the U.S. Geological Survey National Geospatial Program and can be reproduced with no restrictions. The data were downloaded from NASA's Land Processes Distributed Active Archive Center [24].

License: See https://lpdaac.usgs.gov/products/gfsad30saafgircev001/

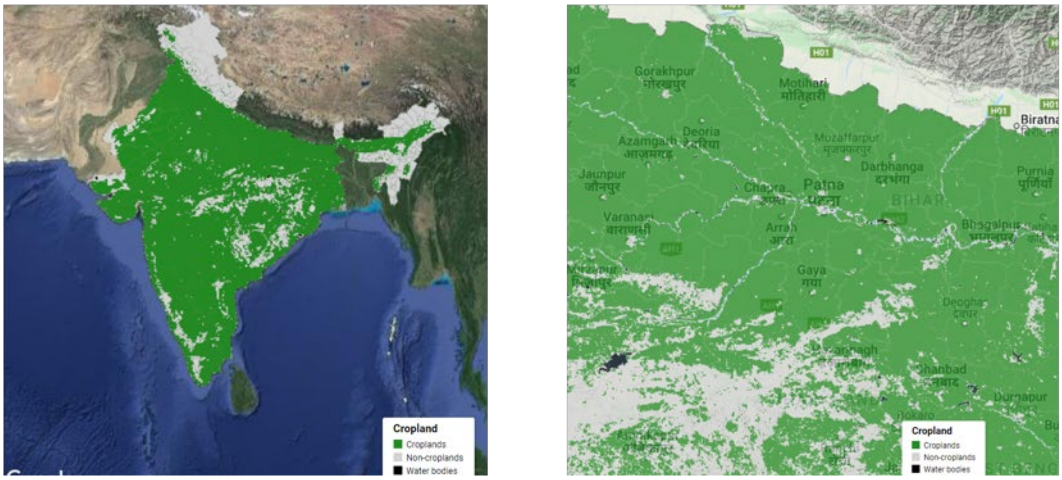


Figure 3: Cropland mask layer (Left: India, Right: Zoomed in on Bihar). The data is color-coded to differentiate between cropland (green), non-cropland (gray), and water bodies (black).

3.2 Elevation

Elevation (in m) was sourced from the Japan Aerospace Exploration Agency (JAXA) Earth Observation Research Center [13] (Fig. 4). The data were prepared by JAXA as part of the ALOS World 3D – 30 m (AW3D30) dataset, a global digital surface model (DSM) dataset with 30 m resolution, and imported directly in Google Earth Engine.

License: See https://earth.jaxa.jp/en/data/policy/index.html

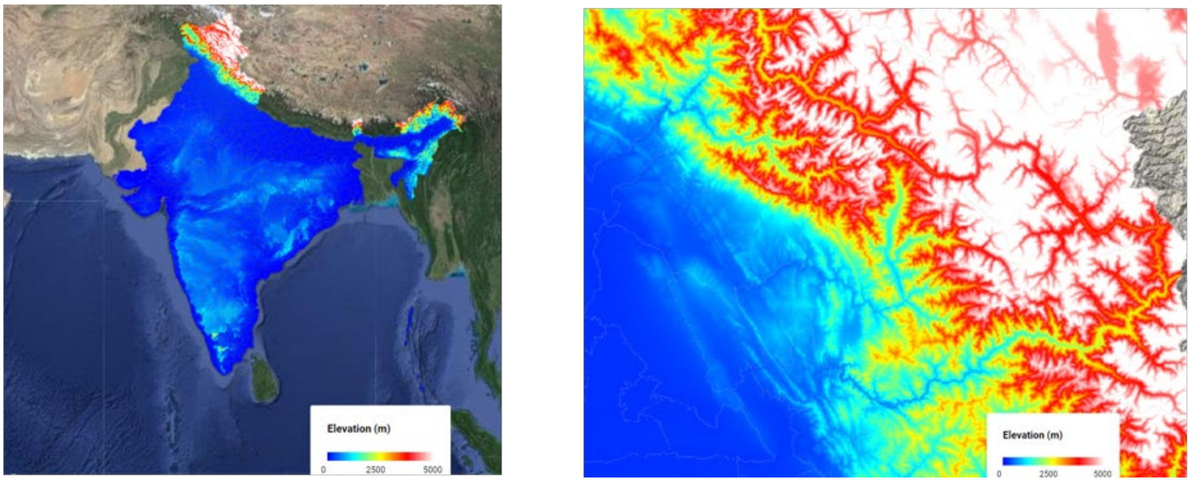


Figure 4: Elevation layer (Left: India, Right: Zoomed in on Himachal Pradesh).

3.3 Predicted electricity network lines

This layer displays the predicted location of electricity network lines, as computed in [25] using satellite imagery of night-time lights and OpenStreetMap data (Fig. 5). The electricity network lines represent (vectorized) predicted distribution and transmission line network, together with existing OpenStreetMap lines. The data were downloaded from [26].

License: Creative Commons Attribution 4.0 International

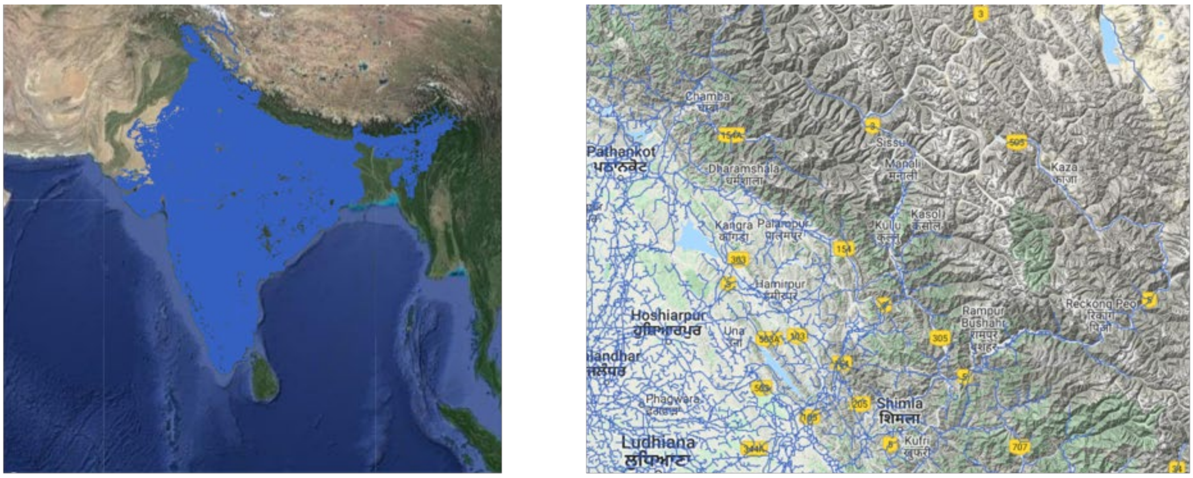


Figure 5: Predicted electricity network lines layer (Left: India, Right: Zoomed in on Himachal Pradesh).

3.4 Agricultural data (ICRISAT data)

The following layers (fruits and vegetable crop production and cold storage distribution) were downloaded from the ICRISAT (International Crop Research Institute for the Semi-Arid Tropics) [27].

License: Creative Commons Attribution 4.0 International

3.4.1 Crop production (fruits and vegetables)

These layers represent district-wise statistics of crop production for fruits and vegetables (Fig. 6). The data were sourced from the State Horticultural Department [27]. Displayed data show perishable crops of interest and refer to the year 2016.

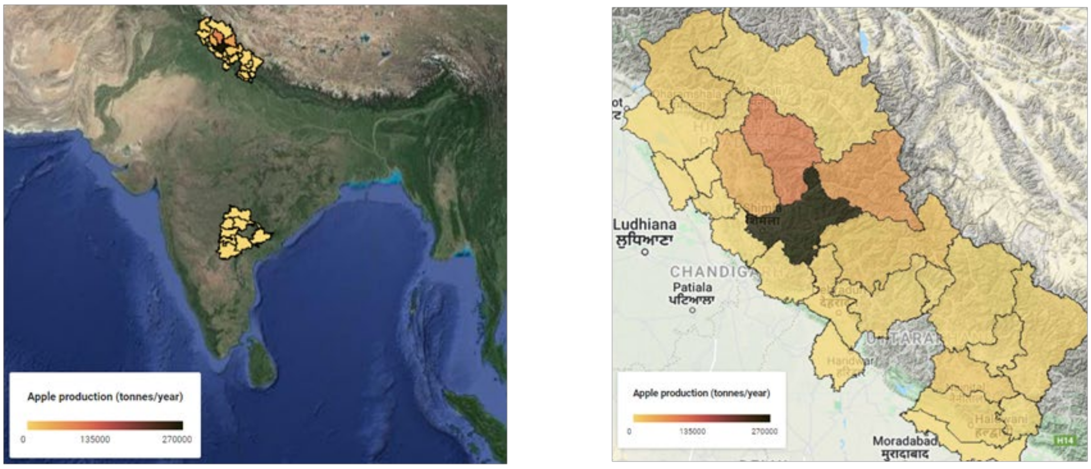


Figure 6: Fruits production layer, an example of apple production (Left: India, Right: Zoomed in on Himachal Pradesh (Terrain background with 80% opacity)).

3.4.2 Cold storage locations

District-wise cold storage data were sourced from the Central Warehousing Corporation [27] (Fig. 7). Displayed data represents the total number of cold storage rooms available in each district for fruits and vegetables or for mixed-use in the year 2009.

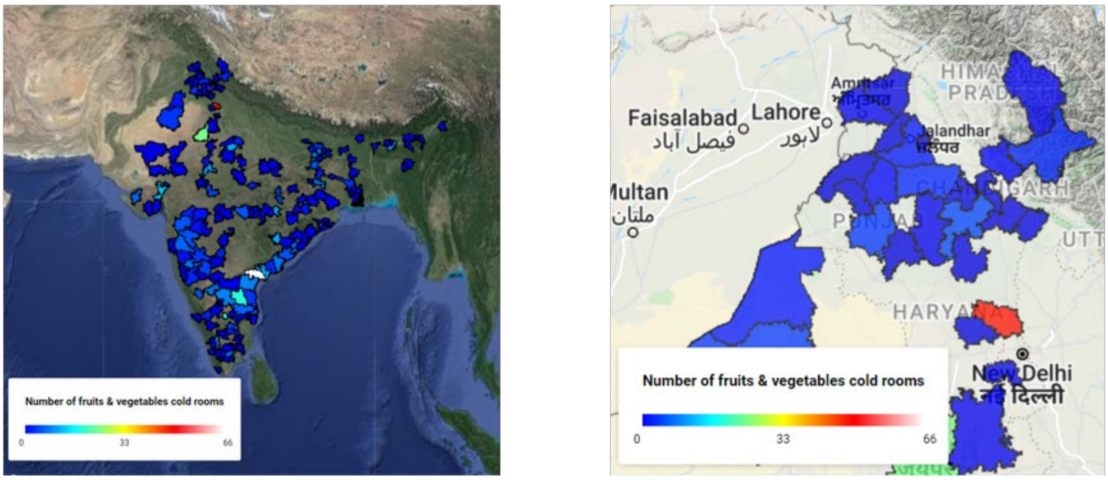


Figure 7: Cold storage locations (Left: India, Right: Zoomed in on Himachal Pradesh (Terrain background with 80% opacity)).

3.5 Harvesting season

The harvesting season of apples, bananas, and potatoes was sourced from [28]–[30] respectively for each state in India (Fig. 8). The original data were images containing state names in rows and months in columns. We manually converted them to texts and merged them with state-level boundary layers (Fig 2a).

License: unknown

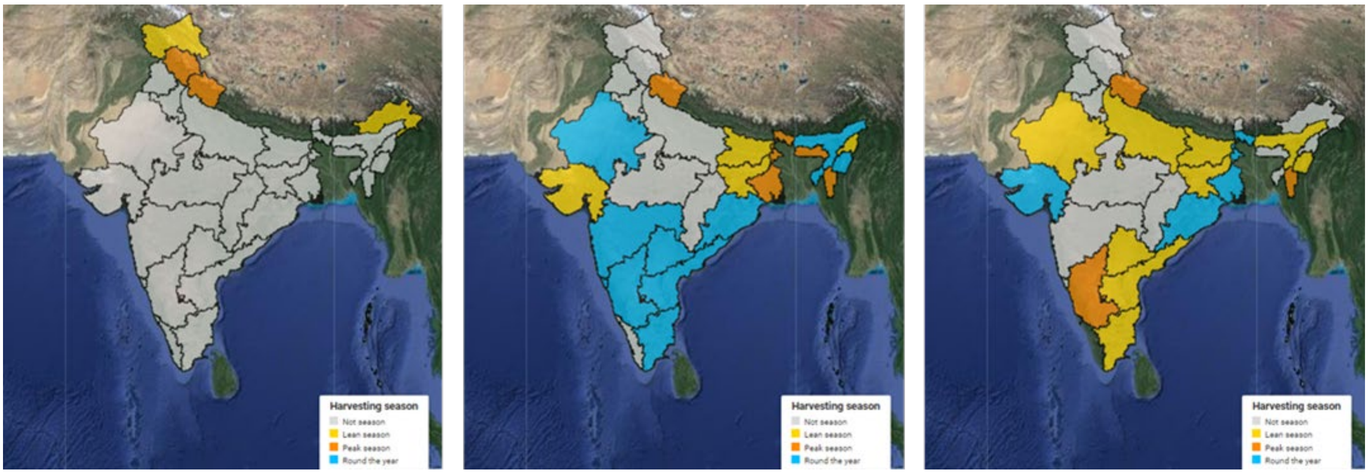


Figure 8: Harvesting season, examples of apples (left), bananas (middle), and potatoes (right) in August. The data is color-coded to differentiate between Not season (gray), Lean season (gold), Peak season (orange), and Round the year (turquoise).

3.6 Climate data

Temperature and solar radiance data were downloaded from the ERA5 database [31]**.** The name of the dataset used is the 'ERA5 monthly averaged data on single levels from 1979 to present'. The names of the extracted parameters are '2m temperature' and 'Total sky direct solar radiation at the surface' for temperature and solar radiance, respectively. The spatial resolution is approximately 30 km (0.25 º latitude and 0.25 º longitude). For this project, we used the data from 2020 as that was the latest year with complete datasets for the whole year.

The format of the downloaded file was NetCDF. We imported it in R (version 4.1.0) [19], clipped it to India (as described in Section 2.1), created a layer for each parameter for each month, and exported them as Geotiff files. We used the 'raster' and 'ncdf4' packages for this processing.

License: See 'License' tab at https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview

3.6.1 Temperature

This layer represents a monthly averaged temperature (in °C) for 2020 (Fig. 9). We converted the original dataset from K to °C.

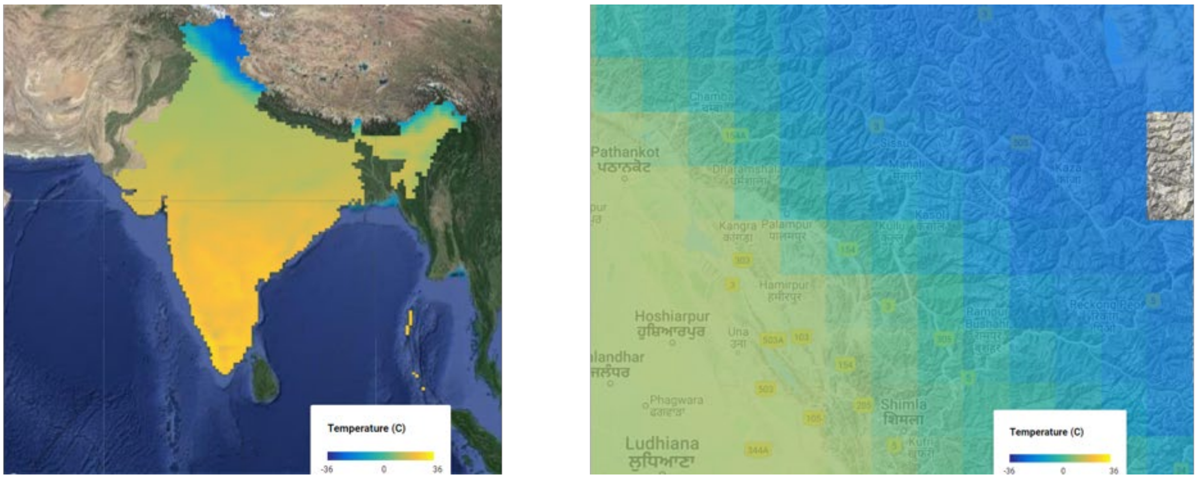


Figure 9: Temperature layer, an example of January 2020 (Left: India, Right: Himachal Pradesh (Terrain background with 80% opacity)).

3.6.2 Solar radiance

The monthly solar radiance of 2020 (in J/m^2) represents the amount of both direct and diffused solar radiation at the earth's surface (Fig. 10).

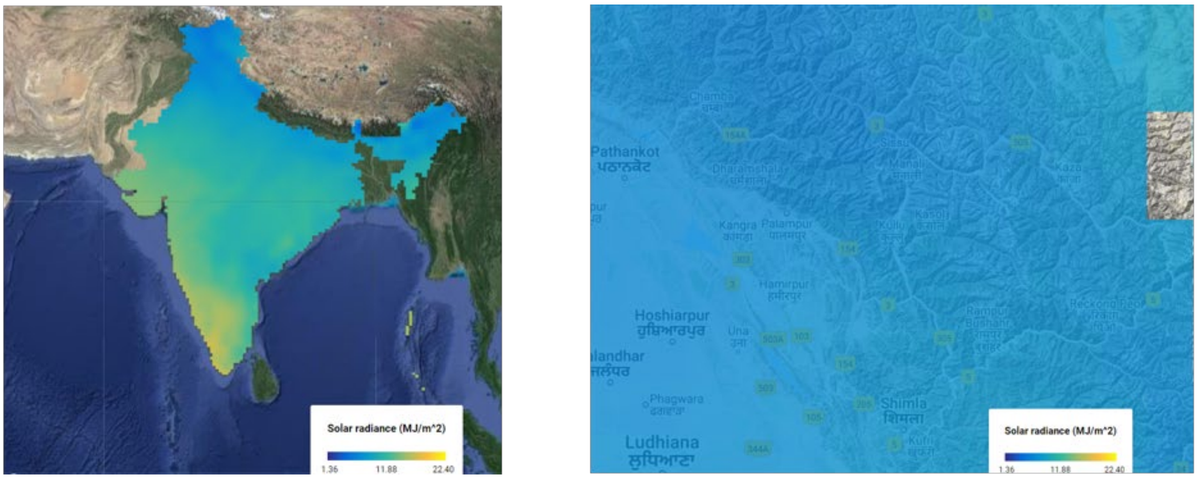


Figure 10: Solar radiation layer, an example of January 2020 (Left: India, Right: Himachal Pradesh (Terrain background with 80% opacity)).

3.7 Roads

This layer represents the road network from OpenStreetMap (Fig. 11). The data were downloaded from [33].

License: Open Database 1.0 License

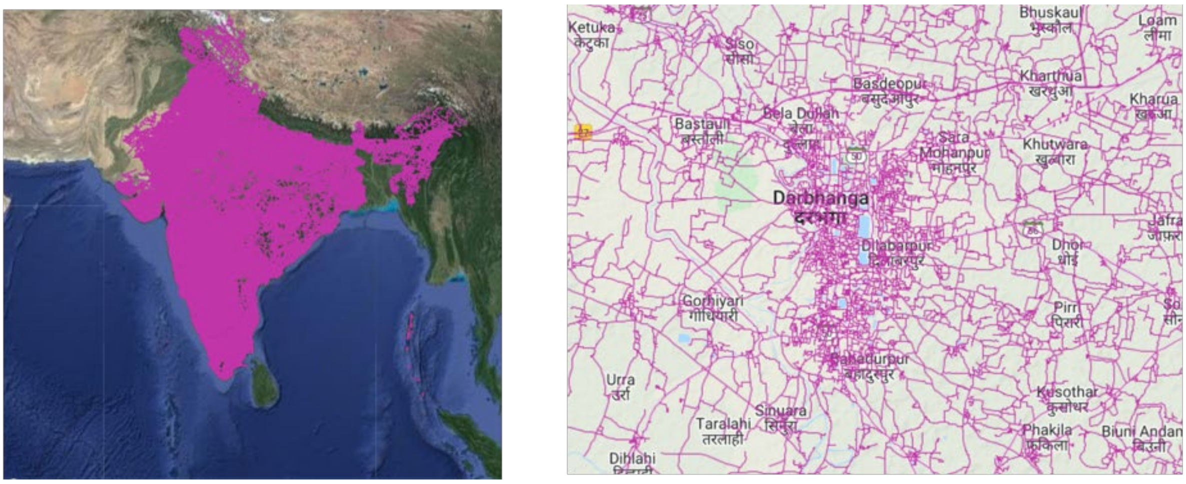


Figure 11: Roads layer (Left: India, Right: Darbhanga, Bihar (Terrain background)).

3.8 Agricultural survey (Mission Antyodaya)

The following parameters (availability of markets, households majorly engaged in farming, distribution of FPOs, and cultivated area) were extracted from the Mission Antyodaya 2019 report [9]. All data were scraped from the web using Python (version 3.8.8) and the Selenium package.

License: unknown

3.8.1 Availability of markets (Mandis)

This layer displays the percentage of villages in each administrative level (district or block) where a Mandi was reported as available (Fig. 12). The data were web scraped from the report 'Villagewise Report on Infrastructure' (the report column 'Availability of markets: Mandis' has been normalized to 'No. of Villages Where Survey completed'; the report also contains columns 'Availability of markets: Regular market' and 'Availability of markets: weekly haat', which we did not take into account for this application).

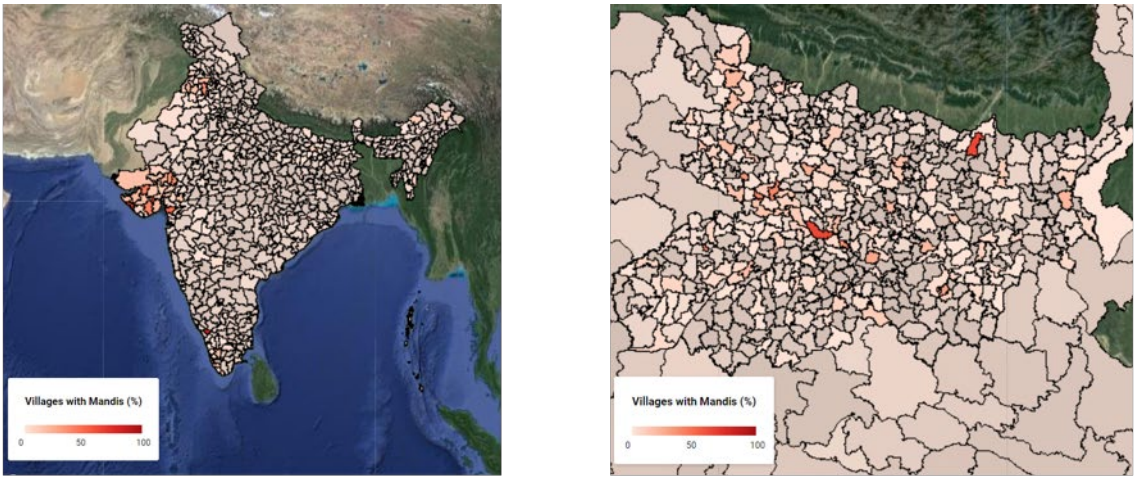


Figure 12: Market availability layer (Left: India, Right: Bihar).

3.8.2 Households engaged in farming

This layer displays the percentage of households (HHs) in each administrative level (district or block) that are majorly involved in farm activities (Fig. 13). The data were web scraped from the report 'Agriculture' (the report column 'Total No. OfHHs engaged majorly in Farm activities' has been normalized to 'Total Household').

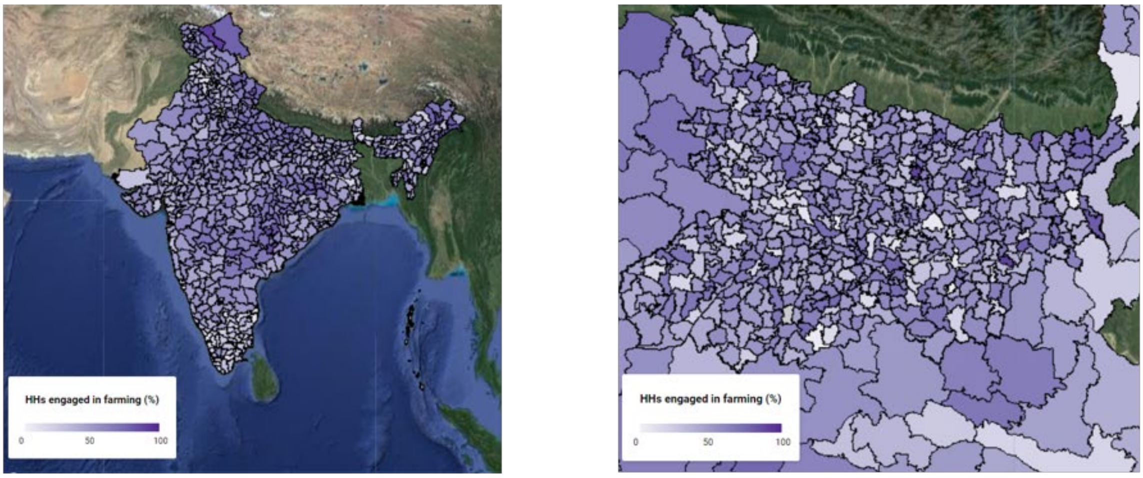


Figure 13: Households engaged in farming layer (Left: India, Right: Bihar).

3.8.3 Distribution of Farmer Producers Organizations (FPOs)

This layer displays the percentage of villages in each administrative level (district or block) that are reported to have at least one FPO (Fig. 14). The data were web scraped from the report ‘Agriculture’, and the columns 'Farmers Collective-FPOs(Count Of Villages)' and 'Farmers Collective-Both(Count Of Villages)' have been added up and normalized to 'No. of Villages Where Survey completed'.

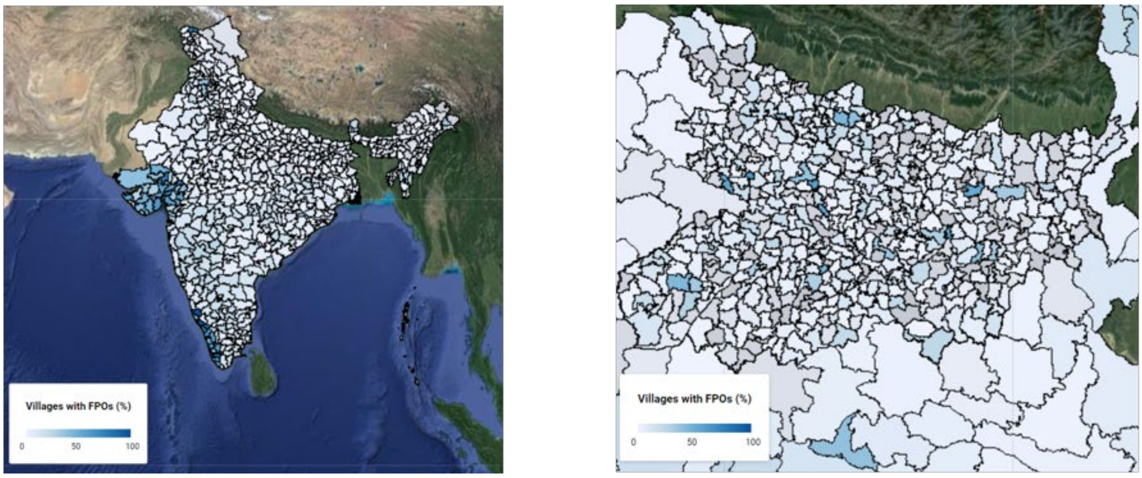


Figure 14: Distribution of Famer Producers organizations layer (Left: India, Right: Bihar).

3.8.4 Cultivable area

This layer displays the percentage of the cultivable area in each administrative level (district or block) (Fig. 15). The data were scraped from the report ‘Land Use and Irrigation’ (report column: 'Total cultivable Area(ha.)'). The original data were in hectares. For visualization and further calculations for potential cooling storage locations (Section 4.2), the data were normalized by the area of the polygons representing each district or block (Section 2.2).

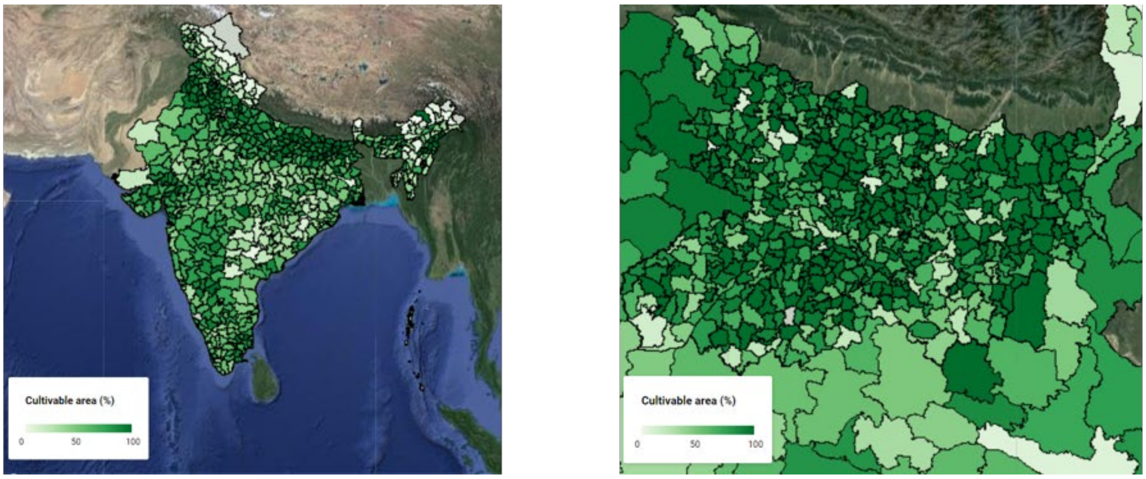


Figure 15: Cultivable area layer (Left: India, Right: Bihar).

3.8.5 Population density

This layer displays the population density (in number of people per km squared) in each administrative level (district or block) (Fig. 16). The data were scraped from the report ‘Villagewise Report on

Infrastructure' (report column: 'Total Population') and normalized by the area of the polygons representing each district or block (Section 2.2).

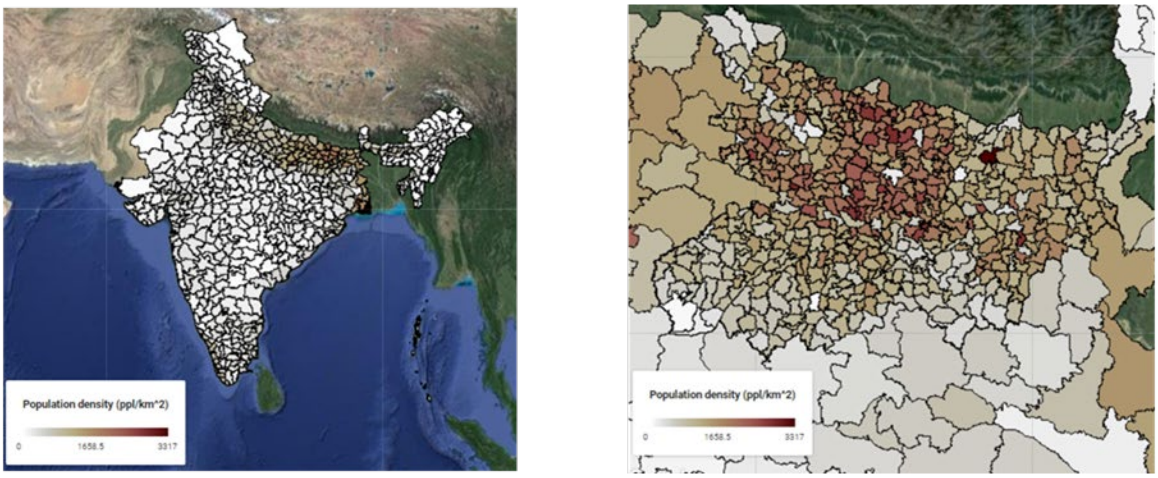


Fig 16. Population density layer (Left: India, Right: Bihar).

3.9 Market locations

This layer presents the amenities identified as 'marketplaces' in OpenStreetMap (Fig. 17). The data were downloaded from [34]. Marketplaces identified as 'waypoints' and 'track points' were merged in QGIS (version 3.18.2) [18] before the layer was cropped. Note that this layer contains around 13,000 markets across India, but is likely not complete: based on the information from the agricultural survey data (see Section 3.7.1), we would expect at least 15,000 Mandis (and around 150,000 markets if we also consider 'regular markets' and 'weekly haat').

License: Open Database 1.0 License

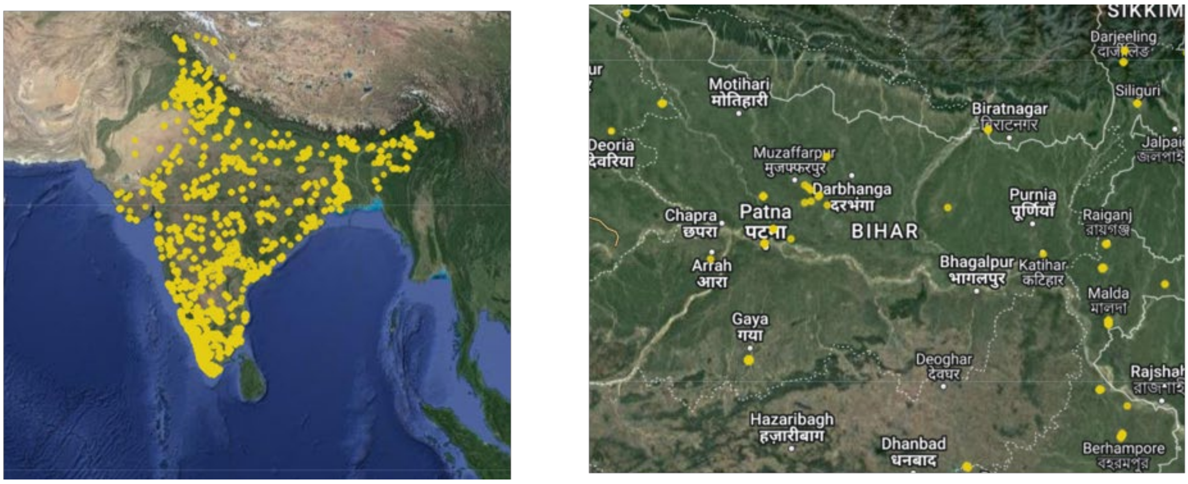


Figure 17: Market locations layer (Left: India, Right: Bihar).

3.10 Mobile broadband coverage

This layer displays the world mobile broadband coverage, which is a 1 km resolution raster representation of cellular mobile wireless Internet access (Fig. 18). The data were downloaded from [35].

License: See https://data.apps.fao.org/map/catalog/srv/eng/catalog.search#/metadata/7ee38f75-605f-4c88-9afc-64779e70e595

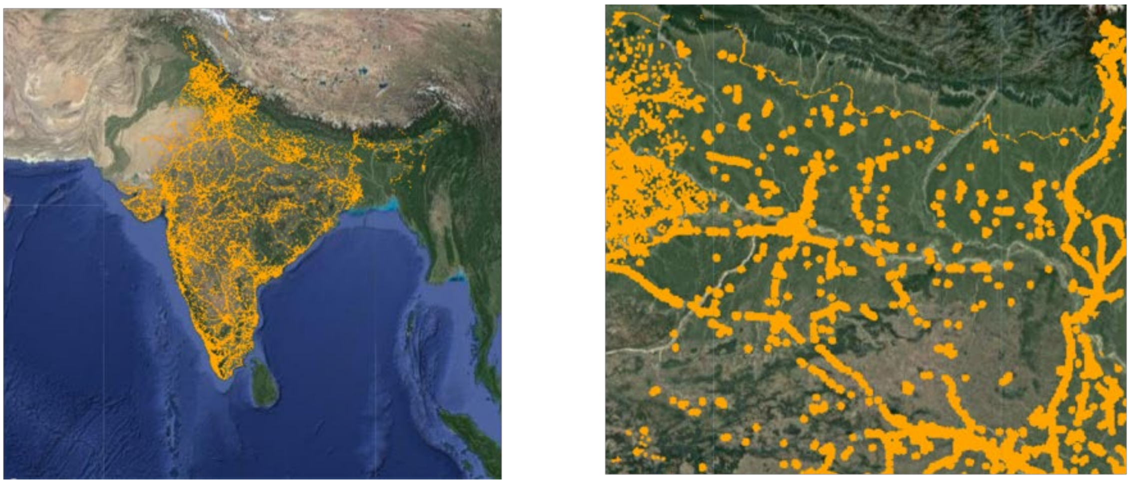


Figure 18: Mobile broadband coverage layer (Left: India, Right: Bihar).

3.11 Shelf life gain

This layer (Fig. 19) represents by how many days the shelf life (i.e. the timespan a commodity can last without spoiling) of selected crops can be extended by storing them at their optimal temperature, compared to keeping them under ambient temperature (Fig. 9). Thus, this layer shows the advantage of storing fresh produce in cooling storage. The calculations were conducted for potatoes, apples, and bananas. The shelf life at the ambient temperature was calculated by using temperature data (Section 3.6.1) as inputs for the temperature-based kinetic rate model (see Section 4.1 for details). The optimal temperature and the shelf life under this optimal temperature for a corresponding crop were obtained from FAO [32]. The details of the model and calculations are in Section 4.1.

License: Creative Commons Attribution 4.0 International

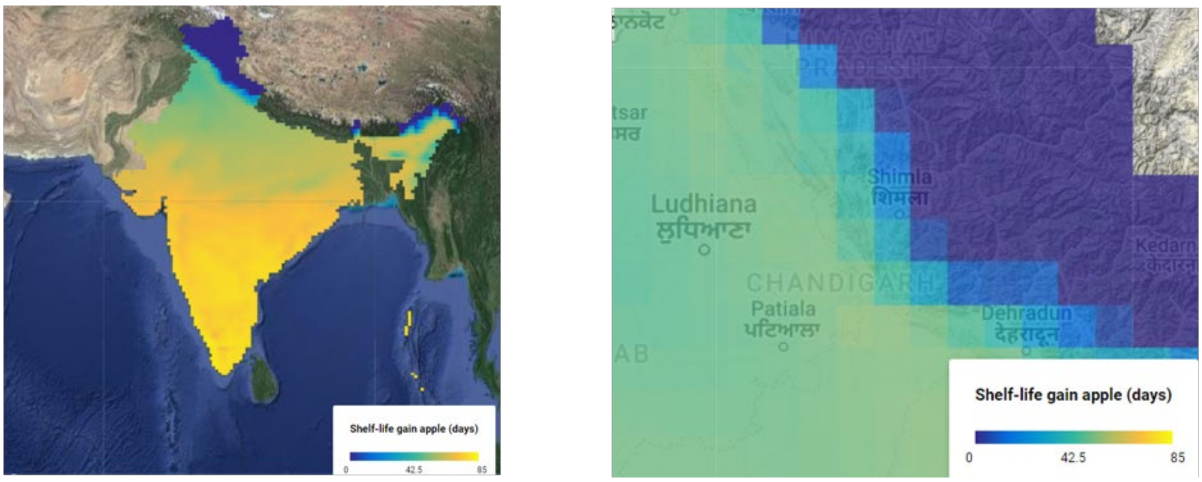


Figure 19: Shelf life gain layer, an example of apple in January 2020 (Left: India, Right: Himachal Pradesh (Terrain background with 80% opacity)).

3.12 Predicted cold room locations

This layer presents areas that could be suitable to install a cooling unit based on the available open-source data (Fig. 20). This information was calculated based on the vicinity to a market, availability of close-by roads and stable network connection, amount of available Mandis and households dedicated to farming, and cultivable area in the region. Please refer to Section 4.2 for details on the computation. Note that this map is case-specific and depends on specific assumptions about layers to consider and their thresholds, which may vary from one stakeholder to another. This prediction is meant as a guiding tool only, and additional and specific on-the-ground insights that are not captured in the available data will be required to make final decisions.

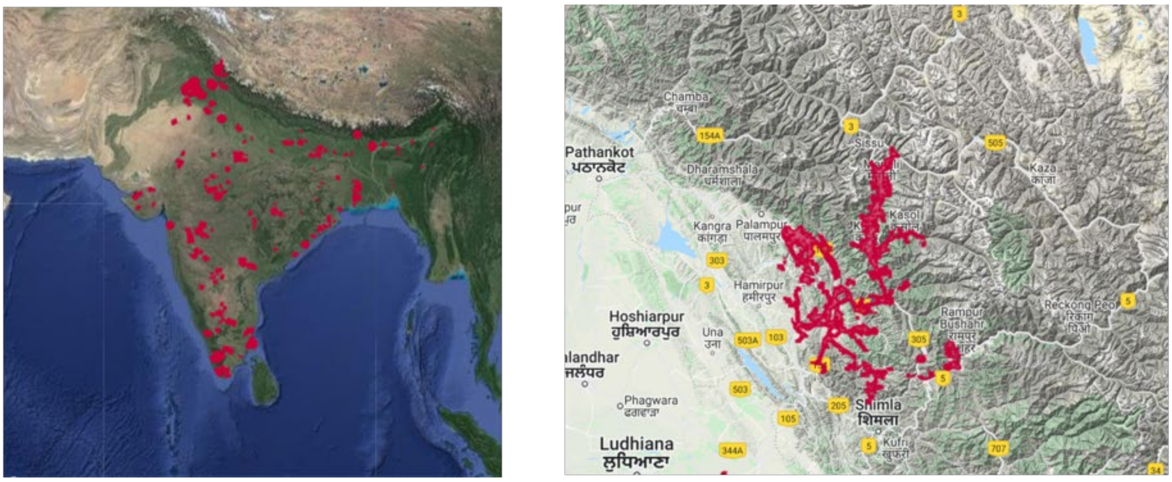


Figure 20: Potential cold room location layer (Left: India, Right: Himachal Pradesh (Terrain background)).

3.13 Cropland predictions in Himachal Pradesh

This layer presents cropland predictions for 2020 made by Li et al. (2022) [16] for three districts in Himachal Pradesh (Mandi, Kullu, and Shimla), illustrated in Fig. 21. The layer is colorized by altitude using the ALOS 30m elevation map (see Section 3.2). Non-croplands predictions are not visualized. The map was generated using a Random Forest classifier and feature engineering from a time series of Sentinel-2 satellite images. The ground truth data used for training, validation, and testing the classifier was manually annotated in QGIS using a combination of field survey reference points and visual interpretation of very high resolution (VHR) imagery. The generalization capability of the model over the three districts has been validated with three hold-out sets, one for each district. The data and code for reproducing the model are available open-source on GitHub.

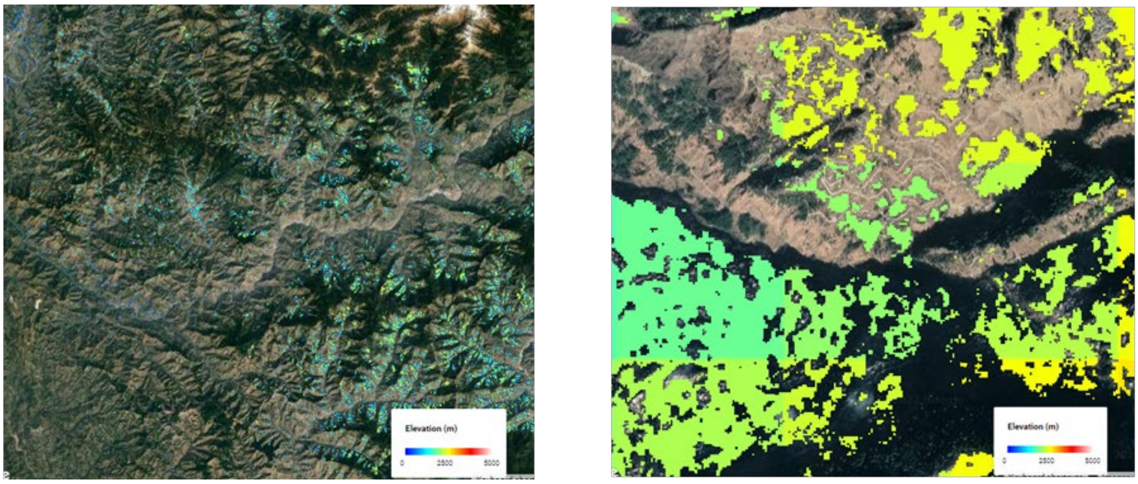


Figure 21: High-altitude 10 m resolution cropland predictions in Himachal Pradesh (Right: Zoomed in on one region in Himachal Pradesh).

4 Supplemental material

4.1 Shelf life gain calculations

Shelf life gain (Section 3.11) was calculated by comparing shelf life (SL_c) at ambient temperature (T) and shelf life at the optimal temperature ($SL_{opt,c}$) (Eq. 1). Note that in Eq. 1 and the following equations, subscript *c* indicates the specific crop. We assume here that the food can be stored in a cold store under these optimal conditions.

$$Shelf\ life\ gain = SL_{opt,c} - SL_c(T) \quad \text{Eq. 1}$$

For the shelf life at optimal condition ($SL_{opt,c}$), the middle point of shelf life under the optimal conditions from FAO [32] was used (Table 1).

Table 1: Shelf life of crops at the optimal temperature, obtained from FAO [32].

Crop	Temperature (°C)	Shelf life (days)
Apple	1.5	105
Banana	14	17.5
Potato(late)	8.75	225

For the shelf life at ambient temperature, the first-order kinetic rate law model (n=1) was used, as described in [36], [37] (Eq. 2).

$$-\frac{dA(t)}{dt} = k(T) \cdot A(t)^n \quad \text{Eq. 2}$$

In Eq. 2, A represents a crop quality index, which is assumed to be 1 at harvest (A_0 [-]).T is ambient temperature [K], which is the monthly temperature from ERA5 (Section 3.6.1), and k is a rate constant, which is temperature-dependent (Eq. 3).

$$k(T) = k_{0,c}\, e^{\frac{-E_{a,c}}{RT}} \quad \text{Eq. 3}$$

In Eq. 3, $k_{0,c}$ is the quality rate constant [s^{-1}], and $E_{a,c}$ is the activation energy [J mol^{-1}]. R is the ideal gas constant (8.314 J mol^{-1} K^{-1}). The values of $k_{0,c,}$ and $E_{a,c}$ were determined based on each crop's temperature coefficient (Q_{10}). Q_{10} represents the rate of quality change for a 10 °C increase in temperature based on biochemical reaction (Eq. 4) [36]–[38]. For fresh produce, the Q_{10} value is commonly 2 to 3 [37]. A Q_{10} value of 3 was used for apples in this project, calculated based on [38], and 2 for banana and potato.

$$Q_{10} = \frac{k(T+10)}{k(T)} \quad \text{Eq. 4}$$

Based on Eq. 3 and Eq. 4 $E_{a,c}$ can be written as Eq. 5.

$$Ea,c \approx \frac{RT^2}{10} ln\, Q_{10} \quad \text{Eq. 5}$$

By rearranging the equations, $k_{0,c}$ was obtained for each crop (Eq. 6)

$$k_{0,c} = \frac{k(t)}{e^{\frac{-E_{a,c}}{RT}}} = \frac{-\frac{1}{SL_{opt,c}}\, ln(\frac{A_{thr}}{A_0})}{e^{\frac{-E_{a,c}}{RT}}} \quad \text{Eq. 6}$$

Note that in Eq. 6, $SL_{opt,c}$ is in seconds. For this project, the threshold of A (A_{thr} [-]) was set to 0.1 to represent that crops are no longer marketable. Finally, the remaining shelf life (SL_c) at ambient temperature was calculated as Eq. 7. For the shelf life layer (Section 3.11), shelf life was converted from seconds to days.

$$SL_c(T) = \frac{-\ ln(\frac{A_{thr}}{A_0})}{k(T)} \quad \text{Eq. 7}$$

The calculations were done in R (version 4.1.0) [19]. As was the case for the climate data (Section 3.6), we created a Geotiff file for each month, using 'raster' and 'ncdf4' packages.

4.2 Potential cooling unit location identification

To calculate potential promising locations to install new cooling units (Section 3.12), we used some of the layers discussed in Section 3 and displayed them in the web map. We restricted ourselves to the following layers:

- Roads (Section 3.7)
- Availability of markets (Mandis) (Section 3.8.1)
- Households engaged in farming (Section 3.8.2)
- Cultivable area (Section 3.8.4)
- Market locations (Section 3.9)
- Mobile broadband coverage (Section 3.10)

We identified locations that are no more than 20 km away from a market, at most 500 m away from a road, have stable network coverage, and in regions (districts or blocks) where the other indicators are above the median for the state (or district in case of the blocks) they belong to. Thus, the thresholds for this filtering process are state-dependent. For the state of Himachal Pradesh, we have acquired proprietary data about market locations [39], which we merged with the market location data from OpenStreetMap and used in the computations. However, only the open-source data are displayed in the market locations layer (Section 3.8).

The model was implemented in QGIS (version 3.18.2) [18] and ran at the state level (Fig. 1). For the states of Odisha and Bihar, the medians of each indicator were calculated for the district that each block belongs to (Fig. 2). Ideally, we would like to replicate the approach used for Odisha and Bihar for all states. However, we were unable to get open-source block boundaries for any other state. Note that this was the same challenge we faced in preprocessing the dataset for visualization (Section 2.2). As the last step, the predictions for each state were merged in a single layer in QGIS (version 3.18.2) [18].

This layer is intended to suggest areas that service providers can explore to place new cooling units. However, the Your VCCA team declines any responsibility for any errors or omissions in the information provided. In a future release of this application, we plan to make the map more interactive by allowing users to compute promising cold room locations by dynamically deciding which layers to include and their thresholds.

Author contributions

TM and TD conceptualized the project and acquired funding; TM and TD did the project administration; KS, RE, and JG collected the data, performed the investigation, and developed the methodology; KS and RE did the processing of the data and analysis; KS, RE, and JG did the visualization, with key input from TM and TD; JG did the initial setup of the web map; JG and RE further developed the functionality of the web map with key inputs from KS, TD, and TM; KS and RE wrote the original draft of the paper; all authors performed critical review and editing of the paper.

References

[1] A. Deshpande and D. Riehle, “The total growth of open source,” *IFIP Int. Fed. Inf. Process.*, vol. 275, no. Open Source Development, Communities and Quality; Barbara Russo, Ernesto Damiani, Scott Hissam, Björn Lundell, Giancarlo Succi; (Boston: Springer), pp. 197–209, 2008.

[2] D. Kumar, S. Narwal, and S. Phougat, “A review of rural development schemes in India,” *Asian J. Sociol. Res.*, vol. 5, no. 4, pp. 18–26, Dec. 2021, Accessed: Jan. 20, 2022. [Online]. Available: https://www.globalpresshub.com/index.php/AJSR/article/view/1373.

[3] FAO, “India at a glance | FAO in India,” *Food and Agriculture Organization of the United Nations*. https://www.fao.org/india/fao-in-india/india-at-a-glance/en/ (accessed Jan. 21, 2022).

[4] G. Nedumaran and M. M, “E-agriculture and rural development in India,” *SSRN Electron. J.*, Jan. 2020, doi: 10.2139/SSRN.3551994.

[5] D. Chandra Misra and P. Kumar Mi, “E-governance and digitalization of Indian rural development,” 2019, doi: 10.1145/3326365.3326438.

[6] P. Dhage, S. Wathore, and V. Jagtap, “E-Gram Panchayat management system,” *Open access Int. J. Sci. Eng.*, vol. 3, no. 12, 2007.

[7] G. S. Raghavendra, M. Shankar Lingam, and J. Vanishree, “ICT for good governance: Evidence from development perspective,” *Inf. Commun. Technol. Compet. Strateg. (ICTCS 2020)*, vol. 190, no. Lecture Notes in Networks and Systems 190, pp. 627–637, 2021, doi: 10.1007/978-981-16-0882-7_55.

[8] J. Chathukulam, J. Manasi, V. Rekha, C. V Balamurali, and T. Thilakan, “Mission Antyodaya: Well envisioned but poorly understood,” *Gandhi Marg Q. 43(2)*, vol. 43, no. 2, pp. 151–186, 2021, Accessed: Jan. 21, 2022. [Online]. Available: http://gandhimargjournal.org/.

[9] M. of R. D. Government of India, “Mission Antyodaya 2019.” https://missionantyodaya.nic.in/ma2019/ (accessed Jun. 15, 2021).

[10] M. I. Khalge, R. R. Chole, and P. B. Bhosale, “Role performance of Grampachayat members,” *Hind Agri-Horticultural Soc.*, vol. 5, no. 1&2, pp. 47–51, 2010, Accessed: Jan. 24, 2022. [Online]. Available: https://www.cabdirect.org/cabdirect/FullTextPDF/2010/20103147652.pdf.

[11] S. K. Seelan, S. Laguette, G. M. Casady, and G. A. Seielstad, “Remote sensing applications for precision agriculture: A learning community approach,” *Remote Sens. Environ.*, vol. 88, no. 1–2, pp. 157–169, Nov. 2003, doi: 10.1016/J.RSE.2003.04.007.

[12] M. S. Moran, Y. Inoue, and E. M. Barnes, “Opportunities and limitations for image-based remote sensing in precision crop management,” *Remote Sens. Environ.*, vol. 61, no. 3, pp. 319–346, Sep. 1997, doi: 10.1016/S0034-4257(97)00045-X.

[13] T. Tadono, H. Ishida, F. Oda, S. Naito, K. Minakawa, and H. Iwamoto, “Precise Global DEM Generation by ALOS PRISM,” *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. II–4, pp. 71–76, Apr. 2014, doi: 10.5194/isprsannals-II-4-71-2014.

[14] H. Hersbach *et al.*, “The ERA5 global reanalysis,” *Q. J. R. Meteorol. Soc.*, vol. 146, no. 730, pp. 1999–2049, Jul. 2020, doi: 10.1002/QJ.3803.

[15] Li, D., Gajardo, J., Volpi, M., & Defraeye, T. (2022). Using Machine Learning to generate an open-access cropland map from satellite images time series in the Indian Himalayan Region. arXiv preprint arXiv:2203.14673.

[16] R. Hijmans, “First-level Administrative Divisions, India, 2015,” *UC Berkeley, Museum of Vertebrate Zoology*, 2015. http://purl.stanford.edu/mw277wc3858 (accessed Jul. 05, 2021).

[16] Google, “Earth Engine Data Catalog | Google Developers.” https://developers.google.com/earth-engine/datasets (accessed Jan. 22, 2022).

[18] QGIS Development Team, “QGIS Geographic Information System. Open Source Geospatial Foundation Project.” 2021.

[19] R Core Team, “R: A Language and Environment for Statistical Computing.” 2020.

[20] R. J. Hijmans, “Second-level Administrative Divisions, India, 2015,” *University of California, Berkeley, Museum of Vertebrate Zoology*, 2015. https://geodata.mit.edu/catalog/stanford-zb090kx3567.

[21] J. Meyers, “Odisha_2021_Official_Boundaries,” *github*. https://github.com/justinelliotmeyers/Odisha_2021_Official_Boundaries (accessed Aug. 18, 2021).

[22] J. Meyers, “official_bihar_india_village_boundary_shapefile,” *github*. https://github.com/justinelliotmeyers/official_bihar_india_village_boundary_shapefile (accessed Aug. 18, 2021).

[23] SeatGeek, “fuzzywuzzy: Fuzzy String Matching in Python.” https://github.com/seatgeek/fuzzywuzzy (accessed Aug. 20, 2021).

[24] Land Processes Distributed Active Archive Center, “Global Food Security-support Analysis Data (GFSAD) Cropland Extent data from the GFSAD30SAAFGIRCE product over Indiar,” *NASA*, 2017. https://lpdaac.usgs.gov/products/gfsad30saafgircev001/ (accessed Oct. 09, 2021).

[25] C. Arderne, C. Zorn, C. Nicolas, and E. E. Koks, “Predictive mapping of the global power system using open data,” *Sci. Data 2020 71*, vol. 7, no. 1, pp. 1–12, Jan. 2020, doi: 10.1038/s41597-019-0347-4.

[26] C. Arderne, C. NIcolas, C. Zorn, and E. E. Koks, “Data from: Predictive mapping of the global power system using open data,” Jan. 2020, doi: 10.5281/ZENODO.3628142.

[27] ICRISAT (International Crop Research Institute for the Semi-Arid Tropics) and IFPRI (International Food Policy Research Institute), “ICRISAT-District Level Data.” http://data.icrisat.org/dld/src/additional.html (accessed Nov. 24, 2021).

[28] APEDA (Agricultural and processed food products Export development authority), “Harvesting Season Apple,” Ministry of Commerce & Industry, Government of India. https://apeda.gov.in/apedawebsite/six_head_product/Harvesting_Season_Apple.htm (accessed May 03, 2021).

[29] APEDA (Agricultural and processed food products Export development authority), “Harvesting Season Banana,” Ministry of Commerce & Industry, Government of India. https://apeda.gov.in/apedawebsite/six_head_product/Harvesting_Season_banana.htm (accessed May 03, 2021).

[30] APEDA (Agricultural and processed food products Export development authority), “Harvesting Season Potato,” Ministry of Commerce & Industry, Government of India. https://apeda.gov.in/apedawebsite/six_head_product/Harvesting_Season_Potato.htm (accessed May 03, 2021).

[31] H. Hersbach *et al.*, “ERA5 monthly averaged data on single levels from 1979 to present,” *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*, 2019. doi: 10.24381/cds.adbb2d47 (accessed Jul. 02, 2021).

[32] A. F. L. Camelo, “Manual for the preparation and sale of fruits and vegetables,” *FAO*, 2004. http://www.fao.org/3/y4893e/y4893e06.htm (accessed Dec. 14, 2020).

[33] “Geofabrik Download Server,” *Geofabrik GmbH, OpenStreetMap*.

http://download.geofabrik.de/asia/india.html (accessed Jun. 30, 2021).

[34] “overpass turbo.” https://overpass-turbo.eu/ (accessed Dec. 06, 2021).

[35] “Mobile Broadband Coverage 2021 (Global - 1km),” *FAO Map Catalog*, Oct. 04, 2021. https://data.apps.fao.org/map/catalog/srv/eng/catalog.search#/metadata/7ee38f75-605f-4c88-9afc-64779e70e595 (accessed Nov. 17, 2021).

[36] K. Shoji, S. Schudel, D. Onwude, C. Shrivastava, and T. Defraeye, “Mapping the postharvest life of imported fruits from packhouse to retail stores using physics-based digital twins,” *Resour. Conserv. Recycl.*, vol. 176, p. 105914, Jan. 2022, doi: 10.1016/J.RESCONREC.2021.105914.

[37] T. Defraeye *et al.*, “Digital twins probe into food cooling and biochemical quality changes for reducing losses in refrigerated supply chains,” *Resour. Conserv. Recycl.*, vol. 149, pp. 778–794, Oct. 2019, doi: 10.1016/j.resconrec.2019.06.002.

[38] W. Wu, P. Cronjé, B. Nicolai, P. Verboven, U. Linus Opara, and T. Defraeye, “Virtual cold chain method to model the postharvest temperature history and quality evolution of fresh fruit – A case study for citrus fruit packed in a single carton,” *Comput. Electron. Agric.*, vol. 144, pp. 199–208, Jan. 2018, doi: 10.1016/j.compag.2017.11.034.

[39] ML Info Map, “Geographic Information Systems (GIS).” https://www.mlinfomap.com/ (acquired Sep. 29, 2021).