



## Madagascar rural observatory surveys, a longitudinal dataset on household living conditions 1995-2015

Velomalala Solo Andrianjafindrainibe, Nicole Andrianirina, Florent Bédécarrats, Isabelle Droy, Jean-Luc Dubois, Jeanne de Montalembert, Bako Nirina Rabevohitra, Rolland Rafidimanana, Patrick Rasolofo, Raphaël Ratovoarinony, et al.

### ► To cite this version:

Velomalala Solo Andrianjafindrainibe, Nicole Andrianirina, Florent Bédécarrats, Isabelle Droy, Jean-Luc Dubois, et al.. Madagascar rural observatory surveys, a longitudinal dataset on household living conditions 1995-2015. 2024. hal-04502779v2

**HAL Id: hal-04502779**

**<https://hal.science/hal-04502779v2>**

Preprint submitted on 2 Apr 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Data Descriptor Template

## Title

Madagascar rural observatory surveys, a longitudinal dataset on household living conditions 1995-2015

## Authors

Velomalala Solo Andrianjafindrainibe<sup>7</sup>, Nicole Andrianirina<sup>7</sup>, Florent Bédécarrats<sup>1,2</sup>, Isabelle Droy<sup>1,2</sup>, Jean-Luc Dubois<sup>1,2</sup>, Jeanne de Montalembert<sup>3</sup>, Bako Nirina Rabevohitra<sup>6</sup>, Rolland Rafidimanana<sup>7</sup>, Patrick Rasolofo<sup>5,2</sup>, Raphaël Ratovoarinony<sup>7</sup>, Lalaso Anjarafara Onivola Ratsaramiarina<sup>6</sup>, Jean Dieudonné Ravelonandro<sup>7</sup>, Voahirana Razanamavo<sup>7</sup>, Mireille Razafindrakoto<sup>1,3</sup>, Bezaka Rivolala<sup>4</sup>, François Roubaud<sup>1,3</sup>, Camille Saint-Macary<sup>1,3</sup>.

## Affiliations

1. French National Research Institute for Sustainable Development (IRD), Bondy, France
2. International Mixed Unit Sustainability and resilience (UMI SOURCE), Guyancourt, France
3. Research Mixed Unit Dauphine Economics Laboratory, Development, institutions and globalization research Team (UMR LEDa/DIAL), Paris, France
4. Malagasy Institute for Planification Techniques (IMATeP), Antananarivo, Madagascar
5. International Institute for Social Sciences (IISS), Antananarivo, Madagascar
6. Unit for emergency prevention and management (CPGU), Prime Minister Office, Antananarivo, Madagascar
7. Independent researcher

Corresponding author: Florent Bédécarrats ([florent.bedecarrats@ird.fr](mailto:florent.bedecarrats@ird.fr))

## Abstract

A Rural Observatory System (ROS) was established in Madagascar to address the lack of socioeconomic data on rural areas. It collected, analyzed, and disseminated data to help formulate and evaluate development policies. From 1995 to 2014, the ROS surveyed a total of 26 areas. The ROS methodology involved annual household panel surveys using consistent questionnaires supplemented by modules covering new themes and enriched questions tailored to specific contexts. Qualitative community surveys were used to understand local features and dynamics. The site selection combined quantitative and qualitative insights to reflect the diversity of Madagascar's rural challenges. Quality control was comprehensive, with measures such as limiting the number of daily surveyor interviews and daily field supervision. By making this data available for 21 consecutive years, along with documentation, metadata, and code with analysis examples, we aim to facilitate their discovery, assessment, and understanding by researchers, policymakers, and social organizations. To our knowledge, this is the only available data for an in-depth analysis of the situation and trends in the rural areas of Madagascar.

## Background & Summary

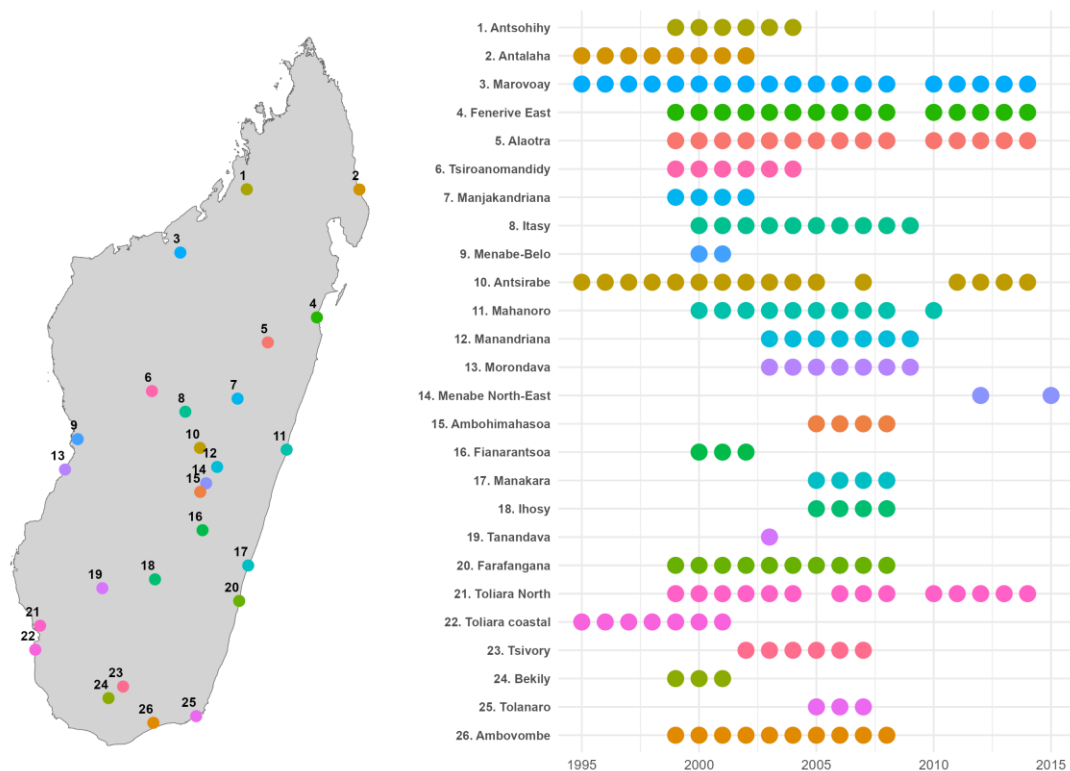
Rural households constitute the majority of the population of the least developed countries (64% in 2022) and account for the bulk of worldwide poverty<sup>1</sup>. Recent reductions in poverty rates in developing countries stem from a decrease in rural poverty<sup>2</sup>. Although improving living conditions in the countryside is one of the keys to enhancing welfare in developing countries, rural areas are generally neglected or poorly grasped by national statistical systems<sup>3-7</sup>. In Madagascar, the population has grown from 12 million in 1993 to 26 million in 2018, with 81%

of the population living in rural areas<sup>8</sup>. In the mid-1990s, Madagascar's rural areas underwent profound transformations. After a socialist period from 1975 to 1985, the country moved towards economic liberalization following the IMF recommendations: liberalization of pricing and marketing structures, and a shift from state-driven policies to market-oriented reforms<sup>9</sup>. The existing statistical system, which had deteriorated during the socialist era, was ill-equipped to capture these changes, leaving a gap in understanding the rural landscape evolution<sup>10</sup>. This situation was particularly concerning given that agriculture employed almost 80% of the active population, and that any analysis of the Malagasy economy would be incomplete without a comprehensive understanding of its rural sector.

This lack of information motivated the creation of four rural observatories in Madagascar in 1995<sup>11,12</sup>, each representing distinct agricultural challenges within the country's diverse ecosystems. The observatories consisted of annual panel survey, dedicated to the continuous collection, analysis and dissemination of socioeconomic data and insights, with the aim of identifying characteristics and dynamics across diverse agroecological contexts to guide the formulation, implementation, and evaluation of rural development policies<sup>13</sup>.

As the ROS emerged as an innovative and cost-effective way to provide a nuanced and continuous observation of rural households, its extension was supported by new funders and partners, such as research institutions, non-governmental organizations or consultancy firms<sup>14</sup>. The institutional anchoring of the rural observatories changed in 2000, moving from the National Statistical Office, with which it was initially associated, to the Rural Development Policy Unit, within the Ministry of Agriculture. A charter was drawn up to promote the methodological unity of the observatories and the ethics of intervention. Each partner involved was required to sign it to ensure methodological consistency across observatories and over the years.

Over the years, the number of observatories increased, reaching 17 in 2004, before contracting to five following the political crisis of 2009<sup>15</sup>. Despite these challenges, the ROS persisted, with three of the original observatories remaining to operate actively until 2014, as shown in Figure 1.



**Figure 1: Coarse location and years of data collection of observatories** (Source: authors). *The figure is composed of two panels. The left panel displays a map of Madagascar with coloured and numbered dots corresponding to the locations of the 26 observatories. The right panel indicates, for each number, the name of the observatory and the survey years.*

The ROS remained operational until 2017. In 2015, a previously surveyed observatory was surveyed again (Menabe North-East, #14 on Figure 1), and three new observatories were initiated (Ambatofinandrahana, Anjozorobe, and Maintirano, not shown in Figure 1). Substantial work is required to clean, harmonize, and document the data from these latest observatories to include them in the dataset. The data from these new observatories will be added to the dataset at a later stage after the publication of this data descriptor.

The ROS adopted a methodology designed to combine consistency and continuous innovation. Annual surveys were conducted using questionnaires with stable parts that remained largely unchanged, ensuring data continuity on key variables over time. This approach allowed to follow-up households, capturing changes in their socioeconomic conditions and their responses to external shocks and policy changes. Over time, new topics were introduced, and existing questions were enriched to adapt to contextual changes and emerging challenges. These additions focused on key areas such as food availability and feeding practices, particularly during lean periods, exposure to natural disasters, damages, and strategies to recover after shocks, and access to basic services such as potable water, sanitation facilities, education, or healthcare. However, the survey faced the challenge of 'questionnaire inflation,' as the eagerness to add new variables without removing outdated ones significantly extended the average interview duration over the years.

Depending on the year, specific modules related to governance issues were added to the core questionnaire. The objective was to collect the population's perception of liberalization reforms, the roles of men and women, educational strategies, and various dimensions of poverty (vulnerability, participation, political voice, etc.). The ROS methodology combined quantitative methods to describe situations and a qualitative perspective to explain them, offering a comprehensive view of rural dynamics. In addition to the household questionnaire,

a community survey was conducted, involving field observations and semi-structured interviews. This mixed-method approach produces general information about the region and provides structural and situational insights on various aspects of community life (agricultural campaigns, social and cultural environment, health, education, security, development support, and product pricing).

Over the years, the ROS data has played a significant role in various academic studies, policy formulations, and development projects. The dataset includes a bibliographic database offering a broad overview of the works derived from this data. The longevity and consistent methodology make it unique among household panel survey experiences. Nevertheless, akin to all survey designs, it has encountered challenges, particularly concerning measurement and sampling, addressed through methodological improvement over time.

## Methods

In a country like Madagascar, where the statistical information system is flawed, the ROS surveys had several advantages that explain why they lasted for several years before successive periods of crisis interrupted the process. They were based on a lightweight, modular organization, using a consistent methodology to generate economies of scale. Rigorous quality control was implemented at every stage, competent supervisors were trained, and institutional capacities were strengthened. Moreover, the ROS was integrated into the national statistical information system, and the results were rapidly published and widely disseminated in the media.

### *Site selection*

The selection of observatory locations was based on a qualitative approach leveraging the expertise of individuals with extensive knowledge of the field to reflect the diverse array of rural situations. Due to the multiplicity of terrains and climates, Madagascar has diverse ecosystems for agriculture. Farming practices are shaped by the ethnocultural distribution of the population, underscoring the need for rural development policies to consider this diversity. A first subset of potential locations was prioritized to characterize a wide range of agroclimatic areas, production systems, population density, accessibility, and support structures. These potential locations were then correlated with development indicators reflecting their economic and social conditions. The observatory locations were selected from each category to constitute a varied portfolio. Each of the 26 observatories consisted of at least two sites, whether hamlet or village, strategically selected to capture the local diversity. This approach aimed to provide an in-depth understanding of the many facets of the rural landscape.

We illustrate this heterogeneity by describing the first four observatories established in 1995. The Toliara observatory (#22 on Figure 1) is located on the coastal Mahafaly Plateau in Toliara Province, an arid and isolated area with a low population density. This observatory was of particular interest due to its landlockedness problem, which is representative of a large part of rural Madagascar (territory and population). However, it is also part of a regional trading system. In addition, the coexistence of two ethnic groups, the Vezo, who rely on fishing, and the Tanalana, who are livestock farmers, is another advantage, as these two populations have different but complementary production systems. The two villages selected, Beheloka and Itampolo, also struggled with primitive living conditions, lack of fresh water, and reduced public services. The area was prone to low and irregular rainfall, leading to frequent droughts and crop failures.

Antalaha observatory (#2), located on the Northeast coast, was characterized by large-scale production of traditional export commodities: vanilla, coffee, pepper, and cloves. The observatory monitored producers' responses to the liberalization of the vanilla trade. The three villages selected, Maromandia, Tampolo, and Ambohitralalana, represent different levels of landlockedness.

The Vakinankaratra observatory (#10) in Antsirabe, located in the Madagascar central plateau, highlighted the challenges faced by family smallholdings that primarily engage in rice cultivation but experience a rice shortage. The central plateau is the most densely populated region in Madagascar, and the peasant farmers have developed a diversified cropping system to maximize the potential of the varied landscape. This observatory is characterized by a production system in which irrigated rice cultivation plays a central role. Polyculture and polyactivity are other significant features (dairy farming, fruit crops, handicrafts, charcoal-making). In addition, seasonal migratory flows, permanent settlements on the "margins of the highlands", and less densely populated areas contribute to the region's complexity. Two villages were selected to reflect the regional diversity: Ambatomena, an old settlement zone, and Vinany, a settlement established by migrants from other regions.

The Marovoay observatory (#3), in the Lower Betsiboka plain, was a significant rice-growing area and belonged to one of the large irrigated perimeters constructed at the beginning of the 20<sup>th</sup> century <sup>16</sup>. This region was mainly populated by migrants from several regions in search of salaried employment and land. A wide variety of communities could be found on the Marovoay plain. The population is young, and the main activity is farming, with almost all households having a secondary activity, often linked to farming or fishing. Farmers are highly integrated into the market economy. This region was known for its productive rice fields, and even some rice varieties were exported to Europe until the 1970s. It experienced challenges during the socialist period and the 1980s crisis, leading local farmers to adapt through retrenchment or diversification strategies. Two villages were selected in 1995: Ampijoroa and Maroala.

As shown in Figure 1, the focus in Toliara shifted to a different location at the beginning of the 2000s (#22 to #21), and the Antalaha observatory (#2) was discarded in 2004. The initial limitation to four observatories was due to resource constraints, the experimental nature of the project, and the number of observatories expanded during the 2000s, maintaining a similar kind of diversity as the one illustrated above.

Name and number on Figure 1	Outstanding features
Alaotra #5	Intermediate plateau, one of Madagascar's main rice baskets.
Ambohimahasoa #15	Tropical highland climate, relatively warm and humid, irrigated rice cultivation, fish farming, some livestock.
Ambovombe #26	Semi-arid region, frequent droughts, recurrent food insecurity, agro-pastoral system in which zebu plays a central role.
Antalaha #2	Family farming with production of rice for family consumption and cash crops (vanilla, coffee, pepper, and cloves).
Antsirabe #10	Small family farming on highlands, tropical highland climate, high.
Antsohihy #1	Diversified agriculture in a semi-humid tropical environment, large rice-growing plains but very isolated region.
Bekily #24	Significant local conflicts and diverse agriculture (rice, maize, peanuts, potatoes, vegetable gardening).

Farafangana #20	Tropical humid climate. Subsistence farming, cash crops (coffee, pepper, cloves), high cyclonic risk.
Fenerive East #4	Humid tropical zone, rice cultivation and cash crops (cloves, vanilla etc.), high cyclonic risk and fluctuation of cash crops prices on the global market.
Fianarantsoa #16	High-altitude tropical climate, rice cultivation, off-season crops, small livestock farming.
Ihosa #18	Tropical highland climate. Some industrial crops (sugar and tobacco), livestock. Affected by fires and deforestation.
Itasy #8	Central plateau, rich volcanic soils, diversified crops and large plains for rice cultivation, land tenure issues.
Mahanoro #11	Fishing, cash crops, tourism, high cyclonic risk.
Manakara #17	Hot and humid tropical climate, cyclones, particularly diverse agricultural production, fishing activities.
Manandriana #12	Tropical highland climate. Low-yield rice areas, complementary subsistence crops.
Manjakandriana #7	Diverse agriculture (rice, potatoes, cassava, taro, sweet potatoes, beans, peas) and cattle farming. Tropical highland climate.
Marovoay #3	Large irrigated rice farming area in plains, dry tropical climate.
Menabe-Belo #9	Semi-arid tropical climate with short rainy season, pastoral system and irrigated agriculture, fishing, water scarcity.
Menabe North-East #14	Tropical climate with a dry season, insecurity (cattle theft), rice cultivation, legume production for export, fish farming.
Morondava #13	Semi-arid tropical climate, affected by bushfires, deforestation, erosion, agriculture (mainly irrigated rice), livestock farming.
Tanandava #19	Former area of administered production of cotton and rice. Hot climate with short rainy season.
Tolanaro #25	Increasing maize cultivation competing with rice, significant coffee production, intermediate climatic zone.
Toliara coastal #22	Fishermen and agro pastoralists in an arid and isolated region.
Toliara North #21	Relatively developed agriculture (rice, cassava, maize, sugarcane) despite a dry climate outside the rainy season.
Tsiroanomandidy #6	Area of recent migration, young population, land issues, persistent insecurity due to armed cattle thieves.
Tsivory #23	Fertile but isolated region, periodic droughts, irrigated rice cultivation, vegetable crops.

**Table 1: List and overview of observatory outstanding features** (Source: authors).

Table 1 presents the diverse range of agricultural and climatic zones encompassed by the ROS, reflecting the country's rich ecological variety. The observatories cover areas ranging from arid and semi-arid regions with predominant activities in dry crops (maize, tubers) and pastoral farming with some irrigated rice perimeters to humid tropical zones where rice cultivation, fishing, and a mix of subsistence and cash crops like vanilla, coffee, and cloves thrive. This network captures the dynamics of different agricultural practices and climatic challenges and addresses unique regional issues such as food insecurity, land conflicts, and exposure to cyclones.

#### *A panel household survey*

For each observatory, an annual, stratified, two-stage survey of households was carried out. The survey adopted the format of a panel study, with the sample of a consistent group of households from year to year. In the initial annual survey conducted in the year of each observatory's creation, at least 500 households situated in a minimum of two sites were selected. In small localities, the sample may have encompassed all households in the locality, while in more populated places, the sample households were randomly drawn from a comprehensive enumeration of all households in the locality. The sample frame was compiled at the beginning of each annual survey to identify the households that had participated the previous year and, if applicable, ascertain the reasons behind the attrition of households (death, moving, refusal to answer). By doing so, the survey was able to account for changes in population from one year to the next, such as households relocating and new households being established. To maintain a consistent sample size of 500 households for each observatory and ensure sample representativeness, households that had relocated were substituted by new households through random selection. These new households were sourced from villages already covered by the survey if their overall population was initially not fully included, or from neighboring villages otherwise.

The unit of observation for the rural observatories is the household, not the farm, as in traditional agricultural surveys. This approach allowed for the comprehensive coverage of all activities undertaken by each individual and to establish a crucial, yet often blurred, distinction between the rural sphere and the realm of agriculture. While most rural households are engaged in activities such as crop farming, livestock husbandry, or fishing, others pursue occupations such as craftsmanship, trade, shopkeeping, or wage labor. Some households combine agricultural and non-agricultural activities. In addition, within each household, a distinction was made between the primary activity, which contributes most to income and consumes most of working time, and the secondary activities. It is common for households to participate in several income-generating activities, and their survival often hinges upon this diversification. Similarly, this approach allows for the inclusion of transfer income from family members engaged in long-term migration (over six months), but whose contributions enable the unit to continue its farming activity. Therefore, focusing on households enables the assessment of secondary activity, transfer revenues and identifies the working population holding jobs not necessarily in the agricultural sector.

Data collection relied on statements provided by the household heads and their partners, who responded to qualitative and quantitative questions. Under this system, the extent of cultivated land, the volumes of agricultural production, and the quantities sold were determined based on the knowledge and recollection of the producers. The survey was based on a comprehensive questionnaire designed to capture the multifaceted nature of the rural environment, encompassing diverse crops, livestock, and both agricultural and non-agricultural activities. The questionnaire included a consistent general section, used each year by all observatories annually, but also special sections tailored to specific observatories or themes, like vanilla cultivation or fishing, or highlighted topics like education, gender roles in household tasks, migration, etc. The general section mainly focused on socioeconomic data, including household living standards, children's education, and food security, as well as income sources, prices, quantities marketed or consumed, and agricultural production factors, offering a holistic view of rural livelihood states and strategies.

### *Team organization*

Conducting surveys across dispersed locations in Madagascar, especially in isolated areas, posed significant logistical challenges. Each observatory's collection team consisted of two to three supervisors, ten to twelve surveyors, and a driver. The primary supervisor for each



observatory, with a background in statistics and survey experience, was supported by one or two deputy supervisors knowledgeable in economics, agricultural science, or geography. Their continuous presence in the villages during survey periods was crucial for effective and reliable data collection.

To foster effective communication, some supervisors and surveyors were locally recruited at each observatory site, ensuring their familiarity with local customs, dialects, and practices. However, surveyors were never assigned to their native villages to maintain confidentiality. Given the temporary nature of their roles and the unavailability of the same team members each year, annual recruitment and training sessions were essential. These sessions, conducted with local partner organizations, included selection processes and intensive training focusing on the questionnaire objectives, teamwork, and household engagement. Supervisors were trained first, followed by surveyors who were trained by supervisors. Each training session ended with a field test survey.

Before beginning the actual survey, surveyors underwent an “integration stage” for a few days in the assigned villages. During this phase, they liaised with local administrative and traditional authorities, often by presenting previous survey results. The team would then compile or update the sample frame and, if necessary, randomly select households to replace those that were absent. This period also enabled surveyors to familiarize themselves with local measurement units, converting them into standard metrics, inventory the crops produced in the village, and acclimate the villagers to the surveyors’ presence.

### *Implementation phasing*

The process was divided into three phases spanning ten months. The preparatory phase, lasting two to three months, ideally between April and July, involved the ROS coordination team refining the general and specific sections of the questionnaire. They also updated user manuals, technical documents for data collection, and budget projections. Supervisors assembled essential kits for field surveyors containing cooking tools, camping gear, and data collection materials. A unified ten-day training session was conducted for all supervisors to ensure consistency in methodologies, concepts, and data collection practices across all observatories.

The fieldwork phase required approximately one month in each observatory, at months adapted to the local context. It entailed filling out and verifying the questionnaires through interviews with the selected households. Prior to departing for their survey sites, the recruited surveyors, approximately ten for each observatory, underwent a week-long training session. Supervisors, who were ultimately accountable for the accuracy of the completed questionnaires, accompanied them and performed first consistency tests on the data collected. This phase was frequently divided, with observatories addressed in groups, considering variations in agricultural schedules, as well as constraints related to human resources and logistics.

The data entry, validation, and analysis phase included actions implemented from June to January. Surveys were conducted in accordance with the agricultural calendar to minimize disruption to rural household livelihoods, and to visit villages during less labor-intensive seasons. Given the variation of the agricultural calendar across agro-climatic zones, surveys were conducted over an extended period. Prior to the first data collection, the production of the input mask, the program for splitting and merging data files and the consistency-testing program were completed. Data from the first observatories to be surveyed were entered before the onset of data collection in observatories with later agricultural schedules.

## Measurement harmonization

The selection of measurement units raised a methodological issue. Among the peasant population, the units used to measure surfaces, weights, and volumes are not standardized and vary from village to village. Requiring peasant farmers and fishermen to answer in standard units, i.e. kilograms, liters, or hectares, was the surest way of losing touch with reality. Hence, the ROS adopted “peasant measurements” which varied from one observatory to another depending on farming practices, and were named differently in regional languages. Output quantities are measured in units like *daba* or *vata* (oilcans), *zinga* (a kind of metal drinking vessel), *sobika* (baskets), or cartloads. Rice-growing acreage can be quantified in terms of the number of transplanters, seed quantities, *kipa* (the number of seedlings to be planted), or carts of rice bales harvested in the field. These local measurement units were identified in each village and converted into standard weights and measures. This conversion process had to be conducted separately for each village due to variations in the meanings of the same name across different locations (for example, the *cans* or *daba* hold different volumes of fluid), and the absence of standardization of “peasant measurements” between villages. After each complete interview, the survey supervisor reviewed the questionnaire and converted the gathered “peasant measurements” into standard weights and measures. This ensured the mitigation of measurement errors that are common in surveys on rural households.

## Data Records

The data collection consists of two datasets and a code repository. The rationale behind having two distinct datasets is to ensure that the ROS metadata and documentation are openly accessible, while also addressing the confidentiality concerns related to the raw data from household surveys. This raw data has been anonymized, yet it contains detailed information about household composition, incomes, and expenses, as well as individuals’ education, activities, and assets. This level of detail poses a risk of re-identifying specific individuals, especially in small rural areas. To mitigate this risk and protect the confidentiality of sensitive information, the raw data from household surveys is stored separately in its own dataset.

Name and location	Content description	Format and details
Metadata and documentation dataset	<ul style="list-style-type: none"><li>- Data catalog</li><li>- Documentation catalog</li><li>- ROS methodological material: surveyor manuals, questionnaires, and other technical documents</li><li>- Full list of publications produced from the ROS data</li><li>- Publications of the ROS project that are not accessible otherwise</li></ul>	<ul style="list-style-type: none"><li>- xlsx and tab</li><li>- txt</li><li>- pdf</li><li>- bibtex and pdf</li><li>- pdf</li></ul>
Raw household survey dataset	<ul style="list-style-type: none"><li>- Data catalog</li><li>- One folder per year with in each one:<ul style="list-style-type: none"><li>o Household composition, housing conditions, asset possession, land tenure, nutrition, food security, incomes and expenses.</li></ul></li></ul>	<ul style="list-style-type: none"><li>- xlsx and tab</li><li>- All raw data provided in two versions:<ul style="list-style-type: none"><li>o Original proprietary format (Stata)</li><li>o Open format (tsv)</li></ul></li></ul>

	<ul style="list-style-type: none"> <li>o Household member activities (with an in-depth analysis of farming practices), and finances</li> <li>o Additional annual or regional data collected in specific modules (e.g. fishing, vanilla production, natural disasters, hygiene, perinatal practices...)</li> </ul>	
Source code repository	<ul style="list-style-type: none"> <li>- Workflow to catalog the data and assess its quality, enhance its metadata and format</li> <li>- Data anonymization workflow</li> <li>- Workflow to geolocate raw data at village level based on toponyms</li> </ul>	Quarto markdown notebook with R code sections for data processing

**Table 2: Outline of the data records' content** (Source: authors). *Formats are abbreviated in the table as follows: tabulation separated values (tsv), Excel Open XML Spreadsheet (xlsx) and portable document format (pdf).*

The two datasets are hosted in the DataSuds platform. It is an instance of the Dataverse open-source research data repository software. It was created and is administered by the French research institute for sustainable development (IRD) to provide research partners in developing countries with open science infrastructures.

The source code is maintained in a GitHub repository. Authors can make further improvements and complements to this source code. Each version of the source code is versioned, allowing users to retrieve the state of the source code at the time of the validation, included in Table 2.

## Technical Validation

This section delves into the processes and methodologies employed in the ROS to ensure high data quality standards. Additionally, we provide a retrospective assessment of panel attrition, an essential aspect for longitudinal studies.

### *Data collection process*

From the inception of the data collection in 1995, measures were implemented at every stage of the survey process to ensure the reliability of the data collected. The ROS was created during a period characterized by the multiplication of observatories in developing countries, often employing multidisciplinary approaches to implement thematic monographies<sup>17</sup>. However, these initiatives often lacked reliable longitudinal data suitable for quantitative analysis. In contrast, the observatories in Madagascar were established from the outset with a clear emphasis on economic and statistical expertise, making the production of consistent data series the main priority. While researchers from various disciplines, such as geography, economics, and anthropology, contributed to the ROS, the project has always emphasized the best practices in survey management, in particular concerning sampling, quality control, and data analysis.

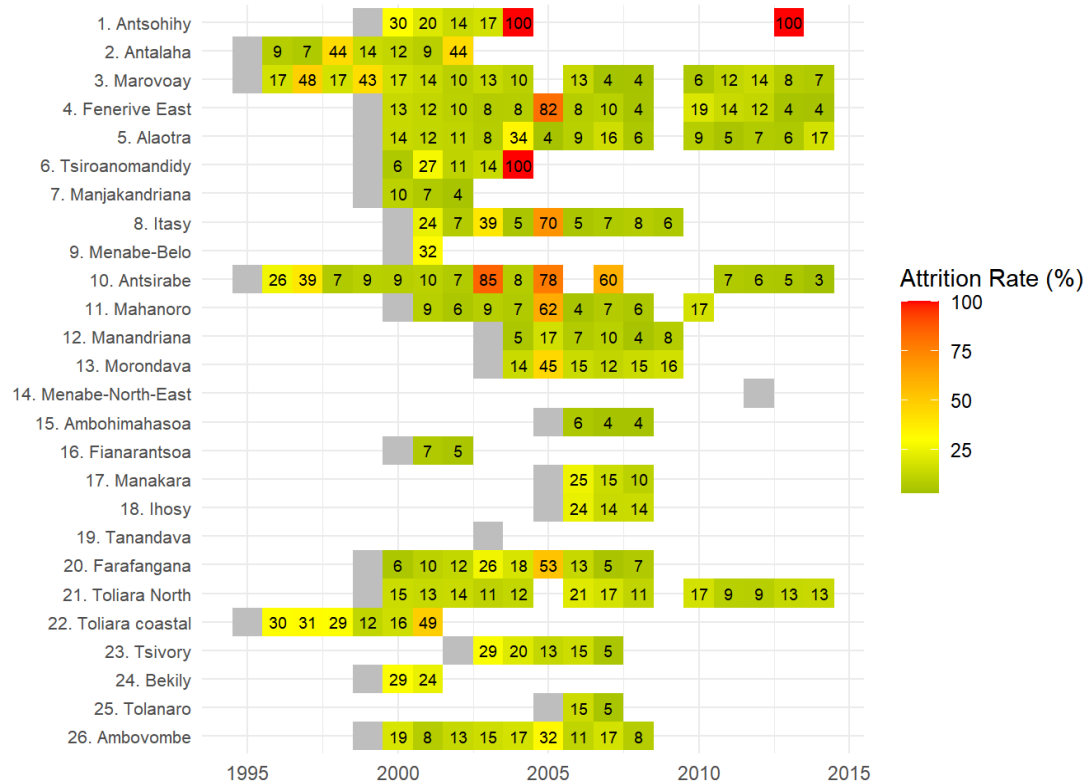
Quality control during survey data collection was paramount. Each surveyor was limited to three daily household interviews, preventing rushing and ensuring comprehensive data collection. These interviews, typically involving the householder and their partner, lasted approximately two hours. At the end of each day, surveyors returned completed

questionnaires to their supervisor, who assigned new interviews and conducted daily operations, including work schedule organization and questionnaire validation. Supervisors actively addressed errors or unclear responses and conducted consistency checks between questionnaire items. This close monitoring, facilitated by concentrating interviews in a limited area, was unique and crucial for maintaining data quality in challenging rural conditions. The continuous presence of surveyors in the villages throughout the survey periods, coupled with the active involvement of supervisors in all stages of fieldwork, from organization to data processing, characterized this network. This feature sets it apart from other rural surveys, where temporary surveyors and supervisors are often less integrated into the workflow. During the survey period, especially in the beginning and middle stages, a researcher supported the team to improve the quality of the data collection.

Additionally, validation tests were periodically conducted to ensure data accuracy. For instance, in 1996, a test compared the measured surface areas of fields with the acreage reported in interviews among a subset of households. This test tends to confirm the absence of consistent bias in interviewees' statements, indicating a balance in any inaccuracies.

### Attrition

Significant levels of attrition are reported from the ROS data<sup>18,19</sup>. It is well-known that such attrition levels are common in any longstanding panel survey<sup>20</sup>. Attrition has been extensively studied and understood by survey experts. There is a strong consensus that despite this phenomenon, panel data is useful and necessary for understanding population living standard dynamics<sup>21</sup>. The attrition rate is the percentage of households interviewed in the previous year that were not re-interviewed in the year of interest.



**Figure 2. Attrition rates in ROS panels by observatory and survey round** (Source: ROS raw data and author's calculations). *This heatmap displays attrition rates for each observatory across survey years from 1995 to 2015. The vertical axis lists the observatories, while the horizontal axis represents the years. Each square at the intersection of an observatory and a*

*survey year denotes the attrition rate, with the color gradient varying from green (0% attrition) to red (100% attrition). The attrition rate for each year is rounded and displayed within the corresponding square. The initial survey year for each observatory is marked in gray, indicating the absence of prior data for attrition calculation.*

Figure 2 depicts the attrition rate per year and per observatory. The six instances of annual attrition rates above 75% were likely induced by reshuffles of household identification codes or shifts in observatory sites. We are working to reconcile old and new identifications, and future dataset updates will include corresponding codes where possible. Excluding these outliers, the average annual attrition rate stands at 15%, aligning with developing countries' standards and resulting in a cumulative attrition rate of 80% over ten years. In 2005, an IRD team conducted a tracking survey in one village of the Marovoay observatory (#3 in Figure 1), successfully locating 59% of the attrited households<sup>19</sup>. Among these, 35% of households had split, and 65% had migrated out of the village. Other factors contributing to attrition, albeit rarely observed by ROS members, include the absence of the household during surveyor visits or the refusal to answer the survey. Nonetheless, it is important to note that attrition can introduce bias, requiring careful consideration in the analysis.

## Usage Notes

Access to the survey documentation and metadata is open, allowing any interested researcher to explore the dataset content, production process and associated materials. However, access to the raw survey data requires researchers to register, specify the purpose of data access, and commit to upholding the confidentiality rules outlined in the associated end-user license.

Given the number of variables and survey rounds, we recommend users to utilize the data catalog provided as a spreadsheet in the documentation and the raw data datasets. The data catalog contains the file name, table title, variable name, variable label and the years in which this variable is present in the dataset. The presence of labels facilitates the findings of relevant information using full-text search.

The GitHub repository associated to this paper serves as a technical appendix. It provides support materials in two formats: raw computational notebooks in markdown with R code chunks, and a web output. The web output is a HTML report featuring formatted text, embedded source code (indented and highlighted for readability), results of data analysis, and associated visualizations. This setup aims to be technically reproducible and didactically user-friendly for researchers exploring this data. The repository contains instructions to replicate all figures in the paper and includes a tutorial on georeferencing the data, useful for various types of analysis.

Users are advised to submit any questions or comments about the source code or data directly on the GitHub repository's issues page ([https://github.com/BETSAKA/Rural\\_obs\\_Madagascar/issues](https://github.com/BETSAKA/Rural_obs_Madagascar/issues)), instead of contacting the paper's corresponding author via email. This approach promotes transparency as all issues and responses are publicly accessible and can be referenced by others facing similar challenges. It also helps build a collaborative knowledge base that benefits the user community.

## Code Availability

The data anonymization, data validation, and guidance for data georeferencing and analysis are available at [https://github.com/BETSAKA/Rural\\_obs\\_Madagascar](https://github.com/BETSAKA/Rural_obs_Madagascar), under the MIT open-source license. This work used *R* 4.3.1 for the general data processing environment<sup>22</sup>, alongside the following packages: *tidyverse* 2.0.0 for data preparation and visualization<sup>23</sup>, *haven* 2.5.3 for reading and converting files from Stata format<sup>24</sup>, *labelled* 2.12.0 for generating the data catalog<sup>25</sup>, *sf* 1.0.14 for spatial data georeferencing<sup>26</sup>, *stringdist* 0.9.10 for toponym matching and data pseudonymization<sup>27</sup>, and *tmap* 3.3.4 for spatial data visualization<sup>28</sup>.

## Acknowledgements

Besides the authors of this paper, the ROS endeavor was made possible by the help and dedication of numerous colleagues at IRD, INSTAT, the University of Antananarivo and other institutions.

The ROS has been financially funded by a variety of donors and NGOs during the 1995-2015 period: European Union, French cooperation, AFD, GTZ, World Bank, Swiss cooperation, Care, and Rio Tinto, and IFAD.

Efforts to complete the preparation of ROS microdata and documentation for publication, along with the elaboration of this data descriptor, were supported by the BETSAKA project (cofunded by the KfW development Bank and the French development agency AFD via the PAIRES Program) and the DT3Ma Makay Observatory project (funded by AFD).

The gold open access fee for this paper was supported by the Development Impact Lab of KfW Development Bank.

The support from the above-mentioned organizations does not imply their endorsement of the interpretations presented herein and they bear no responsibility for the content of this publication or for the associated dataset.

## Author contributions

VSA: data curation and validation.

NA: ROS administration and supervision, investigation, data curation, and validation.

FB: formal analysis, software (geolocation, anonymization and data formatting) visualization and original draft writing of the data descriptor.

ID: ROS conceptualization and methodology design, funding acquisition, ROS administration, supervision, investigation, and validation.

JLD: ROS organization and investigation.

JdM: data curation, original draft writing of the data descriptor.

BNR: ROS administration, supervision and data curation.

RRafi: data curation and validation.

PR: funding acquisition, ROS administration, supervision, investigation and validation.

RRato: ROS administration, supervision and investigation.

LAOR: ROS administration and supervision, investigation, data curation and validation.

JDR: ROS administration and supervision, software (data entry setup), data curation and validation. VR: ROS administration and supervision, investigation, data curation and validation.

MR: methodology and investigation.

BR: funding acquisition, ROS administration, supervision, investigation, and validation.

FR: ROS conceptualization and methodology design, funding acquisition, ROS administration, supervision, investigation, and validation.

CSM: data curation and validation.

All authors revised and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## References

1. Castañeda, A. *et al.* A New Profile of the Global Poor. *World Development* **101**, 250–267 (2018).
2. *World Development Report 2008: Agriculture for Development*. (World Bank, Washington DC).
3. Jerven, M. Random Growth in Africa? Lessons from an Evaluation of the Growth Evidence on Botswana, Kenya, Tanzania and Zambia, 1965–1995. *Journal of Development Studies* **46**, 274–294 (2010).
4. Sandefur, J. & Glassman, A. L. The Political Economy of Bad Data: Evidence from African Survey & Administrative Statistics. *Available at SSRN 2466028* (2014).
5. Randall, S. & Coast, E. Poverty in African households: the limits of survey and census representations. *The Journal of Development Studies* **51**, 162–177 (2015).
6. Östberg, W., Howland, O., Mduma, J. & Brockington, D. Tracing improving livelihoods in rural Africa using local measures of wealth: A case study from central Tanzania, 1991–2016. *Land* **7**, 44 (2018).
7. Ansoms, A., Marijnen, E., Cioffo, G. & Murison, J. Statistics versus livelihoods: questioning Rwanda’s pathway out of poverty. *Review of African Political Economy* **44**, 47–65 (2017).
8. INSTAT. *Résultats Globaux Du Troisième Recensement Général de La Population et de l’habitation de 2018 à Madagascar (RGPH-3)*. (Institut national de la statistique, Antananarivo, 2020).
9. Roubaud, F. La question rizicole à Madagascar: les résultats d’une décennie de libéralisation. *Revue Economie de Madagascar* 37–62 (1997).
10. Roubaud, F. Le projet Madio à Madagascar : objectifs, démarche, résultats. *Statéco* 5–24 (2000).
11. Razafindrakoto, M. & Roubaud, F. Two original poverty monitoring tools: the 1-2-3 surveys and the Rural Observatories. in *New international poverty reduction strategies* 335–361 (Routledge, 2003).
12. Droy, I., Ratovoarinony, R. & Roubaud, F. Rural observatories in Madagascar: An innovative mechanism for monitoring the countryside. *Inter-Stat* 65–91 (2004).
13. Grislain, Q. Promesses et effets des observatoires fonciers: une géographie politique de dispositifs sociotechniques pour le développement en contextes africains et sénégalais (1980/2022). (Paris 1, 2022).
14. Dubois, J.-L. & Droy, I. L’observatoire: un instrument pour le suivi de la pauvreté. in *Les outils de mesure et de suivi/évaluation des politiques de sécurité alimentaires et de lutte contre la pauvreté* 22–46 (AEDES-IRAM, Brussels, 2002).
15. Saint-Macary, C. Le suivi des dynamiques de pauvreté en milieu rural: retour d’expérience des observatoires ruraux à Madagascar. *Statéco* 63–67 (2018).
16. Droy, I. Que sont les greniers à riz devenus ? (Madagascar). *Autrepart* **7**, 89–110 (1998).
17. *Observatoires du développement, observatoires pour le développement: actes du séminaire, septembre 1994*. (Éd. de l’Orstom, Paris, France, 1998).
18. Gubert, F. & Robilliard, A.-S. Risk and Schooling Decisions in Rural Madagascar: A Panel Data-Analysis. *Journal of African Economies* **17**, 207–238 (2008).

- 570 19. Vaillant, J. Attrition and Follow-Up Rules in Panel Surveys: Insights from a Tracking  
571 Experience in Madagascar. *Review of Income and Wealth* **59**, 509–538 (2013).
- 572 20. Alderman, H., Das, J. & Rao, V. *Conducting Ethical Economic Research Complications from*  
573 *the Field*. (The World Bank, 2013). doi:10.1596/1813-9450-6446.
- 574 21. Ashenfelter, O., Deaton, A. & Solon, G. *Collecting Panel Data in Developing Countries: Does*  
575 *It Make Sense?* (World Bank, Washington DC, 1985).
- 576 22. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for  
577 Statistical Computing, Vienna, Austria, 2023).
- 578 23. Wickham, H. *et al.* Welcome to the tidyverse. *Journal of Open Source Software* **4**, 1686  
579 (2019).
- 580 24. Wickham, H., Miller, E. & Smith, D. *Haven: Import and Export ‘SPSS’, ‘Stata’ and ‘SAS’ Files*.  
581 (2023).
- 582 25. Larmarange, J. *Labelled: Manipulating Labelled Data*. (2023).
- 583 26. Pebesma, E. Simple Features for R: Standardized Support for Spatial Vector Data. *The R*  
584 *Journal* **10**, 439–446 (2018).
- 585 27. van der Loo, M. P. The stringdist package for approximate string matching. *The R Journal*  
586 **6**, 111–122 (2014).
- 587 28. Tennekes, M. tmap: Thematic Maps in R. *Journal of Statistical Software* **84**, 1–39 (2018).