



**HAL**  
open science

## **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 (v2), last revised 10 Oct 2024 (v4)

**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%

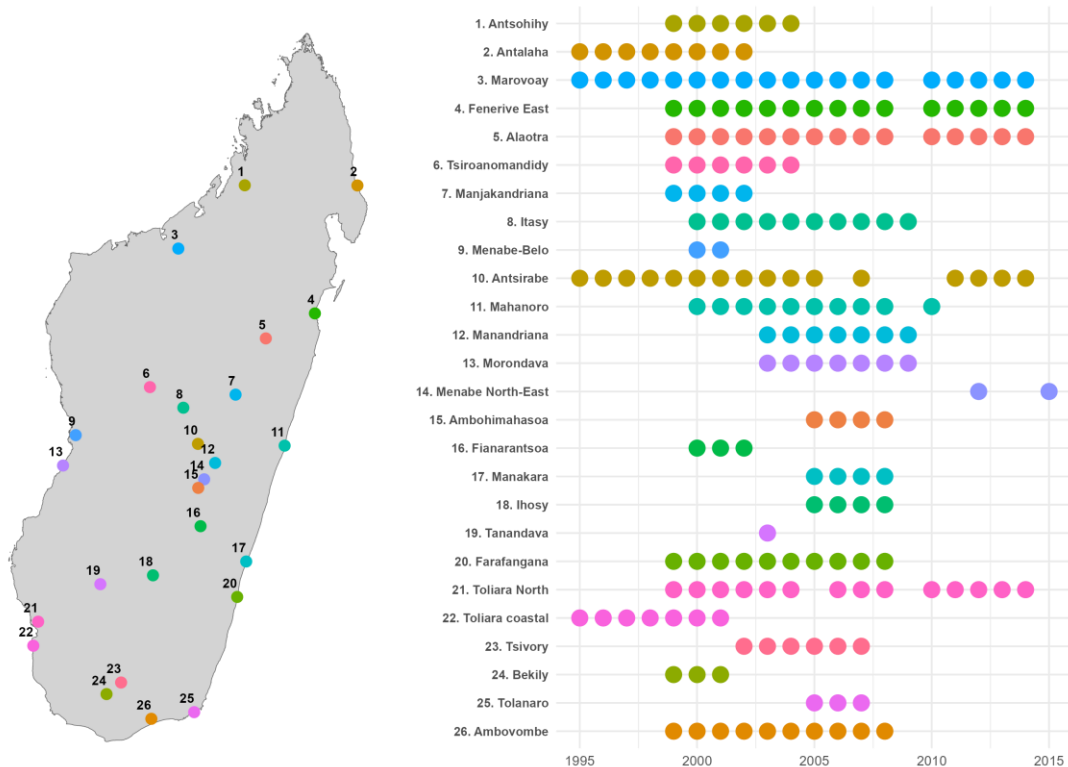
52 of the population living in rural areas<sup>8</sup>. In the mid-1990s, Madagascar's rural areas underwent  
53 profound transformations. After a socialist period from 1975 to 1985, the country moved  
54 towards economic liberalization following the IMF recommendations: liberalization of pricing  
55 and marketing structures, and a shift from state-driven policies to market-oriented reforms<sup>9</sup>.  
56 The existing statistical system, which had deteriorated during the socialist era, was ill-  
57 equipped to capture these changes, leaving a gap in understanding the rural landscape  
58 evolution<sup>10</sup>. This situation was particularly concerning given that agriculture employed almost  
59 80% of the active population, and that any analysis of the Malagasy economy would be  
60 incomplete without a comprehensive understanding of its rural sector.

61  
62 This lack of information motivated the creation of four rural observatories in Madagascar in  
63 1995<sup>11,12</sup>, each representing distinct agricultural challenges within the country's diverse  
64 ecosystems. The observatories consisted of annual panel survey, dedicated to the continuous  
65 collection, analysis and dissemination of socioeconomic data and insights, with the aim of  
66 identifying characteristics and dynamics across diverse agroecological contexts to guide the  
67 formulation, implementation, and evaluation of rural development policies<sup>13</sup>.

68  
69 As the ROS emerged as an innovative and cost-effective way to provide a nuanced and  
70 continuous observation of rural households, its extension was supported by new funders and  
71 partners, such as research institutions, non-governmental organizations or consultancy firms<sup>14</sup>.  
72 The institutional anchoring of the rural observatories changed in 2000, moving from the  
73 National Statistical Office, with which it was initially associated, to the Rural Development  
74 Policy Unit, within the Ministry of Agriculture. A charter was drawn up to promote the  
75 methodological unity of the observatories and the ethics of intervention. Each partner  
76 involved was required to sign it to ensure methodological consistency across observatories  
77 and over the years.

78  
79 Over the years, the number of observatories increased, reaching 17 in 2004, before  
80 contracting to five following the political crisis of 2009<sup>15</sup>. Despite these challenges, the ROS  
81 persisted, with three of the original observatories remaining to operate actively until 2014, as  
82 shown in Figure 1.

83



84  
 85 **Figure 1: Coarse location and years of data collection of observatories** (Source: authors). *The*  
 86 *figure is composed of two panels. The left panel displays a map of Madagascar with coloured*  
 87 *and numbered dots corresponding to the locations of the 26 observatories. The right panel*  
 88 *indicates, for each number, the name of the observatory and the survey years.*

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

95  
 96 The ROS adopted a methodology designed to combine consistency and continuous innovation.  
 97 Annual surveys were conducted using questionnaires with stable parts that remained largely  
 98 unchanged, ensuring data continuity on key variables over time. This approach allowed to  
 99 follow-up households, capturing changes in their socioeconomic conditions and their  
 100 responses to external shocks and policy changes. Over time, new topics were introduced, and  
 101 existing questions were enriched to adapt to contextual changes and emerging challenges.  
 102 These additions focused on key areas such as food availability and feeding practices,  
 103 particularly during lean periods, exposure to natural disasters, damages, and strategies to  
 104 recover after shocks, and access to basic services such as potable water, sanitation facilities,  
 105 education, or healthcare. However, the survey faced the challenge of 'questionnaire inflation,'  
 106 as the eagerness to add new variables without removing outdated ones significantly extended  
 107 the average interview duration over the years.

108  
 109 Depending on the year, specific modules related to governance issues were added to the core  
 110 questionnaire. The objective was to collect the population's perception of liberalization  
 111 reforms, the roles of men and women, educational strategies, and various dimensions of  
 112 poverty (vulnerability, participation, political voice, etc.). The ROS methodology combined  
 113 quantitative methods to describe situations and a qualitative perspective to explain them,  
 114 offering a comprehensive view of rural dynamics. In addition to the household questionnaire,

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

120  
121 Over the years, the ROS data has played a significant role in various academic studies, policy  
122 formulations, and development projects. The dataset includes a bibliographic database  
123 offering a broad overview of the works derived from this data. The longevity and consistent  
124 methodology make it unique among household panel survey experiences. Nevertheless, akin  
125 to all survey designs, it has encountered challenges, particularly concerning measurement and  
126 sampling, addressed through methodological improvement over time.

127

## 128 **Methods**

129

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

138

### 139 *Site selection*

140

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

154

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

166

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

172  
 173 The Vakinankaratra observatory (#10) in Antsirabe, located in the Madagascar central plateau,  
 174 highlighted the challenges faced by family smallholdings that primarily engage in rice  
 175 cultivation but experience a rice shortage. The central plateau is the most densely populated  
 176 region in Madagascar, and the peasant farmers have developed a diversified cropping system  
 177 to maximize the potential of the varied landscape. This observatory is characterized by a  
 178 production system in which irrigated rice cultivation plays a central role. Polyculture and  
 179 polyactivity are other significant features (dairy farming, fruit crops, handicrafts, charcoal-  
 180 making). In addition, seasonal migratory flows, permanent settlements on the "margins of the  
 181 highlands", and less densely populated areas contribute to the region's complexity. Two  
 182 villages were selected to reflect the regional diversity: Ambatomena, an old settlement zone,  
 183 and Vinany, a settlement established by migrants from other regions.

184  
 185 The Marovoay observatory (#3), in the Lower Betsiboka plain, was a significant rice-growing  
 186 area and belonged to one of the large irrigated perimeters constructed at the beginning of the  
 187 20<sup>th</sup> century <sup>16</sup>. This region was mainly populated by migrants from several regions in search  
 188 of salaried employment and land. A wide variety of communities could be found on the  
 189 Marovoay plain. The population is young, and the main activity is farming, with almost all  
 190 households having a secondary activity, often linked to farming or fishing. Farmers are highly  
 191 integrated into the market economy. This region was known for its productive rice fields, and  
 192 even some rice varieties were exported to Europe until the 1970s. It experienced challenges  
 193 during the socialist period and the 1980s crisis, leading local farmers to adapt through  
 194 retrenchment or diversification strategies. Two villages were selected in 1995: Ampijoroa and  
 195 Maroala.

196  
 197 As shown in Figure 1, the focus in Toliara shifted to a different location at the beginning of the  
 198 2000s (#22 to #21), and the Antalaha observatory (#2) was discarded in 2004. The initial  
 199 limitation to four observatories was due to resource constraints, the experimental nature of  
 200 the project, and the number of observatories expanded during the 2000s, maintaining a similar  
 201 kind of diversity as the one illustrated above.

202

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.
Ihoso #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.

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

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

212

213 *A panel household survey*

214

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

231  
232 The unit of observation for the rural observatories is the household, not the farm, as in  
233 traditional agricultural surveys. This approach allowed for the comprehensive coverage of all  
234 activities undertaken by each individual and to establish a crucial, yet often blurred, distinction  
235 between the rural sphere and the realm of agriculture. While most rural households are  
236 engaged in activities such as crop farming, livestock husbandry, or fishing, others pursue  
237 occupations such as craftsmanship, trade, shopkeeping, or wage labor. Some households  
238 combine agricultural and non-agricultural activities. In addition, within each household, a  
239 distinction was made between the primary activity, which contributes most to income and  
240 consumes most of working time, and the secondary activities. It is common for households to  
241 participate in several income-generating activities, and their survival often hinges upon this  
242 diversification. Similarly, this approach allows for the inclusion of transfer income from family  
243 members engaged in long-term migration (over six months), but whose contributions enable  
244 the unit to continue its farming activity. Therefore, focusing on households enables the  
245 assessment of secondary activity, transfer revenues and identifies the working population  
246 holding jobs not necessarily in the agricultural sector.

247  
248 Data collection relied on statements provided by the household heads and their partners, who  
249 responded to qualitative and quantitative questions. Under this system, the extent of  
250 cultivated land, the volumes of agricultural production, and the quantities sold were  
251 determined based on the knowledge and recollection of the producers. The survey was based  
252 on a comprehensive questionnaire designed to capture the multifaceted nature of the rural  
253 environment, encompassing diverse crops, livestock, and both agricultural and non-  
254 agricultural activities. The questionnaire included a consistent general section, used each year  
255 by all observatories annually, but also special sections tailored to specific observatories or  
256 themes, like vanilla cultivation or fishing, or highlighted topics like education, gender roles in  
257 household tasks, migration, etc. The general section mainly focused on socioeconomic data,  
258 including household living standards, children's education, and food security, as well as  
259 income sources, prices, quantities marketed or consumed, and agricultural production factors,  
260 offering a holistic view of rural livelihood states and strategies.

261

### 262 *Team organization*

263

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



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

271

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

281

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

289

### 290 *Implementation phasing*

291

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

300

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

310

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

319

320 *Measurement harmonization*

321

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

340

341 **Data Records**

342

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

351

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

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

355  
356 The two datasets are hosted in the DataSuds platform. It is an instance of the Dataverse open-  
357 source research data repository software. It was created and is administered by the French  
358 research institute for sustainable development (IRD) to provide research partners in  
359 developing countries with open science infrastructures.

360  
361 The source code is maintained in a GitHub repository. Authors can make further  
362 improvements and complements to this source code. Each version of the source code is  
363 versioned, allowing users to retrieve the state of the source code at the time of the validation,  
364 included in Table 2.

## 365 366 **Technical Validation**

367  
368 This section delves into the processes and methodologies employed in the ROS to ensure high  
369 data quality standards. Additionally, we provide a retrospective assessment of panel attrition,  
370 an essential aspect for longitudinal studies.

### 371 372 *Data collection process*

373  
374 From the inception of the data collection in 1995, measures were implemented at every stage  
375 of the survey process to ensure the reliability of the data collected. The ROS was created  
376 during a period characterized by the multiplication of observatories in developing countries,  
377 often employing multidisciplinary approaches to implement thematic monographies<sup>17</sup>.  
378 However, these initiatives often lacked reliable longitudinal data suitable for quantitative  
379 analysis. In contrast, the observatories in Madagascar were established from the outset with  
380 a clear emphasis on economic and statistical expertise, making the production of consistent  
381 data series the main priority. While researchers from various disciplines, such as geography,  
382 economics, and anthropology, contributed to the ROS, the project has always emphasized the  
383 best practices in survey management, in particular concerning sampling, quality control, and  
384 data analysis.

385  
386 Quality control during survey data collection was paramount. Each surveyor was limited to  
387 three daily household interviews, preventing rushing and ensuring comprehensive data  
388 collection. These interviews, typically involving the householder and their partner, lasted  
389 approximately two hours. At the end of each day, surveyors returned completed

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

401

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

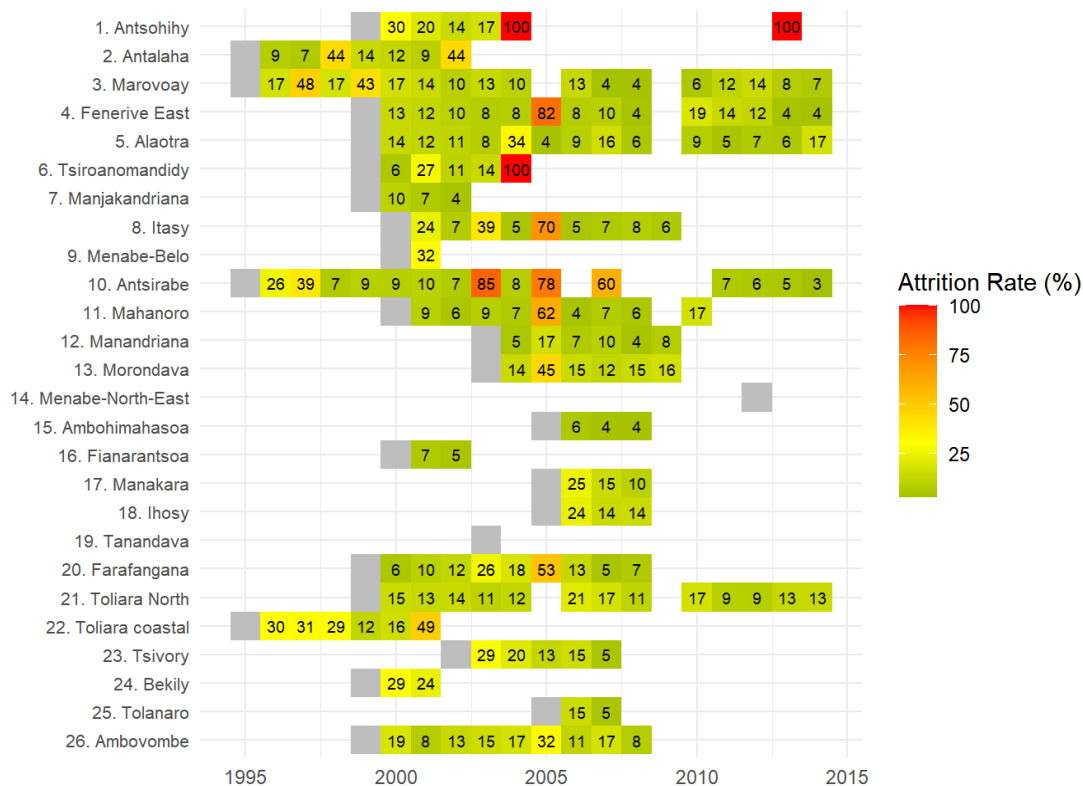
406

### 407 *Attrition*

408

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

415



416

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

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

425

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

438

439

## 440 **Usage Notes**

441

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

446

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

452

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

461

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

468

## 469 **Code Availability**

470

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

478

479

## 480 **Acknowledgements**

481

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

485

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

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

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

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

498

## 499 **Author contributions**

500

501 VSA: data curation and validation.

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

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

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

507 JLD: ROS organization and investigation.

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

509 BNR: ROS administration, supervision and data curation.

510 RRafi: data curation and validation.

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

512 RRato: ROS administration, supervision and investigation.

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

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

516 MR: methodology and investigation.

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

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

520 CSM: data curation and validation.

521 All authors revised and approved the final manuscript.

522

## 523 **Competing interests**

524

525 The authors declare no competing interests.

526

## 527 **References**

528

- 529 1. Castañeda, A. *et al.* A New Profile of the Global Poor. *World Development* **101**, 250–267  
530 (2018).
- 531 2. *World Development Report 2008: Agriculture for Development*. (World Bank,  
532 Washington DC).
- 533 3. Jerven, M. Random Growth in Africa? Lessons from an Evaluation of the Growth Evidence  
534 on Botswana, Kenya, Tanzania and Zambia, 1965–1995. *Journal of Development Studies*  
535 **46**, 274–294 (2010).
- 536 4. Sandefur, J. & Glassman, A. L. The Political Economy of Bad Data: Evidence from African  
537 Survey & Administrative Statistics. *Available at SSRN 2466028* (2014).
- 538 5. Randall, S. & Coast, E. Poverty in African households: the limits of survey and census  
539 representations. *The Journal of Development Studies* **51**, 162–177 (2015).
- 540 6. Östberg, W., Howland, O., Mduma, J. & Brockington, D. Tracing improving livelihoods in  
541 rural Africa using local measures of wealth: A case study from central Tanzania, 1991–  
542 2016. *Land* **7**, 44 (2018).
- 543 7. Ansoms, A., Marijnen, E., Cioffo, G. & Murison, J. Statistics versus livelihoods: questioning  
544 Rwanda’s pathway out of poverty. *Review of African Political Economy* **44**, 47–65 (2017).
- 545 8. INSTAT. *Résultats Globaux Du Troisième Recensement Général de La Population et de*  
546 *l’habitation de 2018 à Madagascar (RGPH-3)*. (Institut national de la statistique,  
547 Antananarivo, 2020).
- 548 9. Roubaud, F. La question rizicole à Madagascar: les résultats d’une décennie de  
549 libéralisation. *Revue Economie de Madagascar* 37–62 (1997).
- 550 10. Roubaud, F. Le projet Madio à Madagascar : objectifs, démarche, résultats. *Statéco* 5–24  
551 (2000).
- 552 11. Razafindrakoto, M. & Roubaud, F. Two original poverty monitoring tools: the 1-2-3 surveys  
553 and the Rural Observatories. in *New international poverty reduction strategies* 335–361  
554 (Routledge, 2003).
- 555 12. Droy, I., Ratovoarinony, R. & Roubaud, F. Rural observatories in Madagascar: An  
556 innovative mechanism for monitoring the countryside. *Inter-Stat* 65–91 (2004).
- 557 13. Grislain, Q. Promesses et effets des observatoires fonciers: une géographie politique de  
558 dispositifs sociotechniques pour le développement en contextes africains et sénégalais  
559 (1980/2022). (Paris 1, 2022).
- 560 14. Dubois, J.-L. & Droy, I. L’observatoire: un instrument pour le suivi de la pauvreté. in *Les*  
561 *outils de mesure et de suivi/évaluation des politiques de sécurité alimentaires et de lutte*  
562 *contre la pauvreté* 22–46 (AEDES-IRAM, Brussels, 2002).
- 563 15. Saint-Macary, C. Le suivi des dynamiques de pauvreté en milieu rural: retour d’expérience  
564 des observatoires ruraux à Madagascar. *Statéco* 63–67 (2018).
- 565 16. Droy, I. Que sont les greniers à riz devenus ? (Madagascar). *Autrepart* **7**, 89–110 (1998).
- 566 17. *Observatoires du développement, observatoires pour le développement: actes du*  
567 *séminaire, septembre 1994*. (Éd. de l’Orstom, Paris, France, 1998).
- 568 18. Gubert, F. & Robilliard, A.-S. Risk and Schooling Decisions in Rural Madagascar: A Panel  
569 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).