



**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-04502779v1

**HAL Id: hal-04502779**

**<https://hal.science/hal-04502779v1>**

Preprint submitted on 13 Mar 2024 (v1), 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.



Distributed under a Creative Commons Attribution 4.0 International License

# Data Descriptor Template

## Title

A rural observatory system surveyed household living conditions in Madagascar each year from 1995 to 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>1,3</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 socio-economic data on rural areas. It collected, analyzed, and disseminated data to help in the formulation and evaluation of 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, complemented by modules covering new themes and enriched questions adapted to specific contexts. Qualitative community surveys provided a qualitative understanding of the local features and dynamics. The site selection combined quantitative and qualitative insights to reflect Madagascar's diverse rural challenges. Quality control was comprehensive, with measures like limiting the number of surveyor interviews per day 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 its 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 major proportion of the population in least developed countries (64% in 2022) and they make up the bulk of worldwide poverty<sup>1</sup>. Recent reductions in poverty rates in developing countries stem from a decline in the poverty rates among the rural population<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

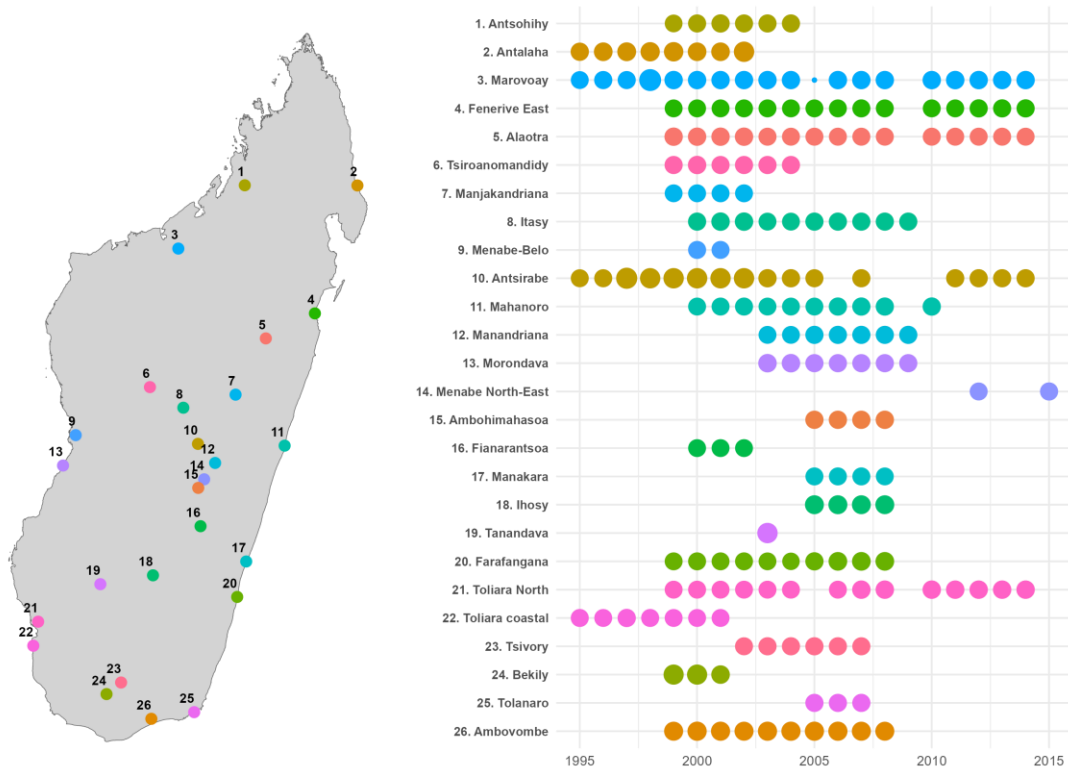
52 million in 1993 to 26 million in 2018 with 81% of the population living in rural areas<sup>8</sup>. In the  
53 mid-1990s, Madagascar's rural areas underwent profound transformations. After a socialist  
54 period from 1975 to 1985, the country embarked on a policy of economic liberalization under  
55 the IMF's recommendations: liberalization of pricing and marketing structures, and shift from  
56 state-driven policies to market-oriented reforms<sup>9</sup>. The existing statistical system, which had  
57 deteriorated during the socialist era, was ill-equipped to capture these changes, leaving a gap  
58 in understanding the evolution of the rural landscape<sup>10</sup>. This situation was particularly  
59 concerning given that agriculture employed almost 80% of the active population, and that any  
60 analysis of the Malagasy economy would be incomplete without a comprehensive  
61 understanding of its rural sector.

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

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

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

84



85  
 86 **Figure 1: Coarse location and years of data collection of observatories composing the ROS**  
 87 *(source: authors). The figure is composed of two panels. The left panel shows a map of*  
 88 *Madagascar with coloured and numbered dots corresponding to the locations of the 26*  
 89 *observatories. The right panel indicates, for each number, the name of the observatory and the*  
 90 *years in which survey data was collected.*

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

98  
 99 The ROS adopted a methodology designed to combine consistency and continuous innovation.  
 100 Annual surveys were conducted using questionnaires with stable parts that remained largely  
 101 unchanged, ensuring the continuity of data on key variables over the years. This approach  
 102 allowed for the tracking of households, capturing changes in their socio-economic conditions  
 103 and their reactions to external shocks and policy changes. Over time, new themes were also  
 104 introduced, and existing questions were enriched to adapt to contextual changes and  
 105 emerging challenges. These additions addressed key areas such as food availability and food  
 106 practices, especially during the lean period, exposure to natural disasters, damages, and  
 107 strategies to recover after shocks, and access to basic services such as potable water,  
 108 sanitation facilities, education, or health services. The survey, however, faced a challenge of  
 109 'questionnaire inflation,' as the eagerness to add new variables without removing outdated  
 110 ones significantly extended the average duration of interviews as years passed.

111  
 112 Depending on the year, different subject-specific modules related to governance issues were  
 113 also appended to the basic questionnaire. The objective was to collect the population's points  
 114 of view on liberalization reforms, the respective roles of men and women, educational  
 115 strategies, etc.), and on the multiple dimensions of poverty (vulnerability, participation,

116 political voice, etc.). The ROS approach combined quantitative methods to describe situations  
117 and a qualitative perspective to explain them, offering a comprehensive view of rural  
118 dynamics. Complementing the household questionnaire, a community survey was conducted,  
119 involving field observations and semi-structured interviews. This quantitative and qualitative  
120 mixed approach provides general information about the region, structural and situational  
121 insights on various aspects of community life, including agricultural campaigns, social and  
122 cultural environment, health, education, security, development support, and product pricing.

123

124 Over the years, the ROS data has played a significant role in various academic studies, policy  
125 formulations, and development projects. The dataset includes a bibliographic database  
126 providing a broad overview of the works produced from this data. The longevity and consistent  
127 methodology of ROS make it unique among household panel survey experiences. However,  
128 like all survey designs, it has faced challenges, particularly in terms of measurement and  
129 sampling, which were addressed through methodological improvement over time.

130

## 131 **Methods**

132

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

141

### 142 *Site selection*

143

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

157

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

167 public services. The area was prone to low and irregular rainfall, leading to frequent droughts  
168 and crop failures.

169

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

175

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

187

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

199

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

205

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/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.

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

207 The Table 1 shows the diverse range of agricultural and climatic zones encompassed by the  
208 ROS, reflecting the country's rich ecological variety. The observatories cover areas ranging  
209 from arid and semi-arid regions with predominant activities in dry crops (maize, tubers) and  
210 pastoral farming with some rice irrigated rice perimeters to humid tropical zones where rice  
211 cultivation, fishing, and a mix of subsistence and cash crops like vanilla, coffee, and cloves  
212 thrive. This network not only captures the dynamics of different agricultural practices and  
213 climatic challenges but also addresses unique regional issues such as food insecurity, land  
214 conflicts, and exposure to cyclones.  
215

216 *A panel household survey*

217

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

235

236 The unit of observation for the rural observatories is the household, not the farm as in  
237 traditional agricultural surveys. This approach allowed for the comprehensive coverage of all  
238 activities undertaken by each individual and to establish a crucial, yet often blurred,  
239 differentiation between the rural sphere and the realm of agriculture. While the vast majority  
240 of rural households are engaged in activities such as crop farming, livestock husbandry, or  
241 fishing, there are others who pursue occupations such as craftsmanship, trade, shopkeeping,  
242 or wage labor. Some households combine agricultural and non-agricultural activities. In  
243 addition, within each household, a distinction was made between the primary activity, which  
244 is the activity that the household considers to contribute most of its income, consume most  
245 of its working time or what the respondent values most, and the secondary activities. It is quite  
246 common for households to participate in several income-generating activities, and the survival  
247 of rural households often hinges upon this diversification. Similarly, this allows the inclusion  
248 of transfer income from family members who are away on longer-term migration (over six  
249 months), but whose contribution enables the unit to continue its farming activity. Therefore,  
250 focusing on households serves as a means to assess the level of secondary activity and transfer  
251 revenues and identify those members of the working population who hold these jobs, which  
252 are not necessarily in the agricultural sector.

253

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

267



268 *Team organization*

269

270 Conducting surveys across dispersed locations in Madagascar, especially in isolated areas,  
271 presented significant logistical challenges. Each observatory's collection team consisted of  
272 two to three supervisors, ten to twelve surveyors, and a driver. The primary supervisor for  
273 each observatory, with a background in statistics and survey experience, was supported by  
274 one or two deputy supervisors knowledgeable in economics, agricultural science, or  
275 geography. Their continuous presence in the villages during survey periods was crucial for  
276 effective and reliable data collection.

277

278 To foster effective communication, some of the supervisors and surveyors were recruited  
279 locally at each observatory site, ensuring that they were familiar with local customs, dialects,  
280 and practices. However, surveyors were never assigned to their native villages to preserve  
281 confidentiality. Given the temporary nature of their roles and the unavailability of the same  
282 team members each year, annual recruitment and training sessions were essential. These  
283 sessions, conducted with local partner organizations, included selection processes and  
284 intensive training focusing on the objectives of the questionnaire, teamwork, and household  
285 engagement. Supervisors were trained first, and then supervisors trained the surveyors. Each  
286 training session ended with a test survey in the field.

287

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

295

296 *Implementation phasing*

297

298 The process was divided into three phases, lasting a total of ten months. The preparatory  
299 phase, which lasted for two to three months, ideally between April and July, involved the ROS  
300 coordination team developing and refining both the general and specific sections of the  
301 questionnaire. They also updated user manuals, technical documents for data collection, and  
302 budget projections. Supervisors assembled essential kits for field surveyors, which included  
303 cooking tools, camping gear, and data collection materials. A unified ten-day training session  
304 was conducted for all supervisors to ensure consistency of methodologies, concepts, and data  
305 collection practices across all observatories.

306

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

316

317 The data entry, validation, and analysis phase included actions implemented at different  
318 moments of the year, from June to January. Surveys were conducted in accordance with the  
319 agricultural calendar in order to minimize disruption to rural household livelihoods, and to visit

320 villages during less labor-intensive seasons. As the agricultural calendar varies by agro-climatic  
 321 zone, surveys were conducted over an extended period of time. Prior to the first data  
 322 collection, the production of the input mask, the program for splitting and merging data files  
 323 and the consistency-testing program were completed. Data from the first observatories to be  
 324 surveyed were entered before the onset of the data collection phase of the observatories with  
 325 a later agricultural schedule.

### 327 *Measurement harmonization*

328  
 329 The selection of measurement units raised a methodological issue. Among the peasant  
 330 population, the units used to measure surfaces, weights, and volumes are not standardized  
 331 and vary from one village to another. Requiring peasant farmers and fishermen to give their  
 332 answers in standard units, i.e. kilograms, liters or hectares, was the surest way of losing touch  
 333 with reality. Therefore, the ROS adopted “peasant measurements” that vary from one  
 334 observatory to another depending on farming practices, and whose names also vary  
 335 depending on regional languages. Output quantities are measured in *daba* or *vata* (oilcans),  
 336 in *zinga* (a kind of metal drinking vessel), in *sobika* (baskets), or in cartloads. Rice-growing  
 337 acreage can be quantified in the number of transplinters, by seed quantities and by *kipa* (the  
 338 number of seedlings to be planted out) or number of carts of rice bales harvested in the field.  
 339 In each village, these local measurement units were identified and converted into standard  
 340 weights and measures. This conversion process had to be conducted separately for each  
 341 village, due to variation in the meanings of the same name across different locations (for  
 342 example, the *cans* or *daba* hold different volumes of fluid), and the absence of standardization  
 343 of “peasant measurements” between villages. Following each complete interview, the  
 344 questionnaire was reviewed by the survey supervisor, who converted the gathered “peasant  
 345 measurements” into standard weights and measures. This ensured overcoming measurement  
 346 errors that are common in surveys on rural households.

### 348 **Data Records**

349  
 350 The data collection consists of two datasets and a code repository. The reason for having two  
 351 distinct datasets is to ensure that the ROS metadata and documentation are freely accessible,  
 352 while also addressing the confidentiality concerns associated with the raw data from  
 353 household surveys. This raw data has been anonymized, but due to the level of detail it  
 354 contains about household composition, incomes, expenses, as well as individuals' education,  
 355 activities, and assets, there is a risk of re-identifying certain individuals, especially in small rural  
 356 areas. To safeguard the confidentiality of such sensitive information, the raw data from  
 357 household surveys is stored separately in its own dataset.

358

Name and location	Content description	Format and details
Metadata and documentation dataset	<ul style="list-style-type: none"> <li>- Data catalog</li> <li>- Documentation overview</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>- tab, xlsx</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:</li> </ul>	<ul style="list-style-type: none"> <li>- tab</li> </ul>

	<ul style="list-style-type: none"> <li>o Household composition, housing conditions, asset possession, land tenure, nutrition, food security, incomes and expenses.</li> <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>	<ul style="list-style-type: none"> <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>
Source code repository	<ul style="list-style-type: none"> <li>- Workflow to catalog the data, 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

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

362  
363 The two datasets are held in the DataSuds platform. It is an instance of the Dataverse open-  
364 source research data repository software. It was created and is administered by the French  
365 research institute for sustainable development (IRD) to provide research partners in  
366 developing countries with open science infrastructures.

367  
368 The source code is maintained in a GitHub repository. The authors might bring further  
369 improvements and complements to this source code. Each version of the source code is  
370 versioned, so the state of the source code at the time of the validation, included in the citation  
371 in Table 2, can always be retrieved.

## 372 373 **Technical Validation**

374  
375 In this section, we delve into the processes and methodologies employed in the ROS to ensure  
376 high data quality standards. We also provide a retrospective assessment of panel attrition, an  
377 important aspect for longitudinal studies.

### 378 379 *Data collection process*

380  
381 From the inception of the data collection in 1995, measures were taken at every stage of the  
382 survey process to ensure the reliability of the data collected. The ROS was created during a  
383 period marked by the multiplication of observatories in developing countries, often with multi-  
384 disciplinary approaches, to implement thematic monographies<sup>17</sup>. However, these initiatives  
385 often fell short of providing reliable longitudinal data suitable for quantitative analysis. In  
386 contrast, the observatories in Madagascar were established from the outset with a clear focus  
387 on economic and statistical expertise, making the production of consistent data series the  
388 main priority. Researchers from other disciplines, such as geography, economics and  
389 anthropology, were involved and their contributions were valued. However, the ROS have  
390 always emphasized the best practices in survey management, in particular regarding sampling,  
391 quality control, and data analysis.

392

393 Quality control during survey data collection was paramount. Each surveyor was limited to  
394 three household interviews per day to prevent rushing and ensure comprehensive data  
395 collection. These interviews, typically involving both the householder and their partner, lasted  
396 about two hours. At the end of the day, surveyors returned completed questionnaires to their  
397 supervisor, who assigned new interviews and conducted daily operations, including work  
398 schedule organization and questionnaire validation. Supervisors actively addressed errors or  
399 unclear responses and conducted consistency checks between questionnaire items. This close  
400 monitoring, facilitated by the concentration of interviews in a limited area, was unique and  
401 crucial for maintaining data quality in challenging rural conditions. The continuous presence  
402 of surveyors in the villages throughout the survey periods, along with the active involvement  
403 of supervisors in all stages of the fieldwork, ranging from organization to data processing, was  
404 a characteristic of this network. This feature distinguishes it from other rural surveys where  
405 temporary surveyors and supervisors are often less integrated into the overall workflow.  
406 During the survey period (mostly in the beginning and in the middle period), a researcher  
407 supported the team to improve the quality of the data collection.

408

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

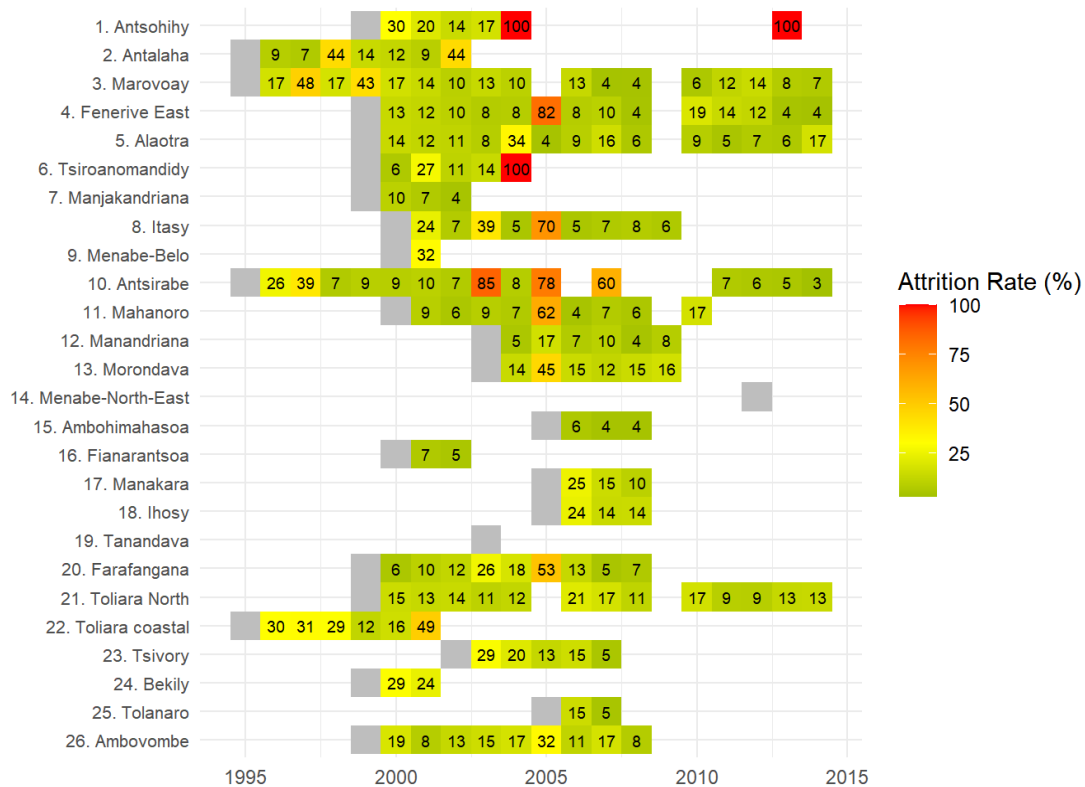
413

#### 414 *Attrition*

415

416 Significant levels of attrition are reported from the ROS data<sup>18,19</sup>. It is well-known that such  
417 levels of attrition are common place in any longstanding panel survey<sup>20</sup>. Attrition has been  
418 extensively studied and understood by survey experts and there is a strong consensus that  
419 despite this phenomenon, panel data is both useful and necessary for understanding  
420 population living standard dynamics<sup>21</sup>. The attrition rate is defined as the percentage of  
421 households interviewed in the previous year that were not re-interviewed in the year of  
422 interest. We calculated the attrition rate per year and per observatory to produce Figure 2.

423



424

425

**Figure 2. Attrition rates in ROS panels by observatory and survey round** (source: ROS raw data and author's calculations). *This heatmap displays the 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 percentage of attrition 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.*

433

434

The six instances of annual attrition rates above 75% for a specific observatory were likely induced by reshuffles of the 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. If we discard these outliers (attrition rates over 75%), we have an average attrition year of 15%, which seems in line with standards in developing countries, leading to a compound attrition rate of 80% over 10 years. In 2005, an IRD team conducted a tracking survey in one village of the Marovoay observatory (#3 in Figure 1) and was able to find 59% of the attrited households<sup>19</sup>. Among these, 35% of households had split, and 65% had migrated out of the village. These findings release part of our concern regarding ROS attrition: if we are interested in local development induced by protected areas, the fate of outgoing migrants is less of an interest, although the migration rate should be monitored somehow. Other motives of attrition observed by ROS members, although rarely, include the absence of the household during surveyor visits or the refusal to answer the survey.

447

448

449

## Usage Notes

450

451

Access to the survey documentation and metadata is open and enables any researcher potentially interested to learn about the dataset content, production process and associated materials. Access to the raw survey data requires researchers to register, specify the purpose

452

453

454 of data access, and commit to upholding the confidentiality rules outlined in the associated  
455 end-user license.

456

457 Due to the number of variables and survey rounds, we recommend users to make use of the  
458 data catalog included as a spreadsheet both in the documentation and the raw data datasets.  
459 The data catalog includes as columns the file name, table title, variable name, variable label  
460 and the years in which this variable is present in the dataset. The presence of labels makes it  
461 easier to find relevant information using full-text search.

462

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

471

472 Users are advised to post any questions or comments about the source code or data directly  
473 on the GitHub repository's issues page ([https://github.com/BETSAKA/Rural\\_obs\\_Madagascar/  
474 issues](https://github.com/BETSAKA/Rural_obs_Madagascar/issues)), rather than contacting the paper's corresponding author via email. This approach  
475 promotes transparency since all issues and responses are publicly visible and can be  
476 referenced by others with similar challenges. It also helps build a collaborative knowledge base  
477 that benefits the entire user community.

478

## 479 **Code Availability**

480

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

488

489

## 490 **Acknowledgements**

491

492 Besides the authors of this paper, the ROS endeavor was made possible by the help and  
493 dedication of a long list of colleagues at IRD, INSTAT, the University of Antananarivo and other  
494 institutions.

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

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

502 The gold open access fee for this paper was supported by the KfW Development Bank,  
503 financed by funds from the German Federal Ministry for Economic Cooperation and  
504 Development (BMZ).

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

508

## 509 **Author contributions**

510

511 VSA: data curation and validation.

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

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

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

517 JLD: ROS organization and investigation.

518 JdM: writing of the original draft of the data descriptor.

519 BNR: ROS administration, supervision and data curation.

520 RRafi: data curation and validation.

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

522 RRato: ROS administration, supervision and investigation.

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

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

526 MR: methodology and investigation.

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

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

530 CSM: data curation and validation.

531 All authors revised and approved the final manuscript.

532

## 533 **Competing interests**

534

535 The authors declare no competing interests.

536

## 537 **References**

538

539 1. Castañeda, A. *et al.* A New Profile of the Global Poor. *World Development* **101**, 250–267  
540 (2018).

541 2. *World Development Report 2008: Agriculture for Development*. (World Bank,  
542 Washington DC).

543 3. Jerven, M. Random Growth in Africa? Lessons from an Evaluation of the Growth Evidence  
544 on Botswana, Kenya, Tanzania and Zambia, 1965–1995. *Journal of Development Studies*  
545 **46**, 274–294 (2010).

546 4. Sandefur, J. & Glassman, A. L. The Political Economy of Bad Data: Evidence from African  
547 Survey & Administrative Statistics. *Available at SSRN 2466028* (2014).

548 5. Randall, S. & Coast, E. Poverty in African households: the limits of survey and census  
549 representations. *The Journal of Development Studies* **51**, 162–177 (2015).

- 550 6. Östberg, W., Howland, O., Mduma, J. & Brockington, D. Tracing improving livelihoods in  
551 rural Africa using local measures of wealth: A case study from central Tanzania, 1991–  
552 2016. *Land* **7**, 44 (2018).
- 553 7. Ansoms, A., Marijnen, E., Cioffo, G. & Murison, J. Statistics versus livelihoods: questioning  
554 Rwanda’s pathway out of poverty. *Review of African Political Economy* **44**, 47–65 (2017).
- 555 8. INSTAT. *Résultats Globaux Du Troisième Recensement Général de La Population et de*  
556 *l’habitation de 2018 à Madagascar (RGPH-3)*. (Institut national de la statistique,  
557 Antananarivo, 2020).
- 558 9. Roubaud, F. La question rizicole à Madagascar: les résultats d’une décennie de  
559 libéralisation. *Revue Economie de Madagascar* 37–62 (1997).
- 560 10. Roubaud, F. Le projet Madio à Madagascar : objectifs, démarche, résultats. *Statéco* 5–24  
561 (2000).
- 562 11. Razafindrakoto, M. & Roubaud, F. Two original poverty monitoring tools: the 1-2-3 surveys  
563 and the Rural Observatories. in *New international poverty reduction strategies* 335–361  
564 (Routledge, 2003).
- 565 12. Droy, I., Ratovoarinony, R. & Roubaud, F. Rural observatories in Madagascar: An  
566 innovative mechanism for monitoring the countryside. *Inter-Stat* 65–91 (2004).
- 567 13. Grislain, Q. Promesses et effets des observatoires fonciers: une géographie politique de  
568 dispositifs sociotechniques pour le développement en contextes africains et sénégalais  
569 (1980/2022). (Paris 1, 2022).
- 570 14. Dubois, J.-L. & Droy, I. L’observatoire: un instrument pour le suivi de la pauvreté. in *Les*  
571 *outils de mesure et de suivi/évaluation des politiques de sécurité alimentaires et de lutte*  
572 *contre la pauvreté* 22–46 (AEDES-IRAM, Brussels, 2002).
- 573 15. Saint-Macary, C. Le suivi des dynamiques de pauvreté en milieu rural: retour d’expérience  
574 des observatoires ruraux à Madagascar. *Statéco* 63–67 (2018).
- 575 16. Droy, I. Que sont les greniers à riz devenus ? (Madagascar). *Autrepart* **7**, 89–110 (1998).
- 576 17. *Observatoires du développement, observatoires pour le développement: actes du*  
577 *séminaire, septembre 1994*. (Éd. de l’Orstom, Paris, France, 1998).
- 578 18. Gubert, F. & Robilliard, A.-S. Risk and Schooling Decisions in Rural Madagascar: A Panel  
579 Data-Analysis. *Journal of African Economies* **17**, 207–238 (2008).
- 580 19. Vaillant, J. Attrition and Follow-Up Rules in Panel Surveys: Insights from a Tracking  
581 Experience in Madagascar. *Review of Income and Wealth* **59**, 509–538 (2013).
- 582 20. Alderman, H., Das, J. & Rao, V. *Conducting Ethical Economic Research Complications from*  
583 *the Field*. (The World Bank, 2013). doi:10.1596/1813-9450-6446.
- 584 21. Ashenfelter, O., Deaton, A. & Solon, G. *Collecting Panel Data in Developing Countries: Does*  
585 *It Make Sense?* (World Bank, Washington DC, 1985).
- 586 22. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for  
587 Statistical Computing, Vienna, Austria, 2023).
- 588 23. Wickham, H. *et al.* Welcome to the tidyverse. *Journal of Open Source Software* **4**, 1686  
589 (2019).
- 590 24. Wickham, H., Miller, E. & Smith, D. *Haven: Import and Export ‘SPSS’, ‘Stata’ and ‘SAS’ Files*.  
591 (2023).
- 592 25. Larmarange, J. *Labelled: Manipulating Labelled Data*. (2023).
- 593 26. Pebesma, E. Simple Features for R: Standardized Support for Spatial Vector Data. *The R*  
594 *Journal* **10**, 439–446 (2018).
- 595 27. van der Loo, M. P. The stringdist package for approximate string matching. *The R Journal*  
596 **6**, 111–122 (2014).



597 28. Tennekes, M. tmap: Thematic Maps in R. *Journal of Statistical Software* **84**, 1–39 (2018).  
598