



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-04502779v3

HAL Id: hal-04502779

<https://hal.science/hal-04502779v3>

Preprint submitted on 21 Aug 2024

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

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



Distributed under a Creative Commons Attribution 4.0 International License

Data Descriptor Template

Title

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

Authors

Velomalala Solo Andrianjafindrainibe⁷, Nicole Andrianirina⁷, Florent Bédécarrats^{1,2}, Isabelle Droy^{1,2}, Jean-Luc Dubois^{1,2}, Jeanne de Montalembert³, Bako Nirina Rabevohitra⁶, Rolland Rafidimanana⁷, Patrick Rasolofo^{5,2}, Raphaël Ratovoarinony⁷, Lalaso Anjarafara Onivola Ratsaramiarina⁶, Jean Dieudonné Ravelonandro⁷, Voahirana Razanamavo⁷, Mireille Razafindrakoto^{1,3}, Bezaka Rivolala⁴, François Roubaud^{1,3}, Camille Saint-Macary^{1,3}.

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)

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 2015, 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. 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 poverty¹. Recent reductions in poverty rates in developing countries stem from a decrease in rural poverty². 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 systems³⁻⁷. 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 areas⁸. 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 reforms⁹.
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 evolution¹⁰. 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.

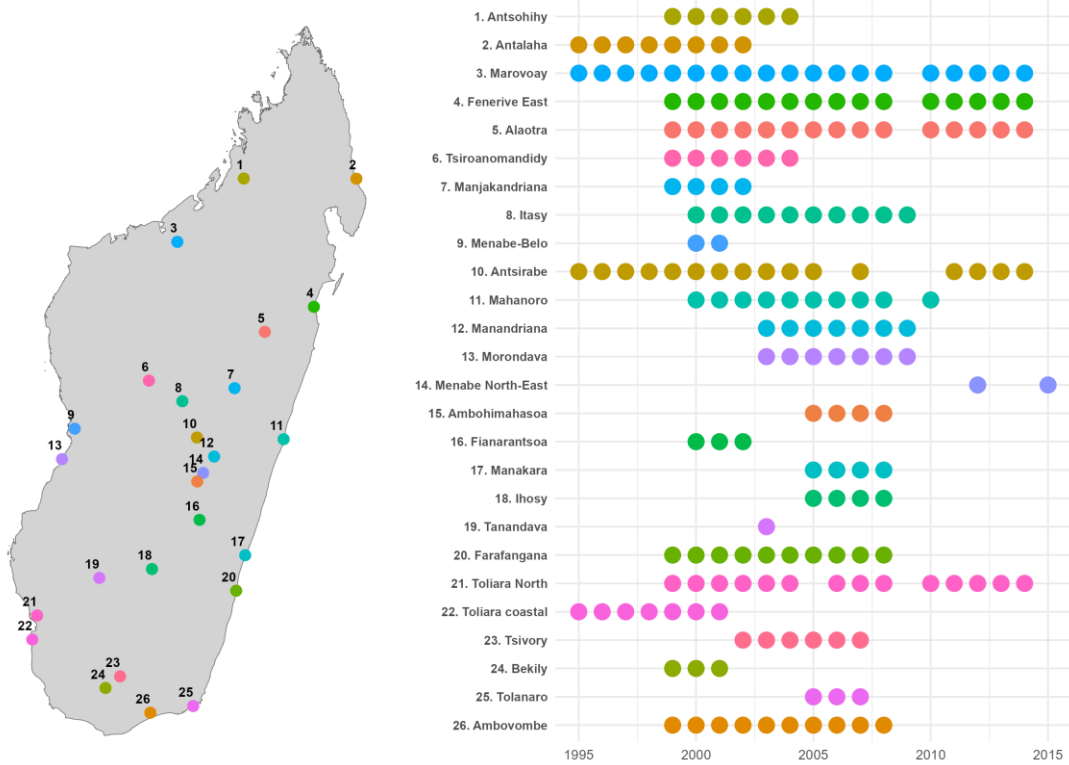
61
62 This lack of information motivated the creation of four rural observatories in Madagascar in
63 1995^{11,12}, each representing distinct agricultural challenges within the country's diverse
64 ecosystems. The observatories consisted of annual panel survey, dedicated to the continuous
65 collection, 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 policies¹³.

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 firms¹⁴.
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 2009¹⁵. 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. In 2015, a previously surveyed observatory was surveyed again (Menabe
83 North-East, #14 on Figure 1), and three new observatories were initiated
84 (Ambatofinandrahana, Anjozorobe, and Maintirano, not shown in Figure 1). Substantial work
85 is required to clean, harmonize, and document the data from these latest observatories to
86 include them in the dataset. The data from these new observatories will be added to the
87 dataset at a later stage after the publication of this data descriptor. The ROS remained
88 operational until 2017

89
90 The closure and relocation of observatories primarily resulted from changing priorities of
91 funding organizations and policymakers¹⁶. For example, the Toliara coastal and Antalaha
92 observatories (#22 and #2) were discontinued when the French Cooperation shifted its focus
93 to new sites established in 1999 and 2000. Toliara Coastal closed because another monitoring
94 system in the area focused on fishing and shrimp production. Similarly, the Antsohihy and
95 Tsiroanomandidy observatories (#1 and #6), funded by the European Union, were halted in
96 2005 due to a strategic shift towards food security, leading to the new observatories in the
97 southern regions. The Bekily observatory (#24) was closed after a German cooperation food
98 security project ended. The Antsirabe site (#10) suspended operations due to the gradual
99 withdrawal of Norwegian cooperation but was reactivated due to the Agricultural Ministry's
100 interest in its agro-pastoral system. Rice production has remained a major strategic priority
101 for Madagascar, so observatories focusing on rice production areas, particularly Alaotra (#5)
102 and Marovoay (#3), were continuously sustained. Internal management issues and difficulties

103 in maintaining competent personnel led to the intermittent cessation of operations, as seen
 104 with the Tsviry observatory (#23).
 105

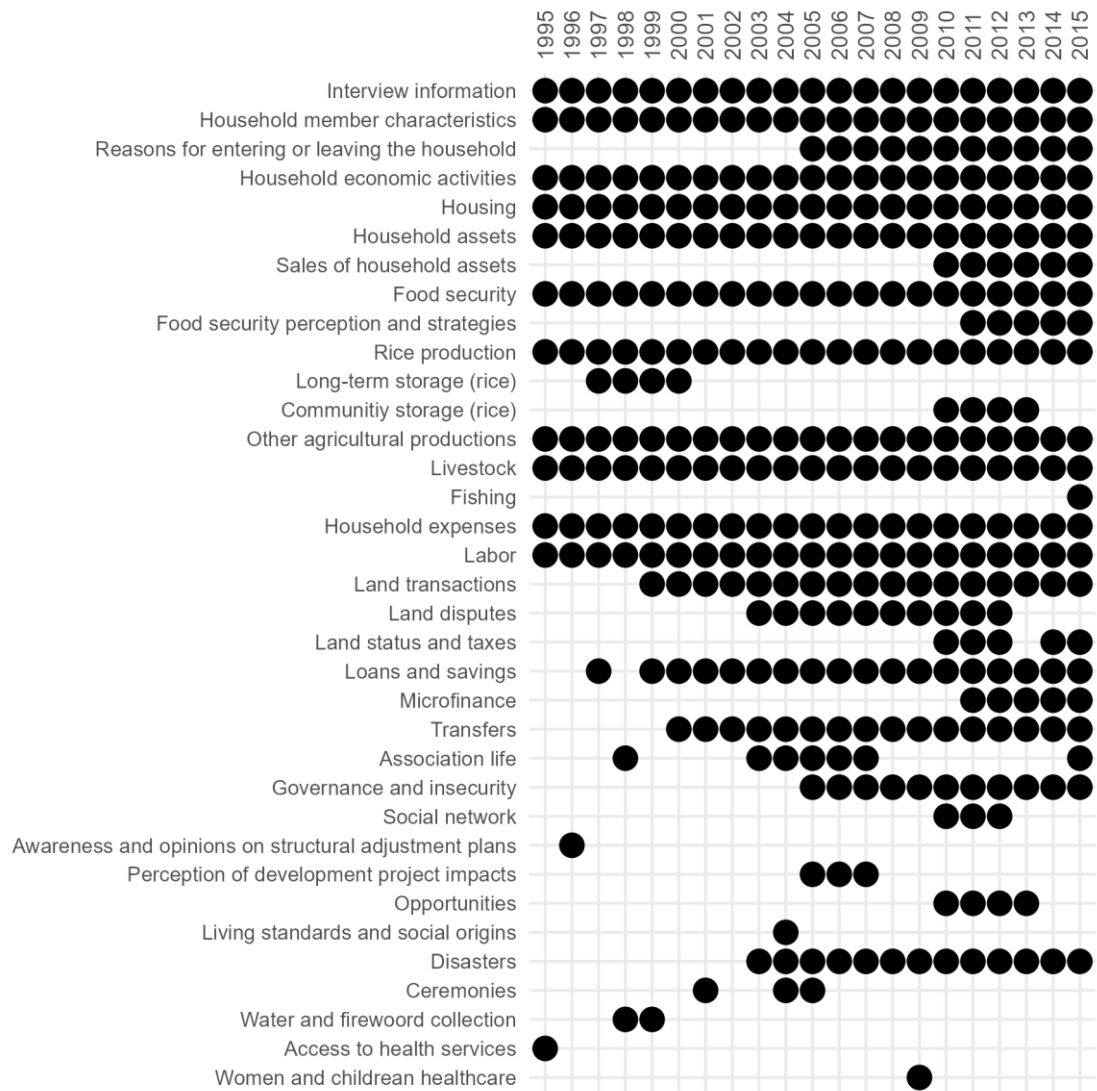


106
 107 **Figure 1: Coarse location and years of data collection of observatories** (Source: authors). *The*
 108 *figure is composed of two panels. The left panel displays a map of Madagascar with colored*
 109 *and numbered dots corresponding to the locations of the 26 observatories. The right panel*
 110 *indicates, for each number, the name of the observatory and the survey years.*

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

123
 124 The ROS methodology combined quantitative methods to describe situations and a qualitative
 125 perspective to explain them, offering a comprehensive view of rural dynamics. In addition to
 126 the household questionnaire, a community survey was conducted, involving field observations
 127 and semi-structured interviews. This mixed-method approach produces general information
 128 about the region and provides structural and situational insights on various aspects of
 129 community life (agricultural campaigns, social and cultural environment, health, education,
 130 security, development support, and product pricing).

131



132

133

Figure 2: Summary of the main modules included in the household ROS questionnaires each year (Source: authors). The chart illustrates the presence of survey modules across years from 1995 to 2015: each black dot represents the inclusion of a specific module in the survey for that year, while the absence of a dot indicates that the module was not included.

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

Over the years, the core questionnaire has been updated by adding new modules and removing of others, as shown in Figure 2. For instance, a module on governance has been included since 2005 to assess the people's experiences and perceptions of the effectiveness of public services (administration, hospitals, schools, etc.) and insecurity. Another example is the inclusion of a module on maternal and child health in 2009. This module was commissioned by donors in the context of a major political crisis that crippled the national health information system. It was not maintained in the subsequent years. Compared to specific health surveys, such as the Demographic Health Survey, the module included only basic questions on the prevalence of diarrhea, cough and fever, and access to health care. Another module on food security administered throughout the entire period also includes relevant information on health determinants, such as food quantity, quality and diversity. With regard to demographic events, live births have been recorded from the outset within the household member roster. However, data on deaths and outmigration were only collected from 2005 onwards, among the section registering the reasons for members to leave the household.

153

154 Retrospective assessments of the cost of the ROS in the 1990s and 2000s have been reported.
155 We consolidated this information and supplemented it with more recent accounts. To ensure
156 comparability with 2023 values, historical prices have been adjusted for inflation and the
157 average exchange rate between the respective survey years and 2023. As a result, we find that
158 the typical costs for data collection at the 2023 value range from €25 to €50 per surveyed
159 household. These costs vary according to several factors, such as the remoteness and
160 accessibility of the sites, the availability of public transport, catering and accommodation, and
161 whether the observatory is a new one. These figures do not include costs associated with
162 upstream activities, such as methodological design, or downstream tasks, such as data
163 cleaning, analysis, and dissemination.

164

165 Over the years, the ROS data have played an important role in various academic studies, policy
166 formulations, and development projects. The dataset includes a bibliographic database of 120
167 references, which provides a broad overview of the work that has been derived from this data.
168 The ROS has primarily served as a national instrument for monitoring changes in the
169 countryside and in the living conditions of rural households, enhancing the rural information
170 systems. Researchers from various disciplines, including sociology, economics and agronomy,
171 have used ROS data with both qualitative and quantitative methodologies. These studies cover
172 a range of topics such as poverty dynamics and vulnerability, family structures and food
173 security, and the impact of climate shocks and agricultural policies. Additionally, studies on
174 agricultural value chains and markets provide insights into the challenges and opportunities
175 facing Malagasy farmers. In general, the analyses show that rural households in Madagascar
176 are very poor, have little room for maneuver, and are highly vulnerable to economic or climatic
177 shocks.

178

179 The longevity and consistent methodology make it unique among households panel survey
180 experiences. Nevertheless, akin to all survey designs, it has encountered challenges,
181 particularly concerning measurement and sampling, addressed through methodological
182 improvement over time.

183

184 **Methods**

185

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

194

195 *Site Selection*

196

197 The selection of observatory locations was based on a qualitative approach leveraging the
198 expertise of individuals with extensive knowledge of the field to reflect the diverse array of
199 rural situations. Due to the multiplicity of terrains and climates, Madagascar has diverse
200 ecosystems for agriculture. Farming practices are shaped by the ethnocultural distribution of
201 the population, underscoring the need for rural development policies to consider this
202 diversity. A first subset of potential locations was prioritized to characterize a wide range of
203 agroclimatic areas, production systems, population density, accessibility, and support
204 structures. These potential locations were then correlated with development indicators

205 reflecting their economic and social conditions. The choice was also guided by the interests of
206 donors operating in a given region. The observatory locations were selected from each
207 category to constitute a varied portfolio. Each of the 26 observatories consisted of at least
208 two sites, whether hamlet or village, strategically selected to capture the local diversity. This
209 approach aimed to provide an in-depth understanding of the many facets of the rural
210 landscape.

211

212 Table 1 and Table 2 present the diverse range of agricultural and climatic zones encompassed
213 by the ROS, reflecting the country's rich ecological variety. The observatories cover areas
214 ranging from arid and semi-arid regions with predominant activities in dry crops (maize,
215 tubers) and pastoral farming with some irrigated rice perimeters to humid tropical zones
216 where rice cultivation, fishing, and a mix of subsistence and cash crops like vanilla, coffee, and
217 cloves thrive. This network captures the dynamics of different agricultural practices and
218 climatic challenges and addresses unique regional issues such as food insecurity, land conflicts,
219 and exposure to cyclones.

220

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

232

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

238

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

250

251 The Marovoay observatory (#3), in the Lower Betsiboka plain, was a significant rice-growing
252 area and belonged to one of the large irrigated perimeters constructed at the beginning of the
253 20th century¹⁸. This region was mainly populated by migrants from several regions in search
254 of salaried employment and land. A wide variety of communities could be found on the
255 Marovoay plain. The population is young, and the main activity is farming, with almost all
256 households having a secondary activity, often linked to farming or fishing. Farmers are highly

257 integrated into the market economy. This region was known for its productive rice fields, and
 258 even some rice varieties were exported to Europe until the 1970s. It experienced challenges
 259 during the socialist period and the 1980s crisis, leading local farmers to adapt through
 260 retrenchment or diversification strategies. Two villages were selected in 1995: Ampijoroa and
 261 Maroala.

262

263 As shown in Figure 1, the focus in Toliara shifted to a different location at the beginning of the
 264 2000s (#22 to #21), and the Antalaha observatory (#2) was discarded in 2004. The initial
 265 limitation to four observatories was due to resource constraints, the experimental nature of
 266 the project. The number of observatories expanded during the 2000s, influenced by the
 267 operational interests of the ROS donors, while seeking to maintain a similar level of diversity
 268 to that illustrated above.

269

| 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 fluctuations of cash crops prices on the global market. |
| Fianarantsoa #16 | High-altitude tropical climate, rice cultivation, off-season crops, small livestock farming. |
| Ihosaloha #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. |

270

Table 1: Description of Observatories with Names Starting with Letters A to L (Source: authors).

271

| Name and number on Figure 1 | Outstanding features |
|-----------------------------|---|
| 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. |

272 **Table 2: Description of Observatories with Names Starting with Letters M to Z** (Source:
273 authors).

274

275 *A Panel Household Survey*

276

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

293

294 The unit of observation for the rural observatories is the household, not the farm, as in
295 traditional agricultural surveys. This approach allowed for the comprehensive coverage of all
296 activities undertaken by each individual and to establish a crucial, yet often blurred, distinction
297 between the rural sphere and the realm of agriculture. While most rural households are
298 engaged in activities such as crop farming, livestock husbandry, or fishing, others pursue
299 occupations such as craftsmanship, trade, shopkeeping, or wage labor. Some households
300 combine agricultural and non-agricultural activities. In addition, within each household, a
301 distinction was made between the primary activity, which contributes most to income and
302 consumes most of working time, and the secondary activities. It is common for households to
303 participate in several income-generating activities, and their survival often hinges upon this

304 diversification. Similarly, this approach allows for the inclusion of transfer income from family
305 members engaged in long-term migration (over six months), but whose contributions enable
306 the unit to continue its farming activity. Therefore, focusing on households enables the
307 assessment of secondary activity, transfer revenues and identifies the working population
308 holding jobs not necessarily in the agricultural sector.

309

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

323

324 The ROS methodology differs in two main ways from other population observatories
325 widespread in Africa, such as the Demographic and Health Surveillance System (HDSS). Firstly,
326 the ROS aims to collect information on socioeconomic factors and agricultural activities while
327 the HDSS focuses primarily on population demographics and health events. Secondly, the
328 sampling strategy is significantly different: the ROS randomly select a sample of 500
329 households per observatory, whereas the HDSS collects information from the entire census of
330 a geographically defined population¹⁹. Nonetheless, both ROS and HDSS face a similar
331 limitation due to the longitudinal nature of their data: the ethical challenge of repeatedly
332 tracking households in the sample over a long period.

333

334 *Team Organization*

335

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

343

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

353

354 Before beginning the actual survey, surveyors underwent an "integration stage" for a few days
355 in the assigned villages. During this phase, they liaised with local administrative and traditional

356 authorities, often by presenting previous survey results. The team would then compile or
357 update the sample frame and, if necessary, randomly select households to replace those that
358 were absent. This period also enabled surveyors to familiarize themselves with local
359 measurement units, converting them into standard metrics, inventory the crops produced in
360 the village, and acclimate the villagers to the surveyors' presence.

361

362 *Implementation Phasing*

363

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

372

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

382

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

391

392 *Measurement Harmonization*

393

394 The selection of measurement units raised a methodological issue. Among the peasant
395 population, the units used to measure surfaces, weights, and volumes are not standardized
396 and vary from village to village. Requiring peasant farmers and fishermen to answer in
397 standard units, i.e., kilograms, liters, or hectares, was the surest way of losing touch with
398 reality. Hence, the ROS adopted "peasant measurements" which varied from one observatory
399 to another depending on farming practices, and were named differently in regional languages.
400 Output quantities are measured in units like *daba* or *vata* (oilcans), *zinga* (a kind of metal
401 drinking vessel), *sobika* (baskets), or cartloads. Rice-growing acreage can be quantified in
402 terms of the number of transplanted, seed quantities, *kipa* (the number of seedlings to be
403 planted), or carts of rice bales harvested in the field. These local measurement units were
404 identified in each village and converted into standard weights and measures. This conversion
405 process had to be conducted separately for each village due to variations in the meanings of
406 the same name across different locations (for example, the *cans* or *daba* hold different
407 volumes of fluid), and the absence of standardization of "peasant measurements" between

408 villages. After each complete interview, the survey supervisor reviewed the questionnaire and
409 converted the gathered “peasant measurements” into standard weights and measures. This
410 ensured the mitigation of measurement errors that are common in surveys on rural
411 households.

412

413 *Ethics*

414

415 The study was approved by the Madagascar National Statistical Office (INSTAT) and conducted
416 in collaboration with INSTAT and the Ministry of Agriculture, ensuring compliance with
417 national statistical regulations and public administration decisions. These institutions are the
418 legitimate authorities responsible for authorizing, administering, and vetting large surveys in
419 the country.

420

421 At the beginning of each interview, informed consent was obtained orally from each
422 respondent. If consent was not granted, the interview did not proceed, and a replacement
423 household was selected. No minors were interviewed. To ensure that respondents are fully
424 aware of the purpose of the ROS, one of the key principles of the system is the systematic
425 presentation of survey results as soon as they become available in the villages concerned. At
426 the same time, this principle of dissemination to the surveyed population ensures that they
427 benefit from the ROS and have the information they need to address difficulties or make their
428 voices heard.

429

430 Several measures were implemented to protect participant data. Direct identifiers were
431 removed and replaced with sequential aliases, pseudonymizing the data. Indirect identifiers
432 (e.g., village name, age, gender, professional activity) remain in the dataset, theoretically
433 allowing re-identification by someone with prior knowledge of the households. However,
434 given the low penetration of information technology and administrative systems in these rural
435 areas, the risk of re-identification is minimal.

436

437 To strengthen confidentiality, the data is available under a Data Usage Agreement (DUA) that
438 requires users to adhere to strict confidentiality rules, prohibits attempts to re-identify
439 participants, and mandates the data destruction after the project concludes. Access to the
440 microdata requires a confidentiality declaration, ensuring that only authorized researchers
441 with legitimate purposes can access the data.

442

443 **Data Records**

444

445 The data collection consists of two datasets and a code repository²⁰. The rationale behind
446 having two distinct datasets is to ensure that the ROS metadata and documentation are openly
447 accessible, while also addressing the confidentiality concerns related to the raw data from
448 household surveys. All direct identifiers have been removed from the raw data, yet it contains
449 detailed information about household composition, incomes, expenses, and assets as well as
450 individuals’ age, gender, education and activities. This level of detail poses a risk of re-
451 identifying specific individuals, especially in small rural areas. To mitigate this risk and protect
452 the confidentiality of sensitive information, the raw data from household surveys is stored
453 separately in its own dataset.

454

| Name and location | Content description | Format and details |
|------------------------------------|--|--|
| Metadata and documentation dataset | <ul style="list-style-type: none"> - Data catalog - Documentation catalog - ROS methodological material: surveyor manuals, questionnaires, and other technical documents - Full list of publications produced from the ROS data - Publications of the ROS project that are not accessible otherwise | <ul style="list-style-type: none"> - xlsx and tab - txt - pdf - bibtex and pdf - pdf |
| Raw household survey dataset | <ul style="list-style-type: none"> - Data catalog - One folder per year with, in each one: <ul style="list-style-type: none"> o Household composition, housing conditions, asset possession, land tenure, nutrition, food security, incomes and expenses. o Individual characteristics of household members: gender, age, education o Household member activities (with an in-depth analysis of farming practices), and finances o Additional annual or regional data collected in specific modules (e.g., fishing, vanilla production, natural disasters, hygiene, perinatal practices...) | <ul style="list-style-type: none"> - xlsx and tab - All raw data provided in two versions: <ul style="list-style-type: none"> o Original proprietary format (Stata) o Open format (tsv) |
| Source code repository | <ul style="list-style-type: none"> - Workflow to catalog the data and assess its quality, enhance its metadata and format - Data pseudonymization workflow - Workflow to geolocate raw data at village level based on toponyms | <ul style="list-style-type: none"> Quarto markdown notebook with R code sections for data processing |

455 **Table 3: Outline of the data records' content** (Source: authors). *Formats are abbreviated in*
456 *the table as follows: tabulation separated values (tsv), Excel Open XML Spreadsheet (xlsx) and*
457 *portable document format (pdf).*
458

459 The two datasets are hosted in the DataSuds platform. It is an instance of the Dataverse open-
460 source research data repository software. It was created and is administered by the French
461 Research Institute for Sustainable Development (IRD) to provide research partners in
462 developing countries with open science infrastructure.
463

464 The source code is maintained in a GitHub repository. Authors can make further
465 improvements and complements to this source code. Each version of the source code is
466 versioned, allowing users to retrieve the state of the source code at the time of the validation,
467 included in Table 3.
468

469 **Technical Validation**

470
471 This section delves into the processes and methodologies employed in the ROS to ensure high
472 data quality standards. Additionally, we provide a retrospective assessment of panel attrition,
473 an essential aspect for longitudinal studies.

474

475 *Data Collection Process*

476

477 From the inception of the data collection in 1995, measures were implemented at every stage
478 of the survey process to ensure the reliability of the data collected. The ROS was created
479 during a period characterized by the multiplication of observatories in developing countries,
480 often employing multidisciplinary approaches to implement thematic monographies²¹.
481 However, these initiatives often lacked reliable longitudinal data suitable for quantitative
482 analysis. In contrast, the observatories in Madagascar were established from the outset with
483 a clear emphasis on economic and statistical expertise, making the production of consistent
484 data series the main priority. While researchers from various disciplines, such as geography,
485 economics, and anthropology, contributed to the ROS, the project has always emphasized the
486 best practices in survey management, in particular concerning sampling, quality control, and
487 data analysis.

488

489 Quality control during survey data collection was paramount. Each surveyor was limited to
490 three daily household interviews, preventing rushing and ensuring comprehensive data
491 collection. These interviews, typically involving the householder and their partner, lasted
492 approximately two hours. At the end of each day, surveyors returned completed
493 questionnaires to their supervisor, who assigned new interviews and conducted daily
494 operations, including work schedule organization and questionnaire validation. Supervisors
495 actively addressed errors or unclear responses and conducted consistency checks between
496 questionnaire items. This close monitoring, facilitated by concentrating interviews in a limited
497 area, was unique and crucial for maintaining data quality in challenging rural conditions. The
498 continuous presence of surveyors in the villages throughout the survey periods, coupled with
499 the active involvement of supervisors in all stages of fieldwork, from organization to data
500 processing, characterized this network. This feature sets it apart from other rural surveys,
501 where temporary surveyors and supervisors are often less integrated into the workflow.
502 During the survey period, especially in the beginning and middle stages, a researcher
503 supported the team to improve the quality of the data collection.

504

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

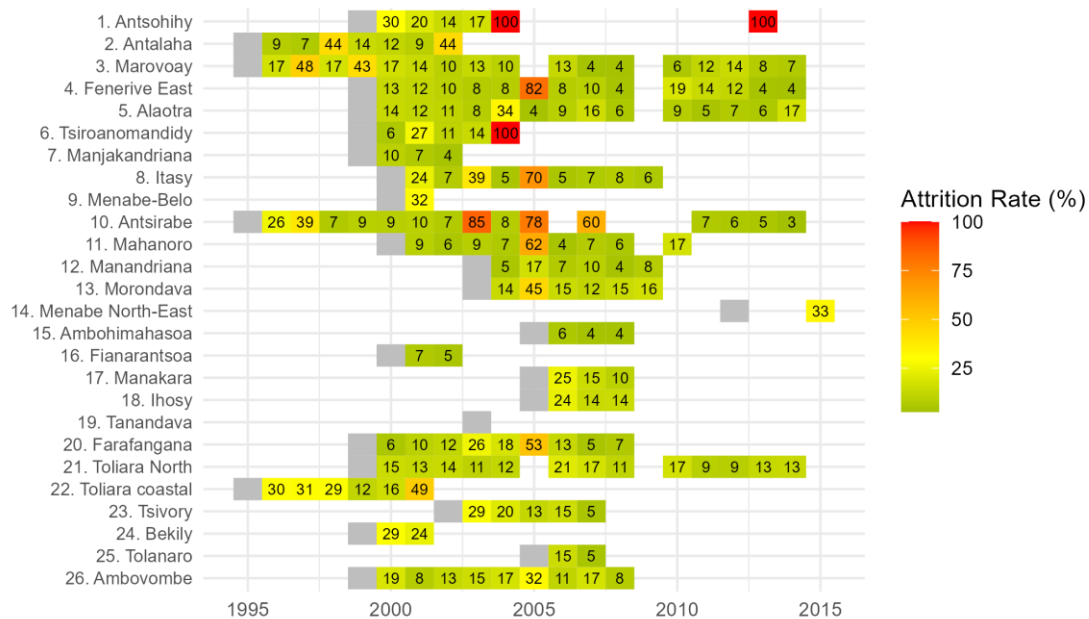
509

510 *Attrition*

511

512 Significant levels of attrition are reported from the ROS data^{22,23}. It is well known that such
513 attrition levels are common in any longstanding panel survey²⁴. Attrition has been extensively
514 studied and understood by survey experts. There is a strong consensus that despite this
515 phenomenon, panel data is useful and necessary for understanding population living standard
516 dynamics²⁵. The attrition rate is the percentage of households interviewed in the previous year
517 that were not re-interviewed in the year of interest.

518



519

520

Figure 3. Attrition rates in ROS panels by observatory and survey round (Source: ROS raw data and author calculations). *This heatmap displays attrition rates for each observatory across survey years from 1995 to 2015. The vertical axis lists the observatories, while the horizontal axis represents the years. Each square at the intersection of an observatory and a survey year denotes the attrition rate, with the color gradient varying from green (0% attrition) to red (100% attrition). The attrition rate for each year is rounded and displayed within the corresponding square. The initial survey year for each observatory is marked in gray, indicating the absence of prior data for attrition calculation.*

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

Usage Notes

544

545

546

547

548

549

550

551

552

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

Given the number of variables and survey rounds, we recommend users to utilize the data catalog provided as a spreadsheet in the documentation and the raw data datasets. The data catalog contains the file name, table title, variable name, variable label and the years in which

553 this variable is present in the dataset. The presence of labels facilitates the findings of relevant
554 information using full-text search.

555

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

564

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

571

572 **Code Availability**

573 The data pseudonymization, data validation, and guidance for data georeferencing and
574 analysis are available at https://github.com/BETSAKA/Rural_obs_Madagascar, under the MIT
575 open-source license. This work used *R* 4.3.1 for the general data processing environment²⁶,
576 alongside the following packages: *tidyverse* 2.0.0 for data preparation and visualization²⁷,
577 *haven* 2.5.3 for reading and converting files from Stata format²⁸, *labelled* 2.12.0 for generating
578 the data catalog²⁹, *sf* 1.0.14 for spatial data georeferencing³⁰, *stringdist* 0.9.10 for toponym
579 matching and data pseudonymization³¹, and *tmap* 3.3.4 for spatial data visualization³².

580

581

582 **Acknowledgements**

583

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

587

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

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

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

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

600

601 **Author contributions**

602

603 VSA: data curation and validation.
604 NA: ROS administration and supervision, investigation, data curation, and validation.
605 FB: formal analysis, software (geolocation, pseudonymization and data formatting)
606 visualization and original draft writing of the data descriptor.
607 ID: ROS conceptualization and methodology design, funding acquisition, ROS administration,
608 supervision, investigation, and validation.
609 JLD: ROS organization and investigation.
610 JdM: data curation, original draft writing of the data descriptor.
611 BNR: ROS administration, supervision and data curation.
612 RRafi: data curation and validation.
613 PR: funding acquisition, ROS administration, supervision, investigation and validation.
614 RRato: ROS administration, supervision and investigation.
615 LAOR: ROS administration and supervision, investigation, data curation and validation.
616 JDR: ROS administration and supervision, software (data entry setup), data curation and
617 validation.
618 VR: ROS administration and supervision, investigation, data curation and validation.
619 MR: methodology and investigation.
620 BR: funding acquisition, ROS administration, supervision, investigation, and validation.
621 FR: ROS conceptualization and methodology design, funding acquisition, ROS administration,
622 supervision, investigation, and validation.
623 CSM: data curation and validation.
624 All authors revised and approved the final manuscript.

625

626 **Competing interests**

627

628 The authors declare no competing interests.

629

630 **References**

631

- 632 1. Castañeda, A. *et al.* A New Profile of the Global Poor. *World Development* **101**, 250–267
633 (2018).
- 634 2. *World Development Report 2008: Agriculture for Development*. (World Bank,
635 Washington DC).
- 636 3. Jerven, M. Random Growth in Africa? Lessons from an Evaluation of the Growth Evidence
637 on Botswana, Kenya, Tanzania and Zambia, 1965–1995. *Journal of Development Studies*
638 **46**, 274–294 (2010).
- 639 4. Sandefur, J. & Glassman, A. L. The Political Economy of Bad Data: Evidence from African
640 Survey & Administrative Statistics. *Available at SSRN 2466028* (2014).
- 641 5. Randall, S. & Coast, E. Poverty in African households: the limits of survey and census
642 representations. *The Journal of Development Studies* **51**, 162–177 (2015).
- 643 6. Östberg, W., Howland, O., Mduma, J. & Brockington, D. Tracing improving livelihoods in
644 rural Africa using local measures of wealth: A case study from central Tanzania, 1991–
645 2016. *Land* **7**, 44 (2018).
- 646 7. Ansoms, A., Marijnen, E., Cioffo, G. & Murison, J. Statistics versus livelihoods: questioning
647 Rwanda’s pathway out of poverty. *Review of African Political Economy* **44**, 47–65 (2017).
- 648 8. INSTAT. *Résultats Globaux Du Troisième Recensement Général de La Population et de*
649 *l’habitation de 2018 à Madagascar (RGPH-3)*. (Institut national de la statistique,
650 Antananarivo, 2020).
- 651 9. Roubaud, F. La question rizicole à Madagascar: les résultats d’une décennie de
652 libéralisation. *Revue Economie de Madagascar* 37–62 (1997).

- 653 10. Roubaud, F. Le projet Madio à Madagascar : objectifs, démarche, résultats. *Statéco* 5–24
654 (2000).
- 655 11. Razafindrakoto, M. & Roubaud, F. Two original poverty monitoring tools: the 1-2-3 surveys
656 and the Rural Observatories. in *New international poverty reduction strategies* 335–361
657 (Routledge, 2003).
- 658 12. Droy, I., Ratovoarinony, R. & Roubaud, F. Rural observatories in Madagascar: An
659 innovative mechanism for monitoring the countryside. *Inter-Stat* 65–91 (2004).
- 660 13. Grislain, Q. Promesses et effets des observatoires fonciers: une géographie politique de
661 dispositifs sociotechniques pour le développement en contextes africains et sénégalais
662 (1980/2022). (Paris 1, 2022).
- 663 14. Dubois, J.-L. & Droy, I. L’observatoire: un instrument pour le suivi de la pauvreté. in *Les*
664 *outils de mesure et de suivi/évaluation des politiques de sécurité alimentaires et de lutte*
665 *contre la pauvreté* 22–46 (AEDES-IRAM, Brussels, 2002).
- 666 15. Saint-Macary, C. Le suivi des dynamiques de pauvreté en milieu rural: retour d’expérience
667 des observatoires ruraux à Madagascar. *Statéco* 63–67 (2018).
- 668 16. David-Benz, H., Benoit-Cattin, M. & Ramboarison, R. *Evaluation Du Réseau Des*
669 *Observatoires Ruraux à Madagascar*. (CIRAD, Montpellier, 2010).
- 670 17. Bédécarrats, F., Dazet, F., Guérin, I. & Razafindrakoto, M. Tracking the Cost of Living, for
671 Whom and at What Price? A Political Economy of Price Indicators in Madagascar.
672 *International Sociology* **38**, 684–703 (2023).
- 673 18. Droy, I. Que sont les greniers à riz devenus ? (Madagascar). *Autrepart* **7**, 89–110 (1998).
- 674 19. Herbst, K. *et al.* The INDEPTH Data Repository: An International Resource for Longitudinal
675 Population and Health Data From Health and Demographic Surveillance Systems. *Journal*
676 *of Empirical Research on Human Research Ethics* **10**, 324–333 (2015).
- 677 20. Andrianjafindraine, V. S. *et al.* Madagascar rural observatory surveys, a longitudinal
678 dataset on household living conditions 1995-2015: microdata and documentation.
679 DataSuds <https://doi.org/10.23708/WNM6LW> (2024).
- 680 21. *Observatoires du développement, observatoires pour le développement: actes du*
681 *séminaire, septembre 1994*. (Éd. de l’Orstom, Paris, France, 1998).
- 682 22. Gubert, F. & Robilliard, A.-S. Risk and Schooling Decisions in Rural Madagascar: A Panel
683 Data-Analysis. *Journal of African Economies* **17**, 207–238 (2008).
- 684 23. Vaillant, J. Attrition and Follow-Up Rules in Panel Surveys: Insights from a Tracking
685 Experience in Madagascar. *Review of Income and Wealth* **59**, 509–538 (2013).
- 686 24. Alderman, H., Das, J. & Rao, V. *Conducting Ethical Economic Research Complications from*
687 *the Field*. (The World Bank, 2013). doi:10.1596/1813-9450-6446.
- 688 25. Ashenfelter, O., Deaton, A. & Solon, G. *Collecting Panel Data in Developing Countries: Does*
689 *It Make Sense?* (World Bank, Washington DC, 1985).
- 690 26. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for
691 Statistical Computing, Vienna, Austria, 2023).
- 692 27. Wickham, H. *et al.* Welcome to the tidyverse. *Journal of Open Source Software* **4**, 1686
693 (2019).
- 694 28. Wickham, H., Miller, E. & Smith, D. *Haven: Import and Export ‘SPSS’, ‘Stata’ and ‘SAS’ Files*.
695 (2023).
- 696 29. Larmarange, J. *Labelled: Manipulating Labelled Data*. (2023).
- 697 30. Pebesma, E. Simple Features for R: Standardized Support for Spatial Vector Data. *The R*
698 *Journal* **10**, 439–446 (2018).

- 699 31. van der Loo, M. P. The stringdist package for approximate string matching. *The R Journal*
700 **6**, 111–122 (2014).
- 701 32. Tennekes, M. tmap: Thematic Maps in R. *Journal of Statistical Software* **84**, 1–39 (2018).
702