

Fieldwork, Economic Theory and Research on Institutions in Developing Countries

Christopher Udry

Yale University

udry@yale.edu

January, 2003

Development economics has been the beneficiary of a rich tradition of field research. Within this broad tradition there is a huge variety of methods, from short qualitative studies in which the primary interaction between the researcher and the participants is relatively unstructured conversation to large-scale surveys designed by and perhaps loosely supervised by economists. In this note, however, I focus on one point in this broad space of research methodologies - iterative field research in which the collection of data through surveys is combined with detailed observation and conversation to elicit knowledge about institutions.

A highly artificial, but I hope useful typology is provided in Figure 1. Typically, empirical work in economics relies on existing data. However, it is becoming more common in development economics to complement existing data with relatively short, often less structured visits to the field site in order to clarify aspects of the data, to better define the economic environment, or to collect limited amounts of complementary data. For example, ICRISAT hosted and provided institutional support for a series of visiting scholars during the collection of the Village Level Surveys. This proved to be a relatively inexpensive mechanism that generated an important sequence of insights regarding economic institutions in India (Rosenzweig (1998 EJ), Pender (1996) are examples of papers emerging from this

practice).

At the forefront of public service in development economics is that set of academic economists who have combined their time and energy with significant outside resources to direct and organize the collection of new large-scale datasets. An important recent example is the Indonesian Family Life Survey (<http://www.rand.org/labor/FLS/IFLS>), which tracks approximately 7,000 households over approximately 7 years.

In this note I discuss a less institutional, more personal form of field research that requires fewer external resources than the large data-collection exercise exemplified by the IFLS. This is a method of intermediate scale in which the collection of data through surveys is combined with detailed observation and conversation. The hallmark of this work is that it engages the researcher in an interactive process of detailed observation, construction of economic models, data collection and empirical testing. An initial hypothesis is refined and clarified through detailed observation, which informs the collection of appropriate data. As the economic environment is clarified during the course of fieldwork, the data collection procedure can be adjusted in response. Finally, the research proceeds to formal statistical analysis and, one hopes, to new hypotheses. This iterative process of moving between theoretical reasoning, informal observation and discussion, data collection and statistical analysis is the locus of creativity in this kind of field research and is its distinguishing feature. The relatively small scale of the research facilitates this iterative process, particularly with respect to the ability of the researcher to modify data collection on the fly. In Figure 1, this is represented by Goldstein and Udry (2002).

Unlike the more purely qualitative or case study research (e.g., Townsend 1995) that can serve to generate hypotheses or clarify aspects of the economic environment for modelling purposes, this kind of research relies on formal statistical reasoning. It is an intensive and time-consuming type of research. It involves extended periods of time in the field, typically six months to two years. During this period, much of the researcher's time is spent fully engaged in the day-to-day details of interviewing and collecting data, NOT writing papers. One is left with data on a sample that is perhaps 1/20 the size of the stunningly rich IFLS surveys. Standard principles of comparative advantage would seem to indicate

that economists should specialize in generating models, writing papers, and encouraging specialists in data collection to collect the sorts of data that will be useful in testing these models, or for generating new ideas about the way the development process works. We are typically not well-trained in supervising interviewers, nor for that matter doing all the other mundane tasks associated with running the equivalent of a small business in a developing country. Under what circumstances is this kind of iterative field research appropriate?

Clearly, it must be the case that the question cannot be addressed using available data. Moreover, even when there is an important gap in available data, in many instances it would be more productive to augment that existing data with targeted supplementary research. The enormous advantage of supplementing existing data is that one could achieve much larger samples for a given cost. If this approach is not feasible because there is no available relevant baseline data, or if there are insurmountable obstacles to coordinated research, then collecting one's own data may be the appropriate strategy.

If existing data are not available and if the research question is sufficiently well-defined, then a conventional program of data collection might suffice. In this case, the data required are determined by a model specified in advance of the field research, and this in turn guides the creation of survey instruments and the sample design (some essential references are Deaton (1997) and Grosh and Glewwe (2000); also see the resources at <http://sticerd.lse.ac.uk/FIELDWORK/>). In Figure 1, this approach is exemplified by Bandiera and Rasul (2002).

A different method is required when the research question is more ambiguous and open-ended. When the question of interest is clear, but the economic environment within which agents live is not well-documented, then iterative field research becomes particularly useful. This inductive process of moving back and forth between hypothesis, observation and testing characterizes much of the research process in applied economics generally. There are many examples of research programs that move through this process over time, beginning with general hypotheses that are clarified through qualitative field research (Townsend (1994) in figure 1), through the construction of more fully-specified models as the environment and incentive structures are clarified, which in turn guide the design of a survey

(<http://www.src.uchicago.edu/users/robt/>). The point of iterative field research is to compress this process: the interaction between inductive and deductive reasoning that might otherwise take place over a sequence of papers, perhaps spread across many researchers, is concentrated in one project.

To summarize, the primary benefit of iterative field research is that the researcher can address questions that are less well-defined than can typically be managed using existing data or through more conventional survey methods. It provides a method for opening up new questions, for being surprised. The most important cost is that for a given sample size it is much more expensive, particularly in terms of the researcher's time, than alternative methods.

Anderson and Baland (2002) provides a vivid recent example of these principles. In this paper, Anderson and Baland examine economic behavior at the nexus of two complex institutions: rotating savings and credit associations (roschas) and households. The authors argue that rosca participation in Kibera, Kenya (a slum of Nairobi) is dominated by married women because such participation is a valuable tool for a women seeking to protect her savings against the immediate claims of her husband. The paper is convincing because of the inclusion of telling details from rosca meetings and from the constitutions of the roschas. Conversations between the authors and members of the groups clarify the importance of confidentiality - specifically confidentiality from husbands. This feature would not have been apparent without direct contacts between the researcher(s) and people who use the institution under investigation. These rich details inform the theoretical model used in the paper, the data collection (although this less so), and the econometric estimation. However, one of the problems of conducting one's own field research is also evident: the sample size is small enough that some important parameters are not very precisely estimated.

HOUSEHOLD ORGANIZATION, LAND TENURE AND SOIL FERTILITY

I frame the rest of the discussion of the potential benefits and limitations of iterative field research around an account of some work that I have been conducting with colleagues in

Ghana (Goldstein and Udry, 2000, 2002; Conley and Udry 2002). One of our objectives when we began planning research was to understand the dynamics of land resource management in an environment characterized by apparently imperfect financial markets and complex land tenure arrangements. This was a context in which iterative field research would seem to be a valuable tool: there is little data available from Africa which combines a rich set of economic information with data on soil fertility. More importantly, even after an extensive literature review, the relevant institutional context was quite obscure. Open-ended and extensive discussions with farmers would be important to clarify the incentives confronting individuals as they managed their land.

On the other hand, an important worry was that the relatively short time scale of the data collection process (two years) might make it difficult to discern movements in soil fertility, which might become apparent only over a longer period.

Initial Theoretical Concerns

Decisions regarding the management of a renewable resource such as land fertility are strongly influenced both by land tenure and by capital market imperfections. It is very difficult to make strong predictions about patterns of investment in land without good knowledge of the incentives faced by individuals in those dimensions. However, the efficient allocation of resources within households implies strong implications for within-household patterns of investment in land fertility, even in the context of imperfections in land and financial markets. The argument leans heavily on the assumption that all potential gains from trade are exhausted within households, and leads to the conclusion that plots with similar inherent physical characteristics cultivated by different individuals within the household should have similar patterns of land investment.

The analysis of land fertility management fits naturally into the well-established literature on optimal extraction of a renewable resource.¹ Let $\phi_{hi}(t)$ be a scalar index of fertility

¹Donovan and Casey, 1998; Clay et al. 1998; Krautkraemer, 1994; Barrett, 1991. The source of the model is Lewis and Schmalensee 1977.

on plot i cultivated by someone in household h at time t . Let $g(\phi)$ be the natural rate of regeneration so that on fallow land $\dot{\phi}_{hi}(t) = g(\phi_{hi}(t))$, with $g(\cdot)$ concave.

The control variable is the anthropogenic change in fertility, which we denote by s (think of s as an index of variable inputs into cultivation). Profits can be increased by increasing the rate of soil deterioration, so profits at time t are $\pi(\phi_{hi}(t), s_{hi}(t))$ where π is increasing in both arguments. The cost of extracting resources from the soil, of course, is diminished fertility in the future: $\dot{\phi}_{hi}(t) = g(\phi_{hi}(t)) - s_{hi}(t)$.

Plot profits, along with other income, are used to purchase streams of consumption for household members. For simplicity, let there be two members and let $v_{hf}(c_h)$ and $v_{hm}(c_h)$ represent their utilities from a given stream of consumption over the life of the household. In a Pareto efficient allocation in the household, consumption and the paths of resource extraction on the household's plots are chosen such that $v_{hf}(c)$ is maximized subject to $v_{hm}(c) \geq \bar{v}_{hm}$ and to a common household resource constraint that might look something like

$$\int_0^{\infty} e^{-t\delta_h(t)} \left[\sum_i \pi_{hi}(\phi_{hi}(t), s_{hi}(t)) - p_h(t)c_h(t) \right] dt \geq 0$$

where $\delta_h(t)$ is a household specific-discount rate at time t . In the simplest models, this would be the interest rate at which the household can borrow or lend, though in environments with more complex financial market imperfections it would emerge endogenously as the shadow price of resources in period t . In either case, the substantive implication for fertility management in an efficient household is unchanged: profits from any of the household's plots are pooled, aggregated over time at a household- (*not* plot-) specific discount rate, and allocated to the consumption of household members. The consequence of this pooling is that fertility management will be similar on similar plots within the household.

The broad outlines of the dynamics of fertility under a fallow system are simple to describe: a plot is cultivated as long as its fertility $\phi_{hi}(t)$ is greater than a critical value and during this period fertility falls. When the critical level of fertility is reached, the plot is fallowed, and fertility recovers until it is sufficiently productive to justify the fixed costs of reestablishing cultivation. An immediate implication of Pareto efficiency within the house-

hold is that for any two currently cultivated (or any two currently fallowed) plots within the household $\phi_{hi}(t) = \phi_{hj}(t) \Rightarrow s_{hi}(t) = s_{hj}(t)$, so that the time paths of fertility on similar plots within the household are identical. Therefore, even in the context of quite imperfect markets, Pareto efficiency within households provides some strong testable restrictions on behavior. However, these predictions are restricted to within-household comparisons; to move beyond that boundary we need a better understanding of the institutional environment. Moreover, even within the household, it would be necessary to clarify property rights and issues of intrahousehold resource allocation if the null hypothesis of Pareto efficiency were to be rejected.

Design of Data Collection

A key feature of this analysis is the current level of soil fertility, ϕ . After a literature review and conversations with soil scientists in Ghana, we decided to attempt to estimate $\phi_{hi}(t)$ by collecting a (short) panel of measurements of soil organic matter (OM) by testing the soil on each of the plots cultivated by members of our sample households in each of the two years of the survey.

The survey was conducted in the Akwapim South District of the Eastern Region of Ghana. We selected four village clusters (comprising 5 villages and two hamlets) with a variety of cropping patterns and market integration. Within each village cluster we selected 60 married couples (or triples - about 5-10 percent of the population is polygynous) for our sample; in three village clusters this was a simple random sample, and in the fourth, we interviewed the entire population of married couples. Each member of the pair or triple was interviewed 15 times during the course of the two years. Every interview was carried out in private, usually by an enumerator of the same gender.

The survey was centered around a core group of agricultural activity questionnaires (plot activities, harvests, sales, credit) that were administered during each visit. We also mapped each plot using a global positioning system. In addition about 35 other modules were administered on a rotating basis.

Finally, we collected data on land tenure by plot. Based on preliminary interviews and a literature review, we decided to take two tracks towards an attempt to understand the complex and overlapping systems of land tenure in the region. On the one hand, we asked a series of questions of each cultivator regarding what they perceived as their rights over the plot. These questions were drawn from earlier surveys in Ghana (Place and Hazell 1993) and there is some evidence that they are related to some types of land investment (Besley 1995). On the other hand, following Berry (1993), we complemented this perceived rights information with questions about the history of the plot: specifically, the process through which the current cultivator acquired the plot. This includes details of formal contracts (sharecropping, rental, purchase) and the more common informal methods (allocated land from the lineage or village, or land passed within the household, usually from husband to wife).

Observation and Adoptive Design of Data Collection

There were three reasons for designing the survey with repeated, partially-varying modules: first, to reduce errors of recall, particularly with respect to plot-level inputs and outputs; second, to generate panel data for consumption, income, time-use, and financial transactions; and third and most importantly to permit modifications and additions to the set of survey instruments as details of the institutional context became clear, and as new hypotheses emerged.

>From conversations during the field research, it became clear that our information on perceived land rights and on plot histories provided an inadequate account of individuals' expectations regarding their future rights over the plot. Most importantly, we learned that virtually all land, regardless of its current tenurial status, can be traced to a source lineage. A respondents' status within his or her lineage might be a determinant of his or her expectations over and above the contractual status of a plot, or his or her current "rights" over the plot. In particular, membership in that lineage, or political power within that lineage might have an important effect on one's security of tenure over the plot. An additional

salient influence over a cultivator's security of tenure seemed to be the relationship of the cultivator to the individual from whom he or she received the plot, and *that* individual's social position.

Therefore, we added questions that recorded the lineage to which each plot can be assigned, and a sequence of questions about the lineage membership and status within the membership of each individual in the sample. In addition, we collected information on the identity of the person from whom land was received.

Preliminary Statistical Analysis

Two important conclusions emerged from our preliminary analysis of the data. First, it became quite clear that our measures of soil organic matter and soil acidity are far from sufficient statistics for current fertility (see Goldstein and Udry (1999) for the gory details). The major problem here is the lack of a deep enough time-series. Changes in OM over the two years of the survey seem to be dominated by a large aggregate shock generated by weather variation (or by a change in the timing of testing relative to the end of the rainy season). The data do not cover a long enough period to observe the anthropogenic changes in OM that might matter. Nor does the measurement of OM seem to be a reasonable summary of the multiple relevant dimensions of soil fertility in the area: in particular, the weed load appears to be an important element of fertility.

Therefore, we turned next to an examination of fallowing choices and land productivity. We found that there are dramatic differences within households in the fallowing behavior of husbands and wives on physically similar plots. Husbands systematically fallow their plots for longer periods than their wives. As a consequence, husbands achieve startlingly higher yields and profits than their wives. This difference in fallowing behavior, in turn, is related to the difference in political power of husbands and wives: men are far more likely to occupy important offices within the lineage or village hierarchy than are their wives. In households in which neither the husband nor his wife hold an office, there is no significant difference in fallowing behavior nor in plot-level profits (Goldstein and Udry (2002)).

Further Observation

In order to explore the reasons for this variation in fallowing behavior within households, after completion of data collection I conducted a sequence of focus group interviews in the sample villages. When confronted with preliminary results relating to the gender differential in plot profits and fallowing behavior, and its relationship to office holders many participants expressed little surprise. A consensus quickly emerged that the primary cause of our finding is a particular kind of uncertainty over land tenure. The worry that emerged in the interviews was not that individuals reduce their investment because of the possibility that they will lose their right to cultivate that plot in the future. Instead, women in particular worry that the very act of investing in the land (that is, leaving it fallow) would reduce their security of tenure. There is simply no danger of losing access to legitimately-acquired land as long as it is under cultivation. However, once fallowed, the right to re-establish cultivation is uncertain even on land that was obtained through a legitimate process (e.g., allocated by the lineage leadership).

In several of the focus groups, the danger of losing one's right to cultivate a plot was related to one's perceived 'need' for that plot. To paraphrase a common view: "the land belongs to the lineage, not to me. If I leave it fallow, someone may say 'she does not need that land, she is just letting it sit there unused' and get use of it for himself."

New Theory

The participants in the focus groups seemed to be describing a land allocation process that was designed to reveal information about one's need for land. This led us to consider a mechanism design problem that had not been apparent at the initiation of the survey. There is strong evidence (from other aspects of our survey) that there is incomplete information within the lineage about the incomes of other members of the lineage, particularly with respect to non-farm income (which is about half of total income). The description that was provided in the focus groups implied that a goal of the leadership of the lineage is to allocate land in such a way as to minimize the number of lineage members whose total

income falls below a certain level.

Suppose that there are two types of individuals in the lineage, one that has a sufficiently lucrative non-farm business and sufficient land under her own control that she can achieve an income above the minimum without additional land from the lineage. The other type needs additional land from the lineage to achieve the minimum acceptable income, given her own land endowment and non-farm income opportunities.

If the lineage leadership had complete information about members' types, it could simply allocate sufficient land to members of the poor type to bring their incomes up to the minimum. However, the types are not observed, so the lineage leadership must devise a contract such that the high types will refuse the additional land; a cultivation requirement serves this purpose. The lineage offers its members a deal akin to the following: you have free access to x units of additional land, however, in order to get this land you must keep at least y units of land under cultivation rather than left fallow.

Suppose that fallowing is productive in the sense that after appropriate discounting a plot fallowed in the first period yields more in aggregate over the relevant horizon than a plot that is continuously cultivated. For appropriately chosen values of x and y , the high type would refuse the additional land from the lineage because it is too costly in terms of the high wage non-farm activity that she would have to sacrifice in order to cultivate it. However, the low type would take the land and its associated cultivation obligation, because the opportunity cost of her time is sufficiently low. A farmer's willingness to accept the requirement reveals that her return to off-farm work is low, and that she therefore needs the additional land to avoid poverty.

If the lineage head has access to the otherwise private information about some individuals' returns to off-farm work, perhaps because these individuals are socially or politically well-connected to the lineage leadership, then for these individuals, the land allocation can be made without the cultivation requirement. Both high- and low- types in this set efficiently fallow their land.

Further Empirical Work and Additions to Data Collection

The key empirical implication is that all plots under the control of an individual are treated similarly. Well-connected individuals about whom the lineage head has full information efficiently follow their entire portfolio of plots. Low types among the set of more isolated individuals reveal their ‘need’ by inefficiently cultivating land that - considered only from the viewpoint of technical productivity - should be fallowed. It is inconsequential what the source of that land is, it could either be the cultivator’s own land or the land provided by the lineage head.

Goldstein and Udry (2002) use our existing data to provide mixed evidence regarding the importance of this hypothesis for decisions regarding investment in land. Unfortunately, the relatively small sample size that we can work with limits the precision of our estimates. It is apparent that new data that identifies the individuals in each lineage and village who have direct influence on land allocation decisions would help resolve the issue. Combining that information with our existing data on flows of information between individuals should enable us to distinguish this hypothesis from other alternatives involving the notion that the security of one’s rights over a plot are determined by one’s social position in the village and lineage. The latter hypothesis is a plausible approximation to much of the literature on land tenure systems in West Africa (Berry; Otsuka and Quisumbing).

CONCLUSION

Iterative field research provides an opportunity, within the context of a unified project, for a flow of work between different research methods: detailed qualitative observation and conversation, theorizing, collection of data through surveys, statistical analysis. While this process requires a larger commitment of resources for a given sample size (or panel duration) than many alternative approaches to research in development economics, it is uniquely valuable in those instances in which the hypotheses to be examined are relatively open-ended or in which the economic environment is not well understood. Direct involvement in field research provides rich opportunities for being surprised, and these surprises can lead

to important insights.

A hopeful development over recent years is that field research of many different types has become less costly. As a consequence, approximately one-third of the microeconomics papers at the 2002 NEUDC conference were based on data collected by an author. It is apparent, however, that there remain important unexploited gains to coordination that could improve the quality and lower the cost of various forms of field research. We can and should make it possible to realize many of the benefits of field research at substantially lower cost.

First, we should develop outlets for discussions of field research methods, including the rich variety of mundane but essential tasks surrounding the data collection in developing countries. Part of this need might be met through publication on the web, as with the “Fieldwork in Development Economics” site mentioned above, but a refereed outlet in one of the field journals would be particularly valuable.

Second, we should strengthen research collaborations between developed and developing country researchers and institutions. These connections are essential for developed country researchers doing fieldwork, and can provide important support for developing country researchers.

Third, we should intensify our work with data collection agencies to encourage collaboration, as for example has been so successful with the Progressa program in Mexico.

Finally, we should facilitate exchanges between major centers of graduate education in development economics to improve graduate students’ knowledge of and access to ongoing research projects in developing countries.

References

- Bandiera, Oriana and Imran Rasul. "Social Networks and Technology Adoption in Northern Mozambique." Mimeo, LSE, 2002.
- Berry, Sara. No Condition is Permanent: The Social Dynamics of Agrarian Change in Sub-Saharan Africa. Madison: University of Wisconsin, 1993.
- Deaton, Angus. The analysis of household surveys: A microeconomic approach to development policy. Baltimore: Johns Hopkins University, 1997.
- Goldstein, Markus and Christopher Udry. "Gender, Land Rights and Agriculture in Ghana." Mimeo, Yale University, 2002.
- Goldstein, Markus and Christopher Udry. "Gender and Land Resource Management in Ghana." Mimeo, Yale University, 1999.
- Grosh, Margaret and Paul Glewwe. Designing Household Survey Questionnaires for Developing Countries. Washington, D.C.: The World Bank, 2000.
- Pender, John L. "Discount Rates and Credit Markets: Theory and Evidence from Rural India." Journal of Development Economics, 1996, (50), pp. 257-96.
- Place, Frank and Peter Hazell. "Productivity Effects of Indigenous Land Tenure Systems in Sub-Saharan Africa." American Journal of Agricultural Economics, 1993, (75), pp. 10-19.
- Quisumbing, Agnes, Jonna Estudillo and Keijiro Otsuka. "Gender Differences in Land Inheritance and Schooling: A Comparative Study of Rural Communities in the Philippines, Indonesia and Ghana." Mimeo: International Food Policy Research Institute, 2001.
- Townsend, Robert M. "Financial Systems in Northern Thai Villages." Quarterly Journal of Economics, 1995, (110), pp. 1011-46.

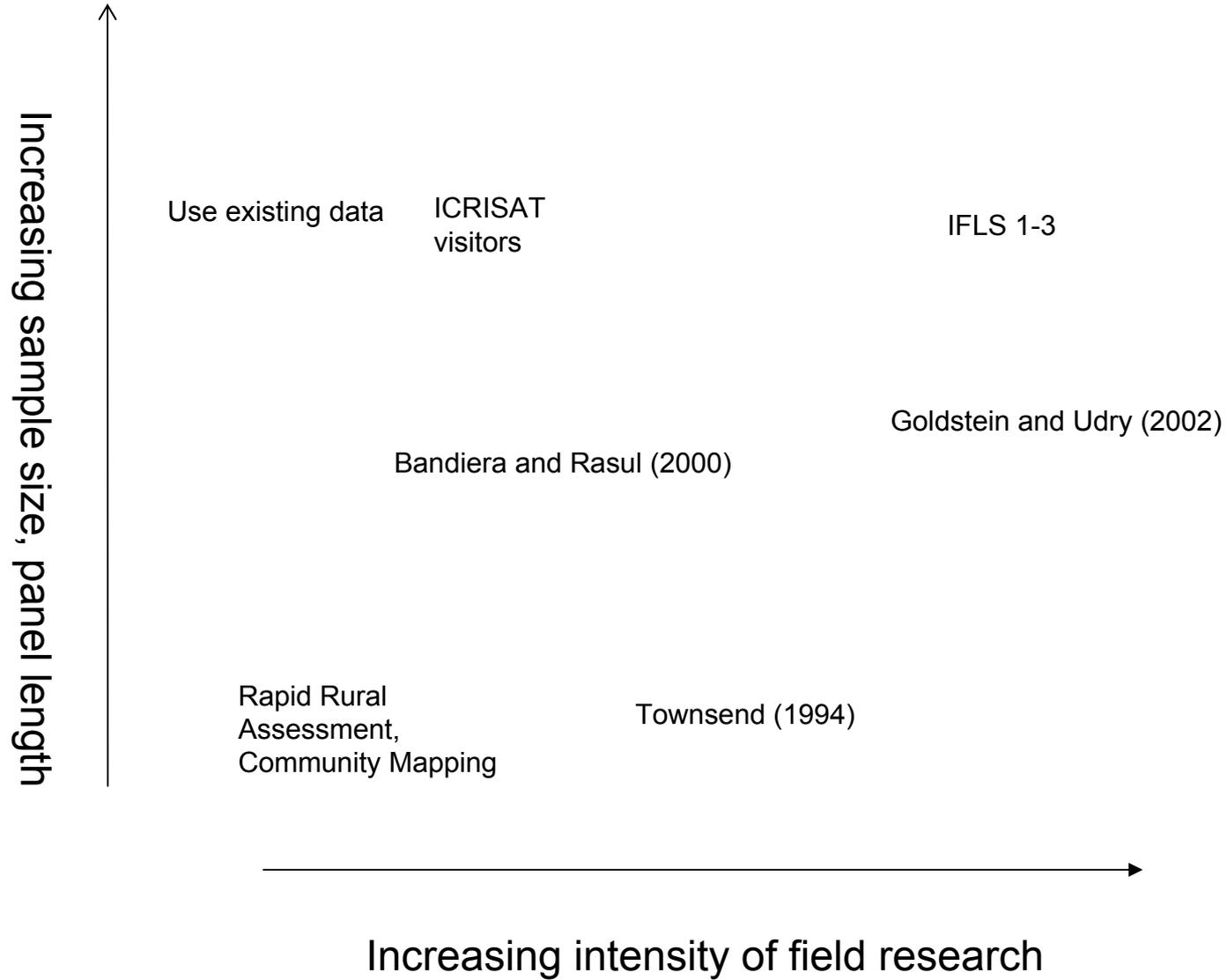


Figure 1