Sampling design for an integrated socioeconomic and ecological survey by using satellite remote sensing and ordination

Michael W. Binford*, Tae Jeong Lee†, and Robert M. Townsend‡§

*Department of Geography, University of Florida, Gainesville, FL 32611; †Department of Economics, Yonsei University, Wonju, Korea 220-710; and ‡Department of Economics, University of Chicago, Chicago, IL 60637

Communicated by Thomas J. Sargent, New York University, New York, NY, April 28, 2004 (received for review April 10, 2003)

Environmental variability is an important risk factor in rural agricultural communities. Testing models requires empirical sampling that generates data that are representative in both economic and ecological domains. Detrended correspondence analysis of satellite remote sensing data were used to design an effective low-cost sampling protocol for a field study to create an integrated socioeconomic and ecological database when no prior information on ecology of the survey area existed. We stratified the sample for the selection of tambons from various preselected provinces in Thailand based on factor analysis of spectral land-cover classes derived from satellite data. We conducted the survey for the sampled villages in the chosen tambons. The resulting data capture interesting variations in soil productivity and in the timing of good and bad years, which a purely random sample would likely have missed. Thus, this database will allow tests of hypotheses concerning the effect of credit on productivity, the sharing of idiosyncratic risks, and the economic influence of environmental variability.

Thailand | sample stratification | correspondence analysis | economic risk | soil

We use satellite land-cover information, specifically Landsat Thematic Mapper (TM) reflectance data, and the results of detrended correspondence analysis (DCA) applied to individual political units to determine whether or not to stratify the sampling of an otherwise random sample of villages. Because surveys typically have sample sizes that are a relatively small fraction of the sampling universe, purely random selection may miss important ecological variation.

An efficient sampling design should capture interesting variations among villages to be surveyed in a least-expensive way. Our sampling protocol is based on analysis of Landsat TM imagery of Thailand, which defines one domain of environmental variation. The goal was to select subcounties, or tambons, as the sampling unit for the usual logistical advantage of a clustered survey. (A tambon is a political division smaller than a county that consists of 4–12 villages). A single Landsat scene consists of ≈6,000 × 6,000 30-m square "pixels," each of which reflects energy from multiple bands of the electromagnetic spectrum. By using pattern recognition methods, each pixel is classified as belonging to one of 25 spectral classes interpreted as land-cover classes. The frequency distribution of the spectral classes in each subcounty is extracted.[¶] DCA is applied to estimate a few latent factors that drive most of the observed variations in land-cover classes within subcounties. Based on these "site scores," this study identifies patterns of clusters of subcounties in three of four provinces, motivating stratified samples in those provinces. Prior knowledge indicates that the stratification is determined by whether subcounties are forested or nonforested.

The subsequent integrated socioeconomic and ecological survey confirms the usefulness of the stratification. Soil samples show that variations in soil productivity are captured successfully by forest versus nonforest strata. The spatial ecological variation encoded in the site scores is also related to the timing of good and bad shocks as recorded in (retrospective) field interviews. Thus, the stratified sampling is useful for constructing a dataset that forms the input to the next phase of the research: to test economic models of credit arrangements and their implications for income growth and inequality and to examine the relationships between environmental variation and economic conditions.

Economic Models and Desired Features of the Sample

Economic models suggest testable hypotheses about growth of aggregate income, inequality in the distribution of income, and uneven access to financial services of various types. The models determine what data would be necessary for such tests.

Although technology may be a way to overcome poor environments, technology also has budget implications. For example, synthetic fertilizers can easily be applied in an area of poor soils to increase crop yields, but synthetic fertilizers cost money [in some cases, fertilizers have become the largest direct production cost (1)]. Many farmers cannot afford to buy fertilizers without loans. One class of models suggests that investment and/or credit may be limited or nonexistent for small businesses or cash-intensive farmers. Banerjee and Newman (2), Aghion and Bolton (3), and Bernhardt and Lloyd-Ellis (4) suggest that investment and credit will vary with wealth because of either limited collateral or weaker incentives for working.

The general weakness of these models in applying them to Thai rural areas is that they postulate a mostly uniform production function, $y_i = f(k_i, n_i, m_i)$, for household i, a mapping of inputs of capital k_i, labor n_i, and raw material m_i to the output y_i, not incorporating environmental variation. Output of a given entrepreneurial farmer could be low not because wealth or assets are low (that is, not because credit-financed inputs are low) but rather because the land of that farmer is relatively unproductive. In short, tests of the credit–wealth nexus should control for the technology in use. Ideally a sample would contain measured variation in the technology; for example, a production function $y_i = f^j(k_i, n_i, m_i)$ allowing variation in soil productivity of types j.

This theory leads in turn to the design of socioeconomic instruments and environmental measurements for field study, focusing primarily on households and small businesses in rural or semiurban communities.

Another class of theoretical models of growth with inequality and uneven financial deepening emphasizes the role of risk and potentially limited means of reallocating that risk. Risk in agricultural communities is often environmental, e.g., floods, droughts, insect outbreaks, and soil degradation. Greenwood and Jovanovic (5) argue that access to commercial banks and the formal financial

Abbreviations: TM, thematic map(per); DCA, detrended correspondence analysis; CEC, cation exchange capacity; OM, organic matter content; FC, field capacity; FN, foliage nitrogen.

§To whom correspondence should be addressed. E-mail: rtownsen@uchicago.edu.

¶The procedure of assigning each pixel to 25 classes and counting the number of pixels belonging to each class for each tambon reduces the dimension of the matrix from 7 × no. of pixels to 25 × no. of tambons.

system allows sharing of idiosyncratic risks (risks specific to households, e.g., local pests) but that aggregate shocks, such as widespread drought, are not ameliorated. To test whether such risk-sharing is better for those in the financial system (e.g., those with access to commercial credit), one needs a sample in which not all shocks are aggregate shocks, i.e., not everyone has bad or good yields at the same time.

For example, suppose that the production function has the form of $y_i = \varepsilon_i \cdot \xi \cdot f^j(k_i, n_i, m_i)$, where the ε_i are idiosyncratic shocks that affect only household i and ξ is the aggregate shock that affects everyone in the same region. The other terms in the equation are as above.

From this specification of the production function, we can derive the exact regression equation to test the "full" risk-sharing hypothesis $\Delta c_i = \Delta I + \beta \Delta I_i + e_i$, where Δc_i is change in the consumption of household i, ΔI is a latent factor capturing common variation in consumption over households induced by variation in aggregate shock ξ, and ΔI_i is a latent or measured factor capturing idiosyncratic variation in income of household i induced by idiosyncratic shock ε_i. First differences have the advantage of eliminating individual fixed effects. In this test, coefficient β should be zero if variation in ε_i were completely shared. If a sample were stratified by variations in land cover (e.g., crops or soil characteristics), then one would be more likely to collect data in which not all variation in income is attributable to a common shock or fixed effect. (Note that for this particular application, crop choice and input use can be endogenous; all that is needed is variation in net incomes.) Thus, one could test whether $\beta = 0$ (for this test and others in levels see refs. 6–9).

The fundamental question for the sampling protocol to determine based on prior information is whether there are any areas that are ecologically distinct from other areas. If there are distinct environmental zones within a sampling area, then some, if not all, of the zones should be sampled to assure a representative sample of each. We want to ensure that not all variation is associated with common factor ΔI, or, in the case of heterogeneity in production, that there are sufficient representatives from each subset of zones to estimate the various technologies, f^j. Samples stratified by zones and then chosen randomly within zones are a feasible option. If, on the other hand, there are no evidently distinct environmental zones, then a totally random sample is the simplest to administer and the most likely to achieve representative variability. Prior information on both environmental and economic variables is necessary to determine the need for stratification. Unfortunately, however, there are no existing land-cover maps at the provincial or larger scale that can be used, so we must create our own land-cover maps by using satellite imagery.

In the process of designing the sampling strategy, we were reminded that there is little that is free from human intervention. The application of fertilizers can affect soil quality, which in turn can alter vegetation and, thus, reflectance. More generally, the distribution of earth surface covered by water can be affected by the construction of dams, canals, and reservoirs. Thus, the sampling scheme stratifies on areas that are likely to be different environmentally, which needs to be confirmed in field research, and is one reason why we restrict ourselves to conservative measures of fertility and conservative sampling strategies. Fortunately, tests of the risk-sharing hypothesis are not sensitive to the endogeneity of income.

Derivation of Land-Cover Classes

Landsat TM scenes were acquired from the data archives of the U.S. Geological Survey's EROS Data Center. TM scenes cover 180×175-km regions with the resolution defined on the ground by 30-m pixels and measure each pixel's reflectance in seven bands of the electromagnetic spectrum (blue, green, red, near infrared, midinfrared, a second midinfrared, and emitted energy in a thermal infrared band). Scenes were chosen to be as cloud-free as possible and to cover as much of the spatial extent of the preselected provinces[‖] as possible. We used TM scenes from November 25, 1990, for the province of Sisaket (World Reference System-2, path 127, rows 49 and 50) and January 20, 1989, for the province of Lopburi (World Reference System-2, path 129, rows 49 and 50). These scenes were the most cloud-free TM scenes of the study areas of the entire data archive (see Fig. 1 for the composite images of Lopburi province with tambon boundaries overlaid). Tambon boundaries were identified on existing commercial maps (1:50,000 scale) and digitized onto vector geographic information system files by using MAPINFO PROFESSIONAL 6.0. For the other two provinces, Chacherngsao and Buriram, one would have to use scenes from two different orbital paths, with data acquired on different days. Unfortunately, none of the combinations of the available data allowed us to conduct adequate image normalization and mosaicing techniques for these two provinces because of the different phenological periods (dates of different vegetation development levels), levels of cloud cover, and other problems with atmospheric conditions. Thus, the analysis of imagery here is restricted to Sisaket and Lopburi.

We applied the unsupervised land-cover classification algorithm ISODATA (10) to each of the two provinces with usable satellite imagery by using the software ERDAS IMAGINE, Versions 7.4, 8.3 and 8.4 (Leica, Deerfield, IL). ISODATA is a self-organizing pattern-recognition method that iteratively groups imagery pixels by their similarities in multidimensional space defined by reflectance values in the seven radiometric bands recorded by the TM instrument. As a result of this analysis, we can assign each pixel on the ground of each province into a land-cover class. A class is defined to be an entity such that members of a class are quantitatively similar to one another in terms of the seven-dimensional (band) reflectance spectrum and are different from members of other classes. After experimentation to determine the optimum number of classes, 25 spectral classes were most representative of spatial variability, which is a somewhat arbitrary number of classes. Specifying >30 land-cover classes in the ISODATA function resulted in many of the classes containing a very small number of cells, and <20 classes resulted in classes with large variances that overlapped other classes in most of the bands.

We identified the land-cover of each tambon separately by overlaying the digitized tambon boundaries onto the results of the land-cover classification. Fig. 1 also shows a larger-scale image of several Lopburi tambon boundaries superimposed on the land-cover data. We counted the number of pixels belonging to each land-cover class within each tambon. In an ecological survey, these land-cover composition data are analogous to species composition data; for this survey, however, they are analogous to land-cover classes in subcounties. Fig. 2 *Right* shows the frequency distribution of 25 land-cover classes in the four tambons shown in *Left*. Silatip seems to be somewhat different from the other three because it has little of land-cover classes 23 and 24 (probably recently harvested paddy fields denoted by light yellow and purple on the image) but much more of classes 2, 6, and 10 (perhaps unharvested agricultural fields denoted by shades of green on the image). Ban Mai Samakkee and Chai Narai are somewhat similar to one another. Yang Rak represents another outlier because of the preponderance of land-cover classes 23, 24, and 14 (possibly paddy fields ready for harvest and denoted by brown on the image). Yang Rak and Silatip are quite

[‖]Funding from the National Institute of Child Health and Human Development and the National Science Foundation was in place with precommitment to survey four provinces: two provinces, Chacherngsao and Lopburi, in the central region and two provinces, Buriram and Sisaket, in the northeast region.

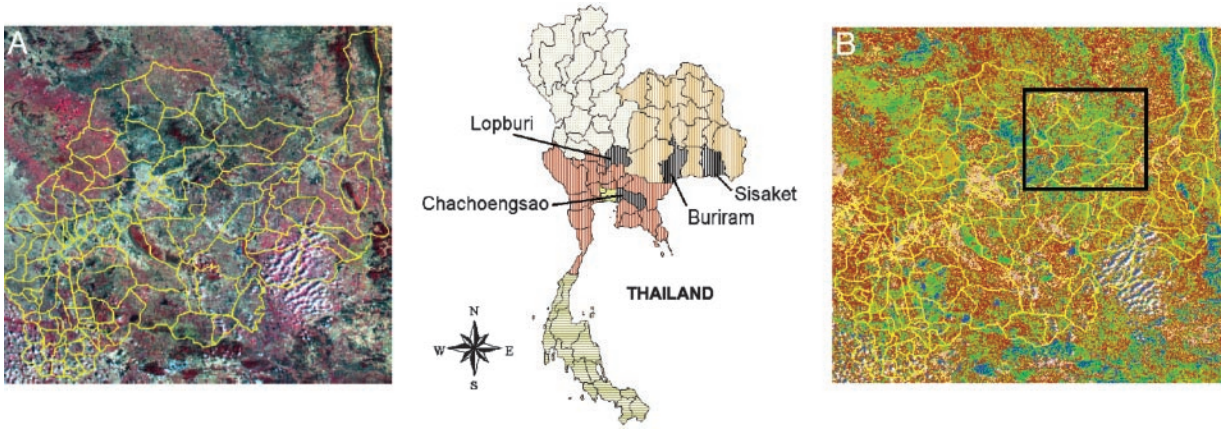

Fig. 1. Composite images for Lopburi with subcounty boundaries overlaid. The composite image was created by assigning the reflectance values for bands 4 (near infrared), 3 (red), and 2 (green) to the red, green, and blue color components of the image.

similar in the relative abundances of classes 6, 10, 12, and 14, which are dominant classes.

Ordination: DCA. By applying DCA (11) to the observed tambon level distribution of land-cover classes, we estimated the latent factors associated with each tambon, called site scores. DCA is a technique often used by ecologists to identify latent factors that drive observed variation in the composition of species across different sites. Here, tambon is the unit for sites, each of the 25 land-cover classes is seen as a species, and the number of pixels belonging to a particular land-cover class in a given tambon is treated as the observed population of a species in that tambon. The DCA algorithm constructs an iterative weighted average, starting with arbitrary initial site scores, where weights are the land-cover composition data. If the distribution of land-cover composition data are unimodal, the iteration is guaranteed to converge to a unique set of values and assign site scores for each tambon. By making use of an orthogonality condition, repeated use of the same algorithm can generate as many latent factors (site scores) as we want.

Because we want to know the within-province variation in the environment to determine the selection of subcounties for each province, we applied DCA to each province separately. (Note that an identified land-cover class in one province from one image is not comparable to the land-cover class in the other province.) The software used was CANOCO 3.1 and CANOCO for Windows, Version 4.0 (12).

Results of DCA Analysis, Sampling, and the Survey. Taking several precautions is necessary when designing the sampling strategy based on the DCA results. First, it is difficult to know what the latent variables (DCA site scores) represent without ground-survey information. Second, a stratified sample requires reweighting if the analysis is to be taken as representative of the larger universe. Third, a purely random selection is most likely to deliver a database that allows for multiple uses, some of which are not envisioned in advance. With these drawbacks in mind, we limit ourselves to a conservative strategy, i.e., purely random selection with stratification based only on salient DCA factor scores with a known interpretation.

The DCA site scores of all tambons in Lopburi are shown in Figs. 3 and 4, and the scores for Sisaket are shown in Figs. 5 and 6. The first two latent factors (or site scores) jointly account for 70% and 76%, respectively, of the total variation of the land-cover distribution of the two provinces. Lopburi tambons are scattered evenly in DCA factor space. The variation is somewhat continuous, and no clusters can be considered as outliers and identified *a priori*. Under our conservative strategy, this result suggests that tambons in Lopburi should be sampled with a simple, entirely random design.

In Fig. 5, a cluster of outlier tambons in DCA Axis 1 stands out. These six outlier tambons were all located in the southeastern part of the province on the Cambodian border and had large areas of forest. This pattern suggests stratification based on forested versus nonforested tambons, then sampling to assure

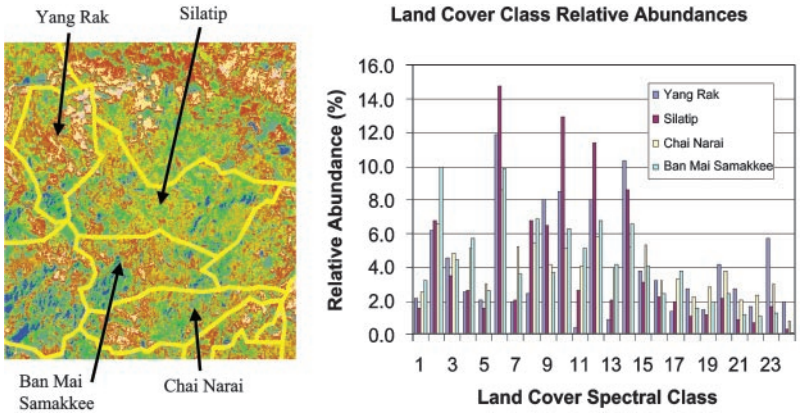

Fig. 2. Twenty-five land-cover class images of several individual subcounties in Lopburi. The 25 land-cover classes are indicated by distinct colors but are not named.

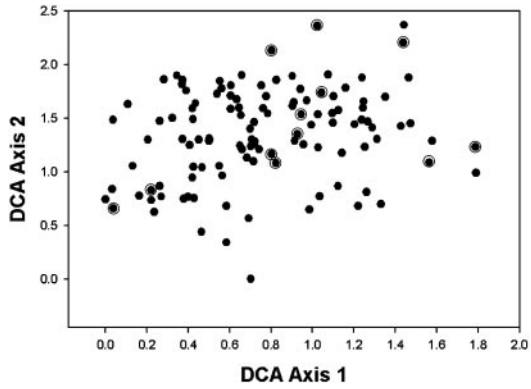

Fig. 3. Plot of first and second DCA site scores for Lopburi. The circled dots in the plot show DCA site scores for the sampled tambons.

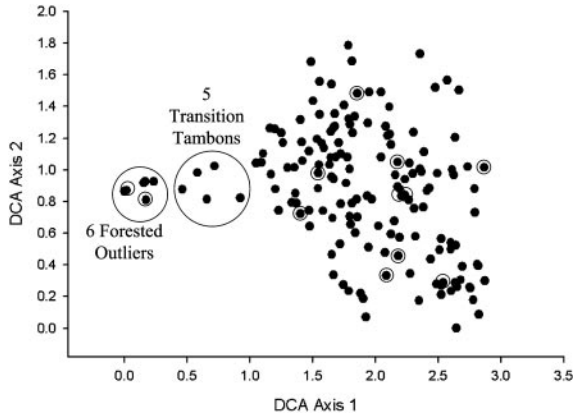

Fig. 5. Plot of first and second DCA site scores for Sisaket. The circled dots in the plot show DCA site scores for the sampled tambons.

representation of each of these two strata. Preliminary field research confirmed the existence of forest on the border and failed to uncover other salient variation that might account for different DCA scores. Although we did not use Landsat data to analyze land cover in Buriram and Chacherngsao, we identified forested areas from both composite images made from the Landsat data and 1:50,000 scale Royal Thai topographic quadrangles. We also chose a subset of forested tambons in the two provinces accordingly.

To summarize, in Lopburi, where there is no forested area, all of the 12 tambons were selected purely at random. The probability of each tambon being selected is 1/no. of tambons in Lopburi, or 12/117. In Sisaket, Buriram, and Chacherngsao, two tambons were selected at random from the set of forested tambons, or 2/6, 2/8, and 2/4, respectively. The remaining 10 nonforested tambons were selected with probabilities 10/153, 10/160, and 10/85, respectively.

Four villages were randomly selected from each of the chosen tambons, and 15 households were then randomly drawn from each of the selected villages. Trained enumerators interviewed not only all of the households chosen in the sample but also all of the headmen and all of the managers of the formal and informal village-level organizations of the chosen villages. In addition, soil samples, vegetation samples, photographs, and plant community descriptions were taken for a plot belonging to each of 10 households randomly chosen from the 15 households of each village. See the Townsend Thai Project web site (http://cier.uchicago.edu/intro.htm) for detailed descriptions of the chosen sample; survey instruments, including the 1997 text on the questionnaire design; and the logistics of the survey.

Evaluation of Sampling Design

In this section, we evaluate with the gathered data whether the stratified sample has delivered a database desirable for our integrated socioeconomic and ecological research. The statistical analyses in this section were conducted with SAS 8.0 and STATA, Versions 5.0 and 6.0 for Windows.

Soil Variation, the Chosen Sample, and DCA Site Scores. Soil characteristics, such as cation exchange capacity (CEC), organic matter content (OM), field capacity (FC), pH, and foliage nitrogen (FN), are with the exception of FN conservative (i.e., they do not change; the properties are conserved or they change very slowly with disturbances, such as agricultural activities or climate change) indicators of soil fertility, and higher values of each variable generally indicate more fertile soil (13). We analyzed the soil samples to measure those characteristics. For laboratory methods of soil analysis, see refs. 14–16. Soil samples were analyzed in the laboratory at the Department of Soil Science, Kasetsaat University, under the supervision of Professor Irb Keorhumramne.

These data allow us to assess how well the forest versus nonforest

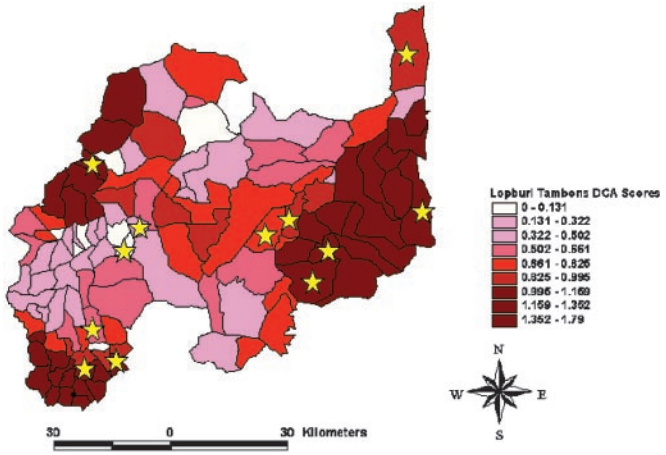

Fig. 4. TM of first DCA site scores for Lopburi. The stars on the TM show the geographic location of the sampled tambons.

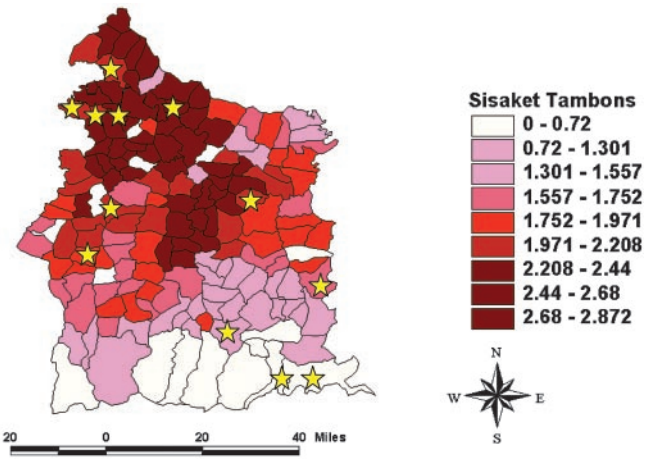

Fig. 6. TM of first DCA site score for Sisaket. The stars on the TM show the geographic location of the sampled tambons.

Table 1. Soil characteristics, forest dummies, and DCA site scores

	pH		CEC		OM		FC		FN	
	Without	With	Without	With	Without	With	Without	With	Without	With
All four provinces										
Region dummy	0.85*	0.31†	21.89*	24.05*	1.82*	2.71*	14.75*	17.88*	0.89*	0.32
Forest dummy	0.05	0.71*	−9.16*	−29.37*	−0.29*	−1.72*	−6.64*	−25.25*	0.52*	1.17*
F statistic for tambon dummies	—	72.79*	—	67.56*	—	30.38*	—	66.73*	—	13.47*
Constant	5.21*	4.49*	6.01*	33.74*	0.65*	0.69*	14.48*	21.22*	1.97*	1.98*
Adjusted R^2	0.130	0.720	0.401	0.791	0.482	0.716	0.478	0.816	0.191	0.402
Sisaket										
First DCA site score	0.06	−0.11†	−0.46*	−0.83*	−0.12*	−0.20*	−0.27	−1.51*	−0.14*	−0.07
Second DCA site score	0.43*	0.11	0.97‡	−0.17	0.21*	0.04	2.95*	0.58	0.53*	0.125
F statistic for tambon dummies	—	4.98*	—	20.56*	—	15.44*	—	20.93*	—	9.72*
Constant	4.61*	6.68*	2.98*	4.37*	0.70*	0.90*	10.48*	14.37*	1.61*	1.94*
Adjusted R^2	0.037	0.134	0.035	0.303	0.050	0.260	0.303	0.303	0.068	0.207

For all four provinces, the following number of observations were made for each measurement: pH, 1,572; CEC, 1,573; OM, 1,573; FC, 1,573; and FN, 1,554. For Sisaket province, the following number of observations were made for each measurement: pH, 410; CEC, 410; OM, 410; FC, 410; and FN, 401. Without, without tambon dummies; with, with tambon dummies.
*, Statistical significance at 1% level; †, statistical significance at 10% level; ‡, statistical significance at 5% level.

stratification strategy captures variation in soil productivity. We can determine whether soil characteristics are consistently different between the forest and nonforest areas by including a forest dummy variable in the following regression equation:

$$S_i = a + b_1 Region_i + b_2 Forest_i + b'_T D_{Ti} + \varepsilon_i. \quad [1]$$

Here, S_i is a soil characteristic of the plot of household i; $Region_i$ is a region dummy equal to 1 if the plot of household i is in the central region and 0 otherwise; $Forest_i$ is a forest dummy equal to 1 if the plot of household i is in a forest area and 0 otherwise; and D_{Ti} are dummies for the tambon of household i. All of the regressions in this paper are done by the software INTERCOOLED STATA 7. This software automatically checks for muticollinearity when there is more than one dummy variable in the equation. When multicollinearity is detected, the software corrects the equation by dropping out some dummies. We confirmed that the software dropped some tambon dummies properly.

As shown in Table 1, soil characteristics are significantly related to forest dummies in all but one case. The exception has to do with pH and can be overturned by including tambon controls. The regression results indicate that CEC, OM, and FC are lower in forested tambons, so we may say that, in general, the forested tambons have less fertile soil. Although we do not report them here, these results do not change, even if we run the regressions separately for each province. Variations in soil productivity are captured successfully by forest versus nonforest strata.

To describe the relationship between DCA site scores and soil characteristics, we ran the following regression equation:

$$S_i = a + b_1 DCA1_i + b_2 DCA2_i + b'_T D_{Ti} + \varepsilon_i, \quad [2]$$

where $DCA1_i$ is the first and $DCA2_i$ is the second DCA site score for the tambon of household i.

The regression coefficients indicate that in Sisaket the tambons with higher first site score, DCA1, tend to have lower values of CEC, OM, FC,** and FN (Table 1). The second site score, DCA2, on the other hand, is significantly positively related to pH, CEC, OM, FC, and FN, although the significance of DCA2 is lost with tambon controls. More generally, as is evident from the TM in Fig. 5, DCA1 increases as one moves from south to north. Now with the soil data in hand, we might contemplate more extensive stratification along the DCA1 axis.

In Lopburi, although we do not report it in detail here, DCA1 is positively and significantly related to all measures of fertility and DCA2 is positively related, except for CEC. As can be seen from the TM in Fig. 3, DCA1 values are geographically clustered. Again, with the soil fertility interpretation in hand, one might have contemplated some kind of stratification on DCA1 values. We emphasize, however, that we did not have the soil measurements before choosing the sample and conducting the survey.

Idiosyncratic Shocks, the Chosen Sample, and DCA Site Scores. Another major objective of the overall research program is to test hypotheses regarding an optimal allocation of risk bearing. As we described earlier, the sample must exhibit a variety of shocks, so that not all shocks are aggregate shocks. To evaluate whether our sample has the necessary mix of shocks, we examined whether the timing of good or bad shocks to villages is related to spatial ecological variation captured by the DCA scores. Answers by village headmen to questions about the history of good and bad years were used as dependent variables in separate regressions. The questions were, Which was the best year in the last 5 years? The worst year in the last 5 years? The best year ever? and The worst year ever? Responses were coded as the Thai year [e.g., *anno Domini* 2000 is 2543 BE (Buddhist Era) in Thailand]. Of course, we did not have actual panel data to confirm the answers about past history.

We regress the timing of shocks onto forest dummies (1 = forested, 0 = nonforested) first by pooling all four provinces but controlling for regional fixed effects and then separately for each of the three provinces with forest areas. The regression results for Sisaket are reported in Table 2. We may infer from the regression coefficients that Sisaket's forest tambons tend to have had good and bad years within the last 5 years but that the worst historical events have happened closer to the present than the past. For Buriram, another survey province in the northeast region, regressions results (data not shown) indicate that forest dummies do matter there: Forested tambons in Buriram tend to have had the worse historical events more recently. However, for Chacherngsao, in the central region, forest dummies do not matter to the time of arrival of good and bad shocks.

Again, one may wonder whether the DCA site scores have anything to do with the timing of good and bad shocks in the survey areas. To describe the relationship of DCA site scores with the timing of good and bad shocks in the survey areas, good and bad years within the last 5 years or over historical memory are regressed

**Apparently, there is an inconsistency in the regression results with forest dummies and DCA site scores: coefficients on forest dummies indicate that the forest areas have less fertile soil, whereas coefficients on DCA1 indicate the contrary. Further research is required to resolve this inconsistency. Our conjecture as of now is that soil quality may have been influenced by human intervention, such as the application of fertilizers to compensate for poor soil.

Table 2. Timing of shocks, forest dummies, and DCA site scores for Sisaket province

	Best year in last 5 years		Best year ever		Worst year in last 5 years		Worst year ever	
Sisaket	Without	With	Without	With	Without	With	Without	With
Regression with forest dummies								
Forest dummy	−0.99*	−0.83	8.94†	6.583‡	−1.32‡	−0.50	14.81†	10.83‡
F statistic for tambon dummies	—	1.66	—	0.71	—	1.16	—	1.98
Constant	20.39†	2.33†	3.70†	6.67*	13.22*	3.00†	7.48†	5.50
Adjusted R^2	0.090	0.214	0.347	0.475	0.062	0.095	0.290	0.417
Regression with DCA scores								
First DCA site score	0.42*	0.58*	−3.15†	−3.46†	0.56‡	0.69	−5.72†	−9.00†
Second DCA site score	−1.12†	−0.92‡	−1.90†	−4.70‡	−0.50	−0.86	−0.16†	−7.66
F statistic for tambon dummies	—	0.44	—	1.97‡	—	0.90	—	2.17‡
Constant	2.43†	2.29†	12.63†	17.40†	2.21*	2.23‡	19.23†	33.57†
Adjusted R^2	0.295	0.199	0.213	0.346	0.063	0.043	0.222	0.369

For the regression with forest dummies, the following number of observations were made for each measurement: best year in last 5 years, 44; best year ever, 45; worst year in last 5 years, 46; worst year ever, 47. For the regression with DCA site scores, the following number of observations were made for each measurement: best year in last 5 years, 40; best year ever, 41; worst year in last 5 years, 42; worst year ever, 43. The number of observations in these regressions is not exactly 48 because some village headman refused to answer the question or answered that they had no particularly good or bad years. *, statistical significance at 5% level; †, statistical significance at 1% level; ‡, statistical significance at 10% level. Without, without tambon dummies; with, with tambon dummies.

onto the first and second DCA site scores for the two changwats of Lopburi and Sisaket by using region and tambon dummies as controls to define the added effect of land-cover itself. In Table 2, the regression results for Sisaket are reported. The results imply that the tambons with higher DCA1 and DCA2 scores had the best and worst years ever in the more remote past and, therefore, knowing the regional DCA patterns, that the best and worst shocks ever happened more recently in the forest tambons of Sisaket. In the regressions for Sisaket of the best and worst shocks in the last 5 years, DCA1 and DCA2 provide a significant explanation with exceptions. We may infer, with caution, that tambons with a higher DCA1 score are likely to have had good or bad years within the last 5 years more recently. Overall, the results are consistent with the regressions with forest dummies.

We have subsequently created an annual household panel for one-third of the original sample, from 1997–2001. By regressing household income changes onto DCA scores, we find that households in tambons with higher DCA2 values in Sisaket had a significantly higher income change from 1999–2000 and a significantly lower income change from 1997–1998, and, thus, vice versa for tambons with lower DCA2 values. These data indicate that even further initial stratification with the DCA scores would appear to have increased the likelihood of obtaining the desired idiosyncratic shocks in the sample.

On the contrary, similar regressions for Lopburi show that DCA site scores hardly have any significant relationship with the timing of shocks. This finding implies that in Lopburi the sample would not have captured idiosyncratic shocks even if we had stratified the sample according to DCA site scores.

Discussions and Conclusion

The analysis with the data gathered in the survey confirms that soil fertility is significantly different across forest and nonforest areas. Thus, the stratified sample succeeded in generating a database that allows us to dissect the wealth, credit, fertilizer, and income nexus as in the first group of models of growth and inequality described in *Economic Models and Desired Features of the Sample*. These data may be applied to a study of the contribution of financial institutions, such as village funds, the Bank for Agriculture and Agricultural Cooperatives, and commercial banks to household income, as well as the effects of the changing economic conditions on land-use and land-cover changes.

The timing of good and bad shocks was also significantly different between forested and nonforested tambons in the northeast region (the provinces of Buriram and Sisaket). If we had not stratified by the forest tambons, we believe we would have lowered the likelihood of generating a sample in which idiosyncratic shocks could be studied. Such a sample is necessary, of course, for testing the second group of models of income growth and inequality outlined in *Economic Models and Desired Features of the Sample*.

We could not have known the order of magnitude of these results on fertility and timing of shocks beforehand, when we designed the sample. However, others may be encouraged by our results and might contemplate bolder stratification strategies on the DCA axes. Here, we suggest the more modest point that satellite remote sensing data and analysis are a relatively inexpensive way to discover variation in the environment that can inform the design of the sampling scheme. When comparable remote sensing data are available for regions or an entire country, their use in sampling schemes can be even more comprehensive.

We thank Khun Sombat Sakuntasathien, Elizabeth Binford, Alan Kolata, the editor, and referees for their assistance and comments. This research was supported by National Institutes of Health Grant 5 R01 HD27638-10, National Science Foundation Grant SES-9987855, the Mellon Foundation, and the University of Chicago.

1. Matson, P. A., Parton, W. J., Power, A. G. & Swift, M. J. (1997) *Science* **277,** 504–509.
2. Banerjee, A. V. & Newman, A. F. (1993) *J. Polit. Econ.* **40,** 274–296.
3. Aghion, P. & Bolton, P. (1997) *Rev. Econ. Stud.* **64,** 151–172.
4. Bernhardt, D. & Lloyd-Ellis, H. (2000) *Rev. Econ. Stud.* **67,** 147–168.
5. Greenwood, J. & Jovanovic, B. (1990) *J. Polit. Econ.* **98,** 1076–1107.
6. Mace, B. (1991) *J Polit. Econ.* **99,** 928–956.
7. Cochrane, J. (1991) *J. Polit. Econ.* **99,** 957–976
8. Townsend, R. (1994) *Econometrica* **62,** 539–591.
9. Deaton, A. (1997) *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy* (The Johns Hopkins Univ. Press, Baltimore).
10. Ball, G. H. & Hall, D. J. (1965) *Novel Method of Data Analysis and Pattern Classification* (Stanford Res. Inst., Menlo Park, CA).
11. Jongman, R. H. G., Ter Braak, C. J. F. & van Tongeren, O. F. R. (1987) *Data Analysis in Community and Landscape Ecology* (Cambridge Univ. Press, Cambridge, U.K.).
12. ter Braak, C. J. F. & Smilauer, P. (1998), *CANOCO Reference Manual and User's Guide to Canoco for Windows: Software for Canonical Community Ordination (Version 4)* (Microcomputer Power, Ithaca, NY).
13. Weischet, W. & Caviedes, C. (1993) *The Persisting Ecological Constraints of Tropical Agriculture* (Blackwell, Oxford).
14. Cassel, D. K. & Nielsen, D. R. (1986) *Methods of Soil Analysis, Part 1: Physical and Mineralogical Methods*, ed. A. Klute (Soil Sci. Soc. Am., Madison, WI), 2nd. Ed.
15. Landon, J. R., ed. (1991) *Booker Tropical Soil Manual* (Pearson Higher Education, Longman, U.K.).
16. Lal, R. & Sanchez, P. A. (1992) *Myths and Science of Soils of the Tropics* (Soil Sci. Soc. Am., Madison, WI).