The empirical analysis is conducted at 3 separate levels of aggregation: across countries, virtual countries and pairs of contiguous cells.

The shapefile contained in the zip file has the following variables:

uniq_cnt25 = is the identifier of each 2.5x2.5 decimal degrees polygons used in the virtual country analysis.

point5_id = is the identifier for each 0.5x0.5 decimal degrees cell used in the pairs of contiguous cells analysis.

suit_new = is the suitability of land for agriculture at each 0.5 by 0.5 decimal degrees cell

pasture = is the share of each 0.5 by 0.5 cell allocated to pasture

maize = is the share of each 0.5 by 0.5 cell allocated to maize cultivation

sorghum = is the share of each 0.5 by 0.5 cell allocated to sorghum cultivation

allcrops = is the share of each 0.5 by 0.5 cell allocated to some type of cultivation

pop95 = is the population density in 1995 of each 0.5 by 0.5 cell.

The geographic and population density variables come from:
http://www.sage.wisc.edu/iamdata/

Cross-Country Analysis

All variables are available. See details in the respected do files, i.e. Tables1a-3a.do and Table3b.do.
Each do file performs the regressions presented in the Tables 1-3b.

Cross-Virtual Analysis

In the cross-virtual regressions I use as dependent variable the languages identified by the Ethnologue, Edition 15th.
The mapping of these languages comes from the World Language Mapping System (WLMS) which can be found at: http://www.worldgeodatasets.com/language/

To construct the number of language per virtual country intersect the language shapefile from WLMS with the shapefile found in the zip file. Then calculate the number of languages per virtual country, i.e. per uniq_cnt25.

The do file "Tables4-7b.do" describes and performs the regressions used in the virtual-country tables 4-7b.

Cross-Pair Analysis

The cross adjacent-cell analysis uses the Ethnologue Edition 15th.
The data on languages may be obtained from the World Language Mapping System available at: http://www.worldgeodatasets.com/language/

To generate the percentage of common languages spoken within pairs of adjacent regions, do the following. First, intersect the language shapefile from the WLMS with the shapefile found in the zip file.
After the intersection calculate the number of languages per cell (point5_id). Identify the 8 neighbors of each cell. Calculate the number of common languages spoken within a pair of adjacent cells. The percentage of common languages is the ratio of the Number of common languages within the pair to the total number of unique languages spoken within the same pair.

The do file "Tables8-9.do" describes the regressions employed in the cross-pair analysis presented in Tables 8 and 9.

Table 10

The regressions of Table 10 can be replicated using the "Tables1a-3a.do" for the cross-country sample, the "Tables4-7b.do" for the cross-virtual country sample and "Tables8-9.do" for the contiguous cells sample.

Table 11

The regressions of Table 11 can be replicated by using Table11.do file.