Caloric and Agricultural Suitability

Oded Galor*and Ömer Özak†

June 21, 2015

1 Caloric Suitability Indices (CSI)

The suitability of land for agriculture (Ramankutty, Foley, Norman and McSweeney, 2002) has become a standard control for the effect of geographical characteristics on comparative economic development. This measure, however, is rather crude and it does not capture the large variation in the potential caloric yield across equally suitable land. In particular, geographical regions that according to this measure are comparable in terms of their suitability for agriculture may differ significantly in their potential caloric output per hectare per year, reflecting the fact that land that is suitable for agriculture is not necessarily suitable for the most productive crops in terms of their caloric return.

In light of the importance of pre-industrial population density in the subsequent course of economic development, and the instrumental role played by caloric yield in sustaining and supporting population growth, it is rather apparent that this commonly used index is not well designed for properly capturing the effect of the suitability of land for agriculture on economic development.

Galor and Özak (2014) introduce a novel methodology that can be used to construct better measures of agricultural productivity based on caloric suitability. This note follows their method to rectify the deficiencies of the agricultural suitability index of Ramankutty *et al.* (2002) and introduce a set of novel indices of agricultural suitability: "The Caloric Suitability Indices" (CSI) that properly capture the variation in potential crop yield across the globe, as measured in calories per hectare per year. Moreover, in light of the expansion in the set of crops that are available for cultivation in the course of the Columbian Exchange Crosby (1972), the CSI indices provide a distinct measure for caloric suitability for the pre-1500 and the post 1500 era.

The CSI indices provide four estimates of caloric suitability for each cell of size $5' \times 5$ in the world:

*Department of Economics, Brown University. E-mail: Oded_Galor@brown.edu

†Department of Economics, Southern Methodist University. E-mail: ozak@smu.edu

1. The maximum potential caloric yield attainable given the set of crops that are suitable for cultivation in the pre-1500 period.
2. The maximum potential caloric yield attainable, given the set of crops that are suitable for cultivation in the post-1500 period.
3. The average potential yields within each cell attainable given the set of crops that are suitable for cultivation in the pre-1500 period.
4. The average potential yields within each cell attainable given the set of crops that are suitable for cultivation in the post-1500 period.

2 The Caloric Suitability Indices Data

These historical measures are constructed based on data from the Global Agro-Ecological Zones (GAEZ) project of the Food and Agriculture Organization (FAO). The GAEZ project supplies global estimates of crop yield and crop growth cycle for 48 crops in grids with cells size of $5' \times 5'$ (i.e., approximately 100 km^2).

The crops available are alfalfa, banana, barley, buckwheat, cabbage, cacao, carrot, cassava, chickpea, citrus, coconut, coffee, cotton, cowpea, dry pea, flax, foxtail millet, greengram, groundnuts, indigo rice, maize, oat, oilpalm, olive, onion, palm heart, pearl millet, phaseolus bean, pigeon pea, rye, sorghum, soybean, sunflower, sweet potato, tea, tomato, wetland rice, wheat, spring wheat, winter wheat, white potato, yams, giant yams, subtropical sorghum, tropical highland sorghum, tropical lowland, sorghum, white yams.

For each crop, GAEZ provides estimates for crop yield based on three alternative levels of inputs – high, medium, and low - and two possible categories of sources of water supply – rain-fed and irrigation. Additionally, for each input-water source category, it provides two separate estimates for crop yield, based on agro-climatic conditions, that are arguably unaffected by human intervention, and agro-ecological constraints, that could potentially reflect human intervention.

In order to capture the conditions that were prevalent during the pre-industrial era, while mitigating potential endogeneity concerns, the indices use the estimates of potential crop yield under low level of inputs and rain-fed agriculture – cultivation methods that characterized early stages of development. Moreover, the estimates of potential crop yield are based on agro-climatic constraints that are largely orthogonal to human intervention. Thus, these restrictions remove the potential concern that the level of agricultural inputs, the irrigation method, and soil quality, reflect endogenous choices that could be potentially correlated with individual preferences or institutional settings. Additionally, the choice of rain-fed conditions is further justified by the fact that, although some societies had access to irrigation prior to the industrial revolution, GAEZ's data only provides estimates based on irrigation infrastructure available during the late twentieth century

The FAO dataset provides for each cell in the agro-climatic grid the potential yield for each crop (measured in tons, per hectare, per year). These estimates account for the effect of temperature and moisture on the growth of the crop, the impact of pests, diseases and weeds on the yield, as well as climatic related "workability constraints".

In order to better capture the nutritional differences across crops, and thus to ensure comparability in the measure of crop yield, the yield of each crop in the GAEZ data (measured in tons, per hectare, per year) is converted into caloric return (measured in millions of kilo calories, per hectare, per year). This conversion is based on the caloric content of crops, as provided by the United States Department of Agriculture Nutrient Database for Standard Reference. Using the estimates of the caloric content for each crop in the GAEZ data (measured in kilo calories per 1g), a comparable measure of crop yield (in millions of kilo calories, per hectare, per year) is constructed for each crop.

Based on these estimates Galor and Özak (2014) construct the maximum potential caloric yield estimate they use in their paper. Here varios additional indices of caloric suitability are constructed and presented. First, for each cell the average caloric yield across all available crops pre- and post-1500CE is computed. Second, for each cell the total caloric yield across all available crops pre- and post-1500CE is computed. Finally, the analysis assigns to each cell the highest potential yield among the available crops pre- and post-1500CE. Additionally, for each caloric index raster the same index is constructed including and excluding cells where no calories can be produced or for averages the crops without caloric output are excluded.[1]

Thus, the research constructs for each type of index, namely *Average*, *Total* and *Maximal* Caloric Suitability, four sets of grids: 1. Caloric Suitability pre-1500CE (without zeros) 2. Caloric Suitability pre-1500CE (with zeros) 3. Caloric Suitability post-1500CE (without zeros) 4. Caloric Suitability post-1500CE (with zeros)

These grids can be used to assess the exogenous effect of agricultural potential on various economic and social outcomes. The next section shows how it can be done and compares with another measure of agricultural suitability.

3 Download Options for Caloric Suitability Indices

The **Caloric Suitability Indices** are available for download as GeoTiff rasters for the whole world and as `STATA` and `comma separated values` files with country-level measures. The data can be downloaded at http://ozak.github.io/Caloric-Suitability-Index/ as a zip file, or individually. The links below can be used to

[1]In particular, this means, that if for a cell c, n of the 48 crops in the FAO GAEZ data are not suitable for the production of calories, in that cell c only $48 - n$ crops will be used in the computations.

download (or you can fork the associated Github repository which contains also an IPython notebook with the computations of this paper).

- All files (zip)[2]
- Pre-1500CE:
  - Average Calories[3]
  - Average Calories (No Zeros)[4]
  - Maximum Calories[5]
  - Maximum Calories (No Zeros)[6]
- Post-1500CE:
  - Average Calories[7]
  - Average Calories (No Zeros)[8]
  - Maximum Calories[9]
  - Maximum Calories (No Zeros)[10]
- Country-level Data:
  - Stata Format
  - CSV Format

If you use the data, please cite:

Oded Galor and Ömer Özak, 2014. "The Agricultural Origins of Time Preference," NBER Working Papers 20438, National Bureau of Economic Research, Inc..

4 Caloric Crop Suitability and Agricultural Suitability

This section plots the various Caloric Suitability Indices constructed following Galor and Özak (2014) and introduced in the previous section. Additionally, it compares them to the agricultural suitability index of Rammankutty, Foley, Norman, and McSweeney (2001).

[2]https://drive.google.com/uc?export=download&id=0By-h7HPv1NhVM1g5aW81TzVRWjQ
[3]https://drive.google.com/uc?export=download&id=0By-h7HPv1NhVeEhsRmdRWkFJX2M
[4]https://drive.google.com/uc?export=download&id=0By-h7HPv1NhVcHgxa1EyOEpURUk
[5]https://drive.google.com/uc?export=download&id=0By-h7HPv1NhVR2dDUm5fU2lMN2c
[6]https://drive.google.com/uc?export=download&id=0By-h7HPv1NhVUzVDTXBST3d4YlE
[7]https://drive.google.com/uc?export=download&id=0By-h7HPv1NhVT05GNGtaZk13S2M
[8]https://drive.google.com/uc?export=download&id=0By-h7HPv1NhVcHVJcmgtb09FTXM
[9]https://drive.google.com/uc?export=download&id=0By-h7HPv1NhVR2ZDemhYd1hqZms
[10]https://drive.google.com/uc?export=download&id=0By-h7HPv1NhVajhjbVcyakFYMHc

4.1 Post-1500CE Caloric Suitability Indices

Below Figure 1 shows the plots of the 4 rasters for the post-1500CE period.

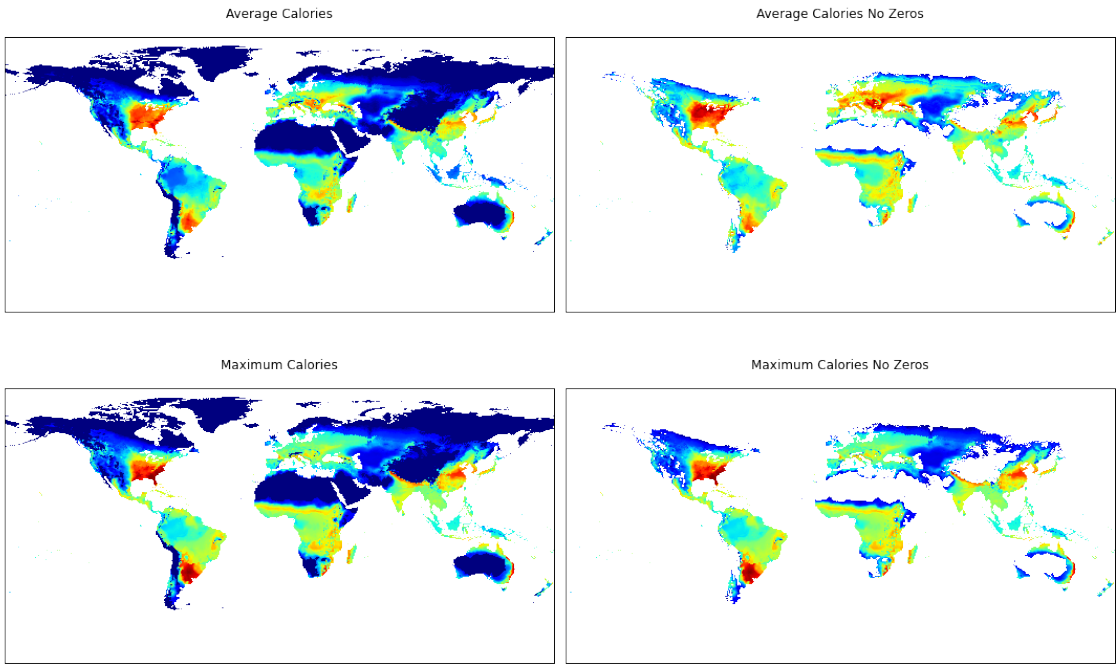


Figure 1: Post-1500CE Caloric Suitability Indices (CSI)

4.2 Pre-1500CE Caloric Suitability Indices

Figure 2 shows the plots of the pre-1500CE CSI data.

5 Agricultural Suitability data

Finally, Figure 3 shows the agricultural suitability data of Ramankutty *et al.* (2002), including the climatic and soil components of their index.

6 Differences between both data sets

6.1 Caloric Suitability Indices of Galor and Özak have finer resolutions

The Ramankutty *et al.* (2002) data is constructed in grids of $0.5° \times 0.5°$ with 360 rows and 720 columns, while the Galor and Özak (2014) has a resolution of $5' \times 5'$ with 2160 rows and 4320 columns. So, each cell in the Ramankutty *et al.* (2002) data is equivalent to 36 cells in the Galor and Özak (2014) dataset. This

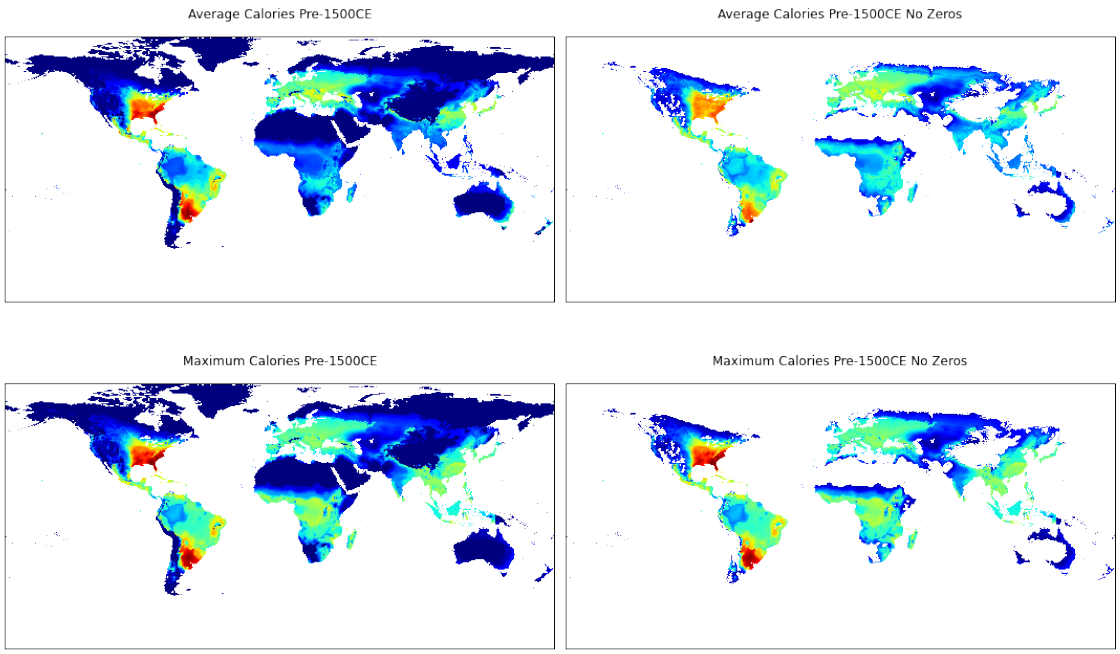


Figure 2: Pre-1500CE Caloric Suitability Indices (CSI)

means one can work at much smaller scales. Additionally, less measurement error will be generated when extracting data for countries or smaller regions.

6.2 Caloric Suitability Indices are exogenous

The CSI data is based on agro-climatic conditions that are mostly orthogonal to human intervention. On the other hand, as can be seen in the figures, most of the variation in the *Agricultural Suitability* data comes from the *Soil* component and not from the *Climatic* component. Thus, the use of the Ramankutty *et al.* (2002) index in most economic research might be problematic, since for most research questions, especially in comparative development and long-run growth, the measure might be affected by human intervention.

6.3 Caloric Suitability Indices have temporal variation

By exploiting the Columbian Exchange, the CSI data capture the changes in productivity generated by the introduction of new crops. This provides temporal variation that can be used in the empirical analysis of economic development. No such variation is present or can be constructed for the agricultural suitability data of Ramankutty *et al.* (2002).

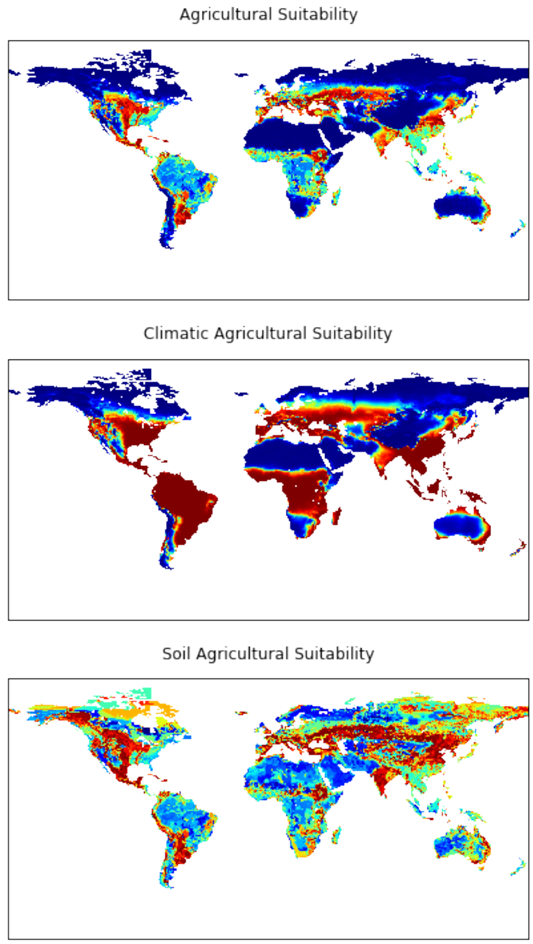
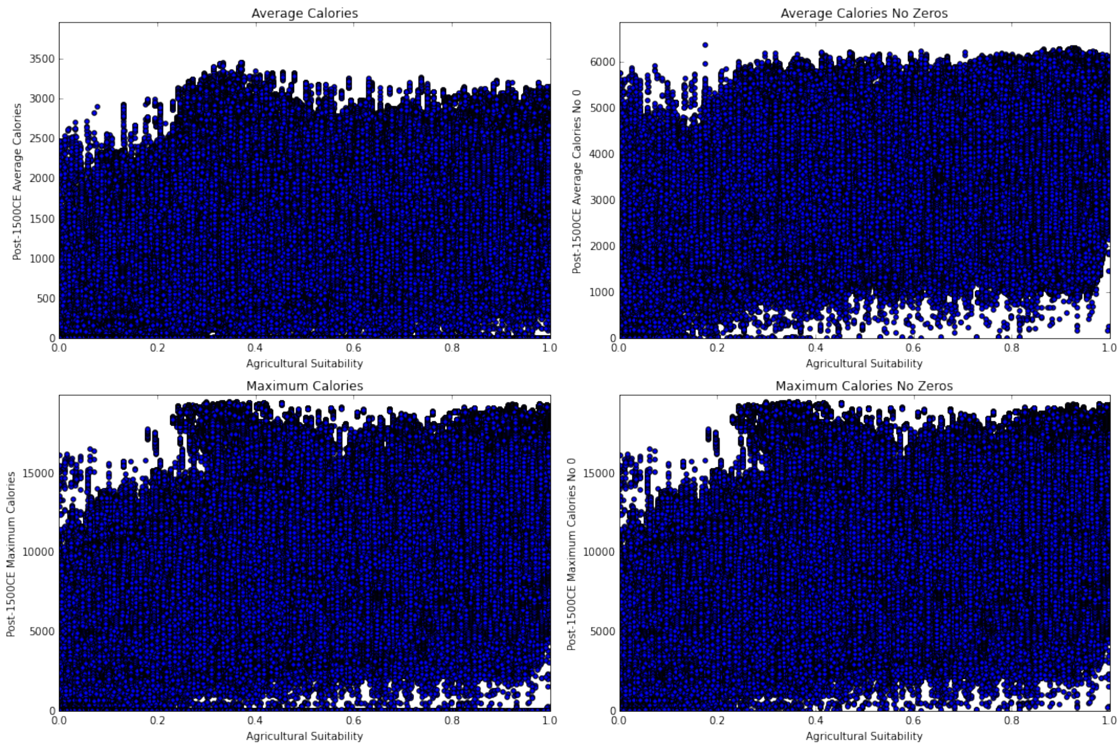


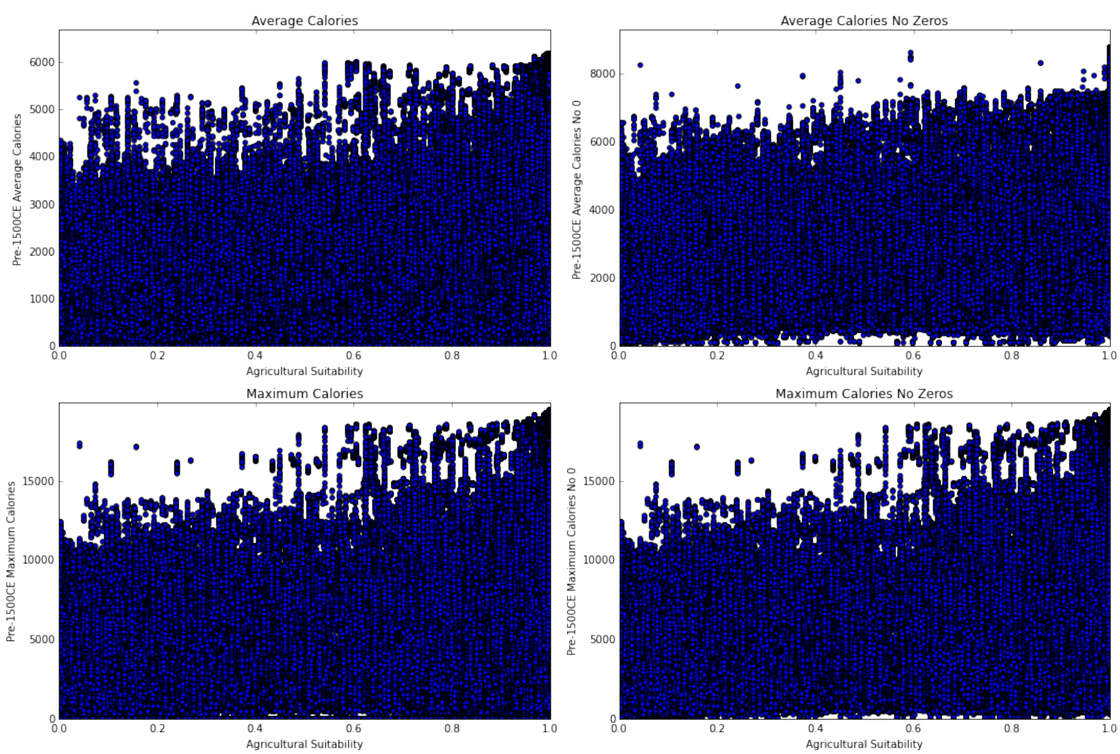
Figure 3: Agricultural Suitability

6.4 Caloric Suitability Indices are different and more essential than and Agricultural Suitability

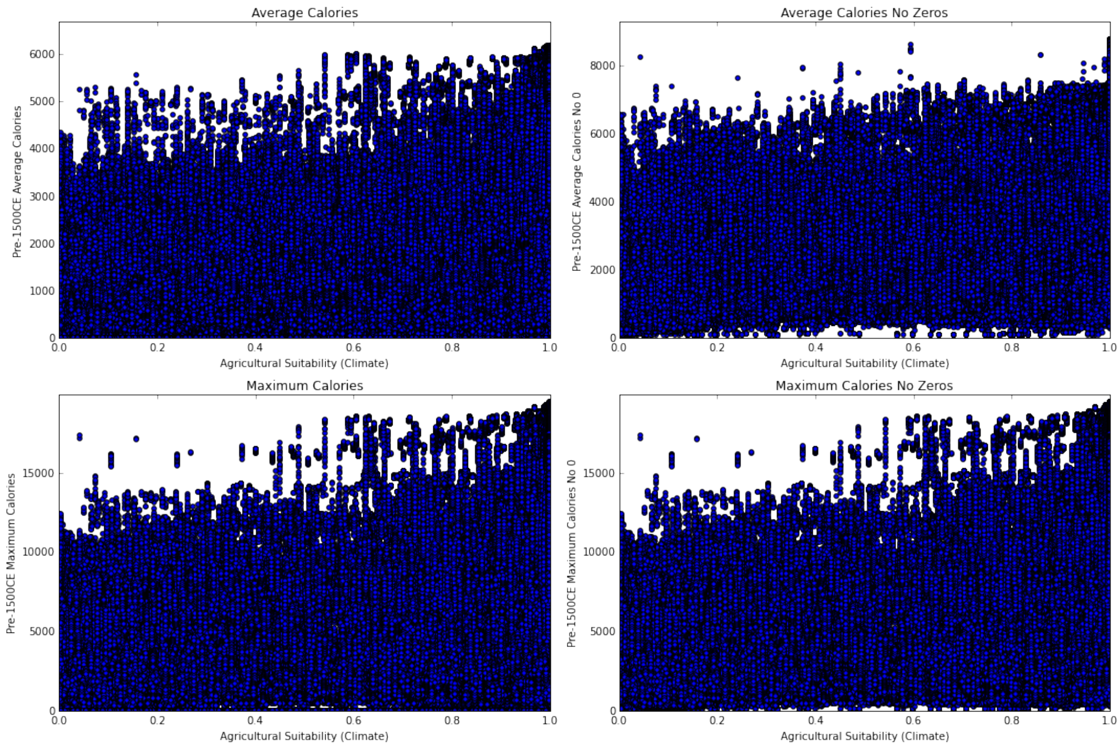
Human existence requires consumption of sufficient calories. Thus, one can expect that mankind would evolve in regions that allow the efficient production of calories. While one can expect that agricultural and caloric suitability be (positively) correlated, they clearly are not the same concept, nor do they measure the same underlying process. In particular, as the following plots show, for any given probability of a cell being suitable for agriculture (as measured by Ramankutty *et al.* (2002)), the Caloric Suitability Indices vary over the full range of their possible values.



Similar results if instead one uses the Pre-1500CE Caloric Suitability Indices



Similar results if instead of *Agricultural Suitability* one uses the *Climatic* component



While these figures show that there is not a strong relation between both sets of indices, it does not show the density or joint probability distibution. The following set of figures show histograms of the joint density of agricultural suitability and CSI.

References

Crosby, A. W. (1972). *The Columbian exchange: biological and cultural consequences of 1492*. Contributions in American studies, no. 2, Westport, Conn.: Greenwood Pub. Co.

Galor, O. and Özak, Ö. (2014). The agricultural origins of time preference. *NBER Working Paper*, (20438).

Ramankutty, N., Foley, J. A., Norman, J. and McSweeney, K. (2002). The global distribution of cultivable lands: current patterns and sensitivity to possible climate change. *Global Ecology and Biogeography*, **11** (5), 377–392.