INSTRUCTIONS FOR NIKITA

- Our first priority is to construct the different versions of indices A1, A2, A3 described in the list below. We want you to generate these indices for (1) subnational region polygons using the shapefiles Arsenij used and then for (2) country polygons. We may later on ask you to repeat the computations for grid-cells and ethnic homeland polygons.

- Before you construct these indices, we need to determine the cells that are suitable for agriculture. These are the ones for which the suitability index (SI) takes a value above the cutoff score of 0.1. These cells should be determined after you normalize the raster values of the suitability index to be between 0 and 1, as Arsenij must have done in his code. Based on that index we should compute the share of raster cells in each region which are suitable for agriculture, as a separate variable.

- Then you need to obtain the raster data showing the amount of calories by crop referred to as caloric suitability index (CSI) for individual crops. These can be downloaded from Omer Ozak's webpage: https://ozak.github.io/Caloric-Suitability-Index/.

- For each of these indices we want to have two versions one based on all 49 available crops and one based on the crops available before 1500. We refer to these two versions as the pre- and the post-Colombian exchange. For the pre-Colombian exchange versions you need to use the list available in the ARSENIJ DB folder which indicates for each crop whether they were native to a given continent/region already prior to 1500 and the discovery of the Americas. The other tab in that file will tell you which continent/broader region each subnational region (with unique region code named "rcode") in the region-level dataset belongs to. Here identification of optimal crops (and the range of potential crops when computing the territorial shares of each crop) should be based on only those crops that were native to the continent/broader region that a subnational region belongs to. I suggest that you take a look at the code Arseniy wrote (under \Crops GIS\crops_suitability\Pre_1500 of the package I shared with you via a DB link) to see if you can adapt what he has done.

FINAL LIST OF CROPS

For our analysis we should use the 46 crops that are covered in both the FAO SI and the CSI database. There should be:

1. Alfalfa	13. Cowpea	25. Olive	37. Sunflower
2. Banana/Plantain	14. Dry pea	26. Onion	38. Sweet potato
3. Barley	**15. Dryland Rice**	27. Pearl millet	39. Switchgrass
4. Buckwheat	16. Flax	28. Phaseolus bean	40. Tea
5. Cabbage	17. Foxtail millet	29. Pigeonpea	41. Tobacco

6. Carrot	18. Gram	30. Rapeseed	42. Tomato
7. Cassava	19. Groundnut	31. Reed canary grass	43. Wetland Rice
8. Chickpea	20. Jatropha	32. Rye	44. Wheat
9. Citrus	21. Maize	33. Sorghum	45. White potato
10. Coconut	22. Miscanthus	34. Soybean	46. Yam
11. Coffee	23. Oat	35. Sugarbeet	
12. Cotton	24. Oil palm	36. Sugarcane	

NEW LIST OF GIS VARIABLES

In order to construct the variables below you will need to use the following GIS raster files:

1. The crop-specific FAO agricultural suitability indices (SI)
2. The crop-specific caloric suitability indices (CSI) by Galor-Ozak
3. The crop-specific FAO production data (SI)
4. The HYDE historical environmental maps
5. The historical trade access data by Michalopoulos-Naghavi-Prarolo
6. The human mobility index of Ozak

You can find the respective raster files in the sub-folder Raster Data (Unprojected)

Part A. Predicted Crop Diversity/Fractionalization Indices

Consider a given polygon *p* which corresponds to a country or a region. This polygon consists of a given number of cells S_p. Among those cells determine the cells *s* that are suitable for agriculture, namely where $max_{c \epsilon C} SI_{ps} \geq 0.1$, based on the FAO SI index. Let's call this subset $\underline{S_p} \equiv \{s \in S_p \mid max_{c \epsilon C} SI_{ps} \geq 0.1\}$. Compute and save also a variable $\boldsymbol{\lambda}_p$ that shows the fraction of cells *s* in S_p that belongs to $\underline{S_p}$.

[Note: This information is for our reference, but we will not be using for the computation of the indices below]

For each of the indices below we should have four versions which are based on one of the following sets of cultivable crops:

I. All 45+1 crops available in the two data sets, denoted as set *C*.

II. The crops historically available in each cell before the Columbian Exchange (pre-1500) denoted as set C_p^{1500}, which are indicated in the spreadsheet we shared earlier.

III. Those crops out of the 45+1 that are at least marginally cultivable ($SI_{psc} \geq 0.1$) in each cell denoted as $\underline{C_{ps}} \equiv \{c \in C \mid SI_{psc} \geq 0.1\}$.

IV. Those crops that are at least marginally cultivable ($SI_{psc} \geq 0.1$) in each cell and historically available before the Columbian Exchange denoted as $\underline{C_{ps}^{1500}} \equiv \{c \in C_p^{1500} \mid SI_{psc} \geq 0.1\}$.

[Note: Sets II and IV are cell-specific, while sets I and III do not vary across the different cells of a polygon.]

A1. Predicted Crop Diversity based on optimal CSI crop [Four Versions: I, II, III, IV.]

For each of the cells in S_p determine the optimal crop *c** out of the four admissible sets of crops, denoted by *AC*, $AC \in \{C, \underline{C_{ps}}, C_{ps}^{1500}, \underline{C_{ps}^{1500}}\}$. This should be the crop that generates the highest amount of calories on that cell provided that this is not zero:

$$CSI_{psc*} \geq max_{c\epsilon AC} CSI_{psc}$$

$$\text{s.t. } CSI_{psc*} > 0$$

If there is such a crop *c** then assume that this crop will be cultivated in the whole area of that cell. Otherwise assume that nothing is cultivated on that cell. Then compute for each of the crops *c* in the admissible set *AC* the total area A_{pc} of polygon p where it will be cultivated. To compute the cultivation shares for each crop we should divide A_{pc} with the total area of cultivation$\sum_{c\epsilon AC} A_{pc}$ in polygon p.Based on these shares we can calculate the predicted crop diversity index as follows:

$$PredCropDiv_p = 1 - \sum_{c\epsilon AC} \left(\frac{A_{pc}}{\sum_{c\epsilon AC} A_{pc}}\right)^2$$

If the total area of cultivation$\sum_{c\epsilon AC} A_{pc}$ in the whole polygon is zero, then we should set the value of the index to -99, so that we can identify these cells.

[Note: When computing the cultivation shares in the polygon, this is not affected by crops and cells where there is no cultivation.]

A2a. Predicted Crop Diversity based on Top3 CSI crops [Four Versions: I, II, III, IV.]

For each of the cells in S_p determine the top 3 set of crops $C3^*$out of the four admissible sets of crops, denoted by *AC*, $AC \equiv \{C, \underline{C_{ps}}, C_{ps}^{1500}, \underline{C_{ps}^{1500}}\}$. This should be the crops that generate the 3 highest amount of calories on that cell provided that this is not zero:

$$If\ c^* \in C3^* \Rightarrow CSI_{psc*} > CSI_{psc}, \forall c \notin C3^*$$

$$\text{s.t. } CSI_{psc*} > 0$$

If there are 3 such crops *c**, then assume that all 3 crops will be cultivated and the whole area of that cell will split equally among them. If the number is less than 3, then the area will be split equally among those. And if the number is zero then nothing is cultivated on that cell. Then compute for each of the crops *c* in the admissible set *AC* the total area A_{pc} of polygon p where it will be cultivated. To compute the cultivation shares for each crop we should divide A_{pc} with the total area of cultivation $\sum_{c \epsilon AC} \quad A_{pc}$ in polygon p.Based on these shares we can calculate the predicted crop diversity index as follows:

$$PredCropDiv_p \quad = 1 - \sum_{c \epsilon AC} \quad \left(\frac{A_{pc}}{\sum_{c \epsilon AC} \quad A_{pc}}\right)^2$$

If the total area of cultivation $\sum_{c \epsilon AC} \quad A_{pc}$ in the whole polygon is zero, then we should set the value of the index to -99, so that we can identify these cells.

A2b. Predicted Crop Diversity based on Top5 CSI crops [Four Versions: I, II, III, IV.]

For each of the cells in S_p determine the top 5 set of crops $C5^*$out of the four admissible sets of crops, denoted by *AC*, $AC \equiv \{C, \underline{C_{ps}}, \; C_{ps}^{1500}, \underline{C_{ps}^{1500}}\}$. This should be the crops that generate the 5 highest amount of calories on that cell provided that this is not zero:

$$If \; c^* \in C5^* \; \Rightarrow CSI_{psc*} \; > CSI_{psc}, \forall c \quad \notin C5^*$$
$$\text{s.t. } CSI_{psc*} > 0$$

If there are 5 such crops *c*,* then assume that all 5 crops will be cultivated and the whole area of that cell will split equally among them. If the number is less than 5, then the area will be split equally among those. And if the number is zero then nothing is cultivated on that cell. Then compute for each of the crops *c* in the admissible set *AC* the total area A_{pc} of polygon p where it will be cultivated. To compute the cultivation shares for each crop we should divide A_{pc} with the total area of cultivation $\sum_{c \epsilon AC} \quad A_{pc}$ in polygon p.Based on these shares we can calculate the predicted crop diversity index as follows:

$$PredCropDiv_p \quad = 1 - \sum_{c \epsilon AC} \quad \left(\frac{A_{pc}}{\sum_{c \epsilon AC} \quad A_{pc}}\right)^2$$

If the total area of cultivation $\sum_{c \epsilon AC} \quad A_{pc}$ in the whole polygon is zero, then we should set the value of the index to -99, so that we can identify these cells.

A3. Predicted Crop Diversity based on Top5 CSI crops using CSI weights
[Four Versions: I, II, III, IV]

For each of the cells in S_p determine the top 5 set of crops $C5^*$out of the four admissible sets of crops, denoted by *AC*, $AC \equiv \{C, \underline{C_{ps}}, \; C_{ps}^{1500}, \underline{C_{ps}^{1500}}\}$. This should be the crops that generate the 5 highest amount of calories on that cell provided that this is not zero:

$$If \; c^* \in C5^* \; \Rightarrow CSI_{psc*} \; > CSI_{psc}, \forall c \quad \notin C5^*$$
$$\text{s.t. } CSI_{psc*} > 0$$

Among the selected crops which could be any number from 0 to 5, assume that the land area of the cell is divided proportionally to their relative CSI values: $CSI_{psc}/\ max_{c\epsilon AC}CSI_{psc}$. So the share of each cell's land area where crop c^* is cultivated is:

$$\frac{CSI_{psc^*}/\ max_{c\epsilon AC}CSI_{psc^*}}{\sum_{C5*}\ (CSI_{psc}/\ max_{c\epsilon AC}CSI_{psc})} = \frac{CSI_{psc^*}}{\sum_{C5*}\ CSI_{psc}}$$

Again if the number of selected crops is zero, then nothing is cultivated on that cell. Then compute for each of the crops *c* in the admissible set *AC* the total area A_{pc} of polygon p where it will be cultivated. To compute the cultivation shares for each crop we should divide A_{pc} with the total area of cultivation$\sum_{c\epsilon AC} A_{pc}$ in polygon p.Based on these shares we can calculate the predicted crop diversity index as follows:

$$PredCropDiv_p = 1 - \sum_{c\epsilon AC} \left(\frac{A_{pc}}{\sum_{c\epsilon AC} A_{pc}}\right)^2$$

If the total area of cultivation$\sum_{c\epsilon AC} A_{pc}$ in the whole polygon is zero, then we should set the value of the index to -99, so that we can identify these cells.

A4. Predicted Crop Diversity based all crops with above mean CSI [Four Versions: I, II, III, IV]

For each of the cells in S_p determine the set of crops C^{Mean*}out of the four admissible sets of crops, denoted by *AC*, $AC \equiv \{C, \underline{C_{ps}},\ C_{ps}^{1500}, \underline{C_{ps}^{1500}}\}$ for which their caloric suitability is above the mean value in that cell, provided that this is not zero. This should be the crops for which

$$If\ c^* \in C^{Mean*} \Rightarrow CSI_{psc*} \geq Mean_{c\epsilon AC}\{CSI_{psc}\}$$
$$\text{s.t. } CSI_{psc*} > 0$$

Among the selected crops which could be any number, assume that the land area of the cell is divided proportionally to their relative CSI values: $CSI_{psc}/\ max_{c\epsilon AC}CSI_{psc}$. So the share of each cell's land area where crop c^* is cultivated is:

$$\frac{CSI_{psc^*}/\ max_{c\epsilon AC}CSI_{psc^*}}{\sum_{C5*}\ (CSI_{psc}/\ max_{c\epsilon AC}CSI_{psc})} = \frac{CSI_{psc^*}}{\sum_{C5*}\ CSI_{psc}}$$

Again if the number of selected crops is zero, then nothing is cultivated on that cell. Then compute for each of the crops *c* in the admissible set *AC* the total area A_{pc} of polygon p where it will be cultivated. To compute the cultivation shares for each crop we should divide A_{pc} with the total area of cultivation$\sum_{c\epsilon AC} A_{pc}$ in polygon p.Based on these shares we can calculate the predicted crop diversity index as follows:

$$PredCropDiv_p = 1 - \sum_{c\epsilon AC} \left(\frac{A_{pc}}{\sum_{c\epsilon AC} A_{pc}}\right)^2$$

If the total area of cultivation$\sum_{c\epsilon AC} A_{pc}$ in the whole polygon is zero, then we should set the value of the index to -99, so that we can identify these cells.

A5. Predicted Crop Diversity based all crops using CSI weights [Four Versions: I, II, III, IV]

For each of the cells in S_p consider all crops in the four admissible sets of crops, denoted by *AC*, $AC \equiv \{C, \underline{C_{ps}},\ C_{ps}^{1500}, \underline{C_{ps}^{1500}}\}$. Assume that each crop is cultivated in the land area of the cell is proportionally to their relative CSI values: $CSI_{psc}/\ max_{c\epsilon AC}CSI_{psc}$. So the share of each cell's land area where crop c^* is cultivated is:

$$\frac{CSI_{psc^*}/\ max_{c\epsilon AC}CSI_{psc^*}}{\sum_{C5*}\ (CSI_{psc}/\ max_{c\epsilon AC}CSI_{psc})} = \frac{CSI_{psc^*}}{\sum_{C5*}\ CSI_{psc}}$$

Again if the number of selected crops is zero, then nothing is cultivated on that cell. Then compute for each of the crops *c* in the admissible set *AC* the total area A_{pc} of polygon p where it will be cultivated. To compute the cultivation shares for each crop we should divide A_{pc} with the total area of cultivation$\sum_{c\epsilon AC} A_{pc}$ in polygon p.Based on these shares we can calculate the predicted crop diversity index as follows:

$$PredCropDiv_p \ = 1 - \sum_{c\epsilon AC} (\frac{A_{pc}}{\sum_{c\epsilon AC} A_{pc}})^2$$

If the total area of cultivation$\sum_{c\epsilon AC} A_{pc}$ in the whole polygon is zero, then we should set the value of the index to -99, so that we can identify these cells.

<u>B. Variation in Agricultural Suitability</u>

Consider a given polygon *p* which corresponds to a country or region. This polygon consists of a given number of cells S_p. Among those cells determine the cells *s* that are suitable for agriculture, namely where $min_{c\epsilon C}SI_{ps} \geq\ 0.1$. Let's call this subset $\underline{S_p}$.

B0. Standard Deviation of Suitability Index across cells by maximum CSI.
[Two Versions: Based on S_p and $\underline{S_p}$]

For each of the 46 crops separately computed the standard deviation of the suitability index SI_{psc} across all cells in either of the two sets $S \equiv \{S_p, \underline{S_p}\}$:

$$StDevB0_{pc} = \sqrt{\frac{1}{S}\sum_{s=1}^{S} [SI_{psc} - (\frac{1}{S}\sum_{s=1}^{S} SI_{psc})]^{2}}$$

[Note: That this measure should be crop specific.]

B1. Weighted Standard Deviations of Suitability Index across cells by maximum CSI.
[Two Versions: Based on S_p and $\underline{S_p}$]

For each of the 46 crops separately computed the standard deviation of the suitability index SI_{psc} across all cells in either of the two sets $S \equiv \{S_p, \underline{S_p}\}$ weighted by its maximum caloric yield in terms of $max_{c\epsilon C} CSI_{psc}$ to proxy for historical population density:

$$WeightStDevB1_{pc} = \sqrt{\sum_{s=1}^{S} w_s [SI_{psc} - (\sum_{s=1}^{S} w_s SI_{psc})]^2}$$

$$WeightStDevB1_{pc} = \sqrt{\frac{1}{S} \sum_{s=1}^{S} [w_s SI_{psc} - (\frac{1}{S} \sum_{s=1}^{S} w_s SI_{psc})]^2}$$

where $w_s = \frac{max_{c\epsilon C} CSI_{psc}}{\sum_{s=1}^{S} (max_{c\epsilon C} CSI_{psc})}$

[Note: That this measure should be crop specific.]

B2. Weighted Standard Deviations of Suitability Index across cells by Population Density. [Two Versions: Based on S_p and $\underline{S_p}$]

For each of the 46 crops separately computed the standard deviation of the suitability index SI_{psc} across all cells in S_p and $\underline{S_p}$ weighted by historical population density in 1500 from raster file popd_1500AD of the HYDE database:

$$WeightStDevB2_{pc} = \sqrt{\sum_{s=1}^{S} w_s [SI_{psc} - (\sum_{s=1}^{S} w_s SI_{psc})]^2}$$

$$WeightStDevB2_{pc} = \sqrt{\frac{1}{S} \sum_{s=1}^{S} [w_s SI_{psc} - (\frac{1}{S} \sum_{s=1}^{S} w_s SI_{psc})]^2}$$

where $w_s = \frac{PopDens_{ps}}{\sum_{s=1}^{S} PopDens_{ps}}$

For historical population density in each cell there is a separate raster file.

[Note: That this measure should be crop specific.]

D. Contemporary Crop Diversity Indices

Consider a given polygon *p* which corresponds to a country or region. This polygon consists of a given number of cells S_p. Among those cells determine the cells *s* that are suitable for agriculture, namely where $min_{c\epsilon C} SI_{ps} \geq\ 0.1$.

D1. Contemporary Crop Diversity based on FAO Cultivation Data

For each of the cells either of the two sets $S \equiv \{S_p, \underline{S_p}\}$ compute the share in terms of hectares used for the cultivation of each of the 27 crop categories (C27) covered by the FAO production database. For consistency compute the shares as:

$$\phi_{pc} = \frac{\sum_s^S \quad hectares_{psc}}{\sum_c^{C27} \quad \sum_s^S \quad hectares_{psc}}$$

Based on these hectare shares we can compute crop diversity as a fractionalization score.

$$ActualCropDivD1_p \ = 1 - \sum_{c\epsilon C27} \quad (\phi_{pc})^2$$

E. Historical Importance of Agriculture

Consider a given polygon *p* which corresponds to a country or a region. This polygon consists of a given number of cells S_p. For each of the raster files below we should compute the average value of variable X across all cells in the polygon:

$$HistAgr_p = \frac{1}{S_p} \sum_{s=1}^{S_p} \quad X_{ps}$$

E0a. Population Density in 1500

Let X be the population density of the gridcell in year 1500 from the HYDE database reported in the raster file popd_1500.asc

E0b. Population Density in 1700

Let X be the population density of the gridcell in year 1700 from the HYDE database reported in the raster file popd_1700.asc

E1a. Cropland Area in 1500

Let X be the share of cropland area of the gridcell in year 1500 under baseline scenario of the HYDE database reported in the raster file cropland1500.asc

E1b. Cropland Area in 1700

Let X be the share of cropland area of the gridcell in year 1700 under baseline scenario of the HYDE database reported in the raster file cropland1700.asc

E2a. Red-Fed Cropland Area in 1500
Let X be the share of rain-fed cropland area of the gridcell in year 1500 under baseline scenario of the HYDE database reported in the raster file tot_rainfed1500.asc

E2b. Red-Fed Cropland Area in 1700
Let X be the share of rain-fed cropland area of the gridcell in year 1700 under baseline scenario of the HYDE database reported in the raster file tot_rainfed1700.asc

E3a. Red-Fed Rice Cropland Area in 1500
Let X be the share of rain-fed rice cropland area of the gridcell in year 1500 under baseline scenario of the HYDE database reported in the raster file rf_rice1500.asc

E3b. Red-Fed Rice Cropland Area in 1700
Let X be the share of rain-fed rice cropland area of the gridcell in year 1700 under baseline scenario of the HYDE database reported in the raster file rf_rice1700.asc

E4a. Red-Fed Non-Rice Cropland Area in 1500
Let X be the share of rain-fed non rice cropland area of the gridcell in year 1500 under baseline scenario of the HYDE database reported in the raster file rf_norice1500.asc

E4b. Red-Fed Non-Rice Cropland Area in 1700
Let X be the share of rain-fed non rice cropland area of the gridcell in year 1700 under baseline scenario of the HYDE database reported in the raster file rf_norice1700.asc

E5a. Grazing Area in 1500
Let X be the share of grazing area of the gridcell in year 1500 under baseline scenario of the HYDE database reported in the raster file grazing1500.asc

E5b. Grazing Area in 1700
Let X be the share of grazing area of the gridcell in year 1700 under baseline scenario of the HYDE database reported in the raster file grazing1700.asc

E6a. Rangeland Area in 1500
Let X be the share of rangeland area of the gridcell in year 1500 under baseline scenario of the HYDE database reported in the raster file rangeland1500.asc

E6b. Rangeland Area in 1700
Let X be the share of rangeland area of the gridcell in year 1700 under baseline scenario of the HYDE database reported in the raster file rangeland1700.asc

E7a. Pasture Area in 1500
Let X be the share of pasture area of the gridcell in year 1500 under baseline scenario of the HYDE database reported in the raster file pasture1500.asc

E7b. Pasture Area in 1700
Let X be the share of pasture area of the gridcell in year 1700 under baseline scenario of the HYDE database reported in the raster file pasture1700.asc

<u>F. Historical Importance of Trade Access</u>

F1a. Number of Historic Trade Routes passing through the Region in 1700.
For each polygon we should have a simple count variable counting the number of trade route points within its borders in the year 1700 based on the shape file trade_1700_full.

F1b. Number of Historic Trade Routes passing through the Region in 1300.
For each polygon we should have a simple count variable counting the number of trade route points within its borders in the year 1300 based on the shape file trade_1300.

F2. Number of Ancient Ports within the Region
For each polygon we should have a simple count variable counting the number of ancient ports within its borders based on the shape file ancient_ports.

F3. Number of Roman Roads within the Region
For each polygon we should have a simple count variable counting the number of roman road points within its borders based on the shape file romanroads_points.

F4. Minimum Distance between Region's Centerpoint and Closest Trade Route in 1700
For each polygon we should compute the minimum geodesic distance on the map between its centerpoint and the closest trade point on any trade route on the map in the year 1700 based on the shape file trade_1700_full.

F4b. Minimum Distance between Region's Centerpoint and Closest Trade Route in 1300
For each polygon we should compute the minimum geodesic distance on the map between its centerpoint and the closest trade point on any trade route on the map in the year 1300 based on the shape file trade_1300.

F5. Minimum Distance between Region's Centerpoint and Closest Ancient Port
For each polygon we should compute the minimum geodesic distance on the map between its centerpoint and the closest ancient port on the map based on the shape file ancient_ports.

F6. Minimum Distance between Region's Centerpoint and Closest Roman Road
For each polygon we should compute the minimum geodesic distance on the map between its centerpoint and the closest roman road point on the map based on the shape file romanroads_points.

F7. Minimum Travel Time between Region's Centerpoint and Closest Trade Route in 1700

For each polygon we should compute the least cost path on the HMI raster between its centerpoint and the closest trade point on any trade route on the map in the year 1700 based on the shape file trade_1700_full.

F7b. Minimum Travel Time between Region's Centerpoint and Closest Trade Route in 1300
For each polygon we should compute the least cost path on the HMI raster between its centerpoint and the closest trade point on any trade route on the map in the year 1300 based on the shape file trade_1300.

F8. Minimum Travel Time between Region's Centerpoint and Closest Ancient Port
For each polygon we should compute the least cost path on the HMI raster between its centerpoint and the closest ancient port on the map based on the shape file ancient_ports.

F9. Minimum Travel Time between Region's Centerpoint and Closest Roman Road
For each polygon we should compute the least cost path on the HMI raster between its centerpoint and the closest roman road point on the map based on the shape file romanroads_points.

<u>G. Historical Accessiblity</u>
Consider a given polygon *p* which corresponds to a country or a region. This polygon consists of a given number of cells S_p. For each cell *s* in each polygon *p* we should define the number of adjacent cells Adj_{ps}, which should be a number from 0 to 8.

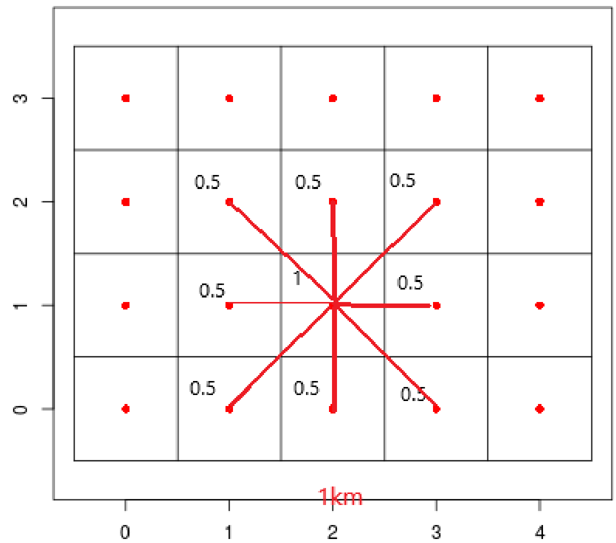
G1. Relative Local Within-Accessibility Index without Seafaring
For each cell *s* in each polygon *p* we should compute the average travel time by land based on the Human Mobility Index of Ozak (HMI) from the centroid of each adjacent cell in Adj_{ps} to the centroid of our given cell. Then we should average across all cell of each polygon p to get an average historical accessibility index for the polygon:

$$HistAccessG1_p = \frac{1}{S_p}\sum_{s=1}^{S_p} \frac{1}{Adj_{ps}}\sum_{s'=1}^{Adj_{ps}} HMI_{pss'}$$

G2. Relative Local Within-Accessibility Index with Coastal Seafaring
For each cell *s* in each polygon *p* we should compute the average travel time by land and coastal seafaring based on the alternative Human Mobility Index of Ozak (HMISea) from the centroid of each adjacent cell in Adj_{ps} to the centroid of our given cell. Then we should average across all cell of each polygon p to get an average historical accessibility index for the polygon:

$$HistAccessG2_p = \frac{1}{S_p}\sum_{s=1}^{S_p} \frac{1}{Adj_{ps}}\sum_{s'=1}^{Adj_{ps}} HMISea_{pss'}$$

In both the above cases we should treat the raster values as applying continuously in space and think of the pairwise index $HMI_{pss'}$ between cells s and s' in polygon p as the average value along the straight line between the two centroids of these two adjacent cells shown in the figure:



To provide an example, take the above grid and assume each cell is a 1km X 1km square. Suppose for simplicity that the value of HMI in a given cell is 1, (so a human needs 1 day to cross it) and 0.5 in all 8 adjacent cells. Then HMI(pss') should be 0.75 if s' are the cells N, S, W and E and should be SQRT(2) * 0.75 if s' are the cells NW, NE, SW and SE. Hence, if we apply the formula for the average pairwise HMI for cell s that should be 0.905.