Strategic Maps for Decision-Making on
Opium-Poppy Cultivation in Afghanistan
(Internal draft 04 May 2020 – formatted as research, not as a policy document)

1) Background information

Afghanistan is the country with the largest areas under opium poppy cultivation worldwide (82 percent of total global area)[1]. Opium poppy cultivation is widespread and takes place in 28 of the 34 provinces (and 178 of the 399 districts) of this country[2]. The production of opiates has become a crucial pillar of Afghanistan's economy and rural communities have become dependent on the income from opium poppy to sustain their precarious livelihoods[3]. At the same time, opiates provide income to anti-government and terrorist groups, mainly through forced "taxation" to farmers, which fuels violence, exacerbates the on-going war, and impedes development in this country. As such, opium poppy cultivation is at present one of the most important axes that sustain the vicious circle of underdevelopment and violence in Afghanistan. Providing reliable solutions to this complex problem requires the availability of robust evidence about which specific (income and non-income) poverty and security-related factors play a (statistically) significant role in opium poppy cultivation. Those factors have proved to be very different based on the geographic location of the affected communities and ideally should be identified per community to be able to deliver effective and sustainable responses.

In general, it is still unknown the magnitude and statistically significance of the effects of several geographic and climatic circumstances, including droughts; security and governability; agricultural and market conditions (such as prices of opium poppy and other licit crops); demographic, socio-economic and the presence of external (dis-) incentives for opium poppy cultivation (for example, access to advanced money for opium poppy cultivation -similar to credit to farmers provided by narcotics traffickers- and awareness campaigns against opium poppy cultivation) have in the extent of opium poppy cultivation. This kind of information is vital for evidence-based drug-control policy design. So, this research aims to contribute to the scarce (geo-) referenced and quantitative evidence on what could work on building resilience against opium poppy cultivation in this country.[4]

2) Goals and Question(s)

The main objective of this research is to determine which factors are statistically significantly associated with opium poppy cultivation in Afghanistan with a view of potentially forecasting the effects that changes in any of those factors (in year 0) may have in future opium poppy areas (in year 1 or beyond). This can inform drug-control policy design, as statistically significantly factors (with large absolute magnitudes) can be prioritized when feasible to curve

[1] UNODC. 2019. "World Drug Report 2018". Vienna, Austria.
[2] UNODC. 2018. "Afghanistan Opium Survey 2018. Cultivation and Production".
[3] UNODC. 2019. "Afghanistan Opium Survey 2018. Challenges to Sustainable Development, Peace and Security".
[4] Several authors have conducted research on the socio-economic factors influencing opium poppy cultivation in Afghanistan, for example: Mansfield (2018). "Turning deserts into flowers: settlement and poppy cultivation in southwest Afghanistan." Third World Quarterly, 39 (2). pp. 331-349; Garcia-Yi. 2017. "Building Resilience to Opium Poppy Cultivation by Strengthening the Design of Alternative Development Interventions: Evidence from Afghanistan. Bulletin on Narcotics, Vol LXI. United Nations. However, few researchers have explicitly included geo-referenced information in their analysis, with the exception of authors such as Mansfield and Fishstein. 2016. "Time to Move on: Developing an Informed Development Response to Opium Poppy Cultivation in Afghanistan". Afghanistan Research and Evaluation Unit (AREU); although they mostly incorporated remote sensing for crop rotation evaluation and location of wells.

illicit crop cultivation (in other words, if for example, awareness campaigns against opium poppy cultivation in year 0 have a statistically significant effect in opium poppy cultivation in year 1, it can be turned into a priority as drug-control policy).

The main research question is:
What factors and shocks are (statistically significantly) associated with opium poppy cultivation at district level[5] ("ceteris paribus" and after controlling for spatial autocorrelation, if needed)?

3) Conceptualization of question(s)
Overall, four types of factors (for which information is publicly available) are likely to have a (significant) effect on farmers' decisions regarding opium poppy cultivation. They are:

a) geographic and climatic conditions
b) security and governability conditions
c) agricultural and market characteristics
d) demographic and socio-economic conditions and presence of external (dis)incentives for opium poppy cultivation

Each of these types of factors and their corresponding hypotheses are explained in more detail below.

Geographic and climatic conditions
Opium poppy, as any other annual crop, is influenced by geographic and climatic conditions. However, little information is publicly available on its (optimal and sub-optimal) cropping characteristics. Conditions such as elevation, temperature, precipitation, and main type of ecoregion are expected to affect its cultivation. Therefore, the specific hypotheses (H) to be evaluated under this category are:

H1 (altitude): High altitude restricts the cultivation of most annual crops. As such, high altitude is expected to decrease opium poppy cultivation.

H2 (temperature): High temperature limits the survival of annual crops. Therefore, high temperature will likely reduce opium poppy cultivation. As well high temperature is used as proxy of drought (in this research).

H3 (precipitation): Low precipitation is expected to reduce opium poppy cultivation. Similarly, to high temperature, low precipitation is used as proxy of drought (in this research).

H4 (ecoregion): There are ten main ecoregions in Afghanistan[6], as defined by WWF (see Table 1 for full reference). However previous research reports[7] suggest that opium poppy is currently

[5] District was finally selected as unit of analysis as it is the smallest geo-referenced unit for which there is information available in the multiple criteria considered in this research (see section 3 for the list of criteria and hypotheses).
[6] After selecting only one ecoregion per district (the most prominent in coverage inside each district). Those ecoregion are: Central Afghan Mountains xeric woodlands, Baluchistan xeric woodlands, Ghorat-Hazarajat alpine meadow, Sulaiman Range alpine meadows, Hindu Kush alpine meadow, Paropamisus xeric woodlands, East Afghan montane conifer forests, Badghyz and Karabil semi-desert, Registan-North Pakistan sandy desert and Central Persian desert basins.
[7] See Mansfield (2018). "Turning deserts into flowers: settlement and poppy cultivation in southwest Afghanistan." Third World Quarterly, 39 (2). pp. 331-349

largely cultivated in desert areas. As such is hypothesized that in desertic ecoregions[8] there will be more opium poppy cultivation than in non-desertic ones.

Security and governability conditions
Anti-government groups tend to obtain profits from opium poppy cultivation. In general, conflicted regions are expected to have more opium poppy cultivation than non-conflicted ones. The specific hypotheses under this category include:

H5 (government influence): lack of government presence is expected to be positively correlated with opium poppy cultivation

H6 (fatalities due to violence): a greater number of fatalities due to incidents of lethal violence are likely to be correlated with opium poppy cultivation inside the same district

H7 (internal displacement): more conflict influences internal displacement. Therefore, a greater number of internally displaced persons (i.e., families leaving the district) is expected to be positively associated with opium poppy cultivation

H8 (on-going aid projects): security conditions affect the continuity of aid projects. In addition, farmers who access aid projects are less likely to be economically dependent on opium poppy cultivation. Therefore, more number of on-going aid projects is expected to be correlated with less opium poppy cultivation inside the same district.

Agricultural and market characteristics
As for any other crop, the extent of opium poppy cultivation is influenced by the availability of agricultural land, water for irrigation, type of soil, and distance to major cities and markets. The hypotheses under this group include:

H9 (agricultural areas): more opium poppy cultivation is likely to occur in district with larger extensions of agricultural areas (as for any other crop, cultivation is not possible if there is no agricultural land). More specifically, access to irrigated agricultural land is likely associated with more opium poppy cultivation (especially considering that cultivation occurs in desert areas, see H4 above).

H10 (rivers): district with more superficial water (rivers) are likely to have more crop cultivation, including opium poppy

H11 (type of soil): There are eighteen main types of soil in Afghanistan[9], as defined by USDA (see Table 1 for full reference). For this research, it is hypothesized that in rocky soil[10] there

[8] Desertic ecoregions include: Badghyz and Karabil semi-desert, Registan-North Pakistan sandy desert and Central Persian desert basins.

[9] After selecting only one type of soil per district (the most prominent in coverage inside each district). Those 18 types of soils are: Haplocambids with Torriorthents, Xerochrepts with Xerorthents, Rocky land with Lithic Haplocryids, Rocky land with Lithic Cryorthents, Rocky land with Lithic Haplocambids, Calcixeralfs with Xerochrepts, Xerorthents with Xeropsamments, Torriorthents with Torrifluvents, Natrixeralfs with Halaquepts, Torripsamments with Dunes, Rocky land with Torriorthents, Haplocambids with Torripsamments, Torripsamments with Torriorthents, Haplocalcids with Torriorthents, Dunes, Torrifluvents, Torrifluvents with Haplogypsids, Torrifluvents with Haplosalids

[10] Rocky soil includes: Rocky land with Lithic Haplocryids, Rocky land with Lithic Cryorthents, Rocky land with Lithic Haplocambids, and Rocky land with Torriorthents.

will be less opium poppy cultivation than in other types of soil (as it is harder to cultivate any crop in general).

H12 (distance to cities/markets): farmers for whom access to cities (and their markets) is easier are likely less dependent on opium poppy income. Therefore, it is expected that shorter distances to cities is correlated with less opium poppy cultivation.

H13 (opium poppy prices): higher opium poppy prices in year 0 is likely correlated with high opium poppy cultivation in year 1

H14 (prices of legal crops): lower legal crop prices in year 0 is likely correlated with high opium poppy cultivation in year 1

Demographic, socio-economic and presence of external (dis)incentives for opium poppy cultivation

It is likely that cheaper (farm) labour (for example, for assisting in harvesting opium poppy, which is very labour intensive) will be available in districts with more population. Farmers also take their decisions on income generating activities based on their compatibility with their usual type of livelihood (for example, farm households who mainly cultivate cash crops would be more likely to cultivate opium poppy than farm households who focus on raising livestock). The main hypotheses under this category are:

H15 (population per district): more number of person inside the district, the larger the opium poppy cultivation

H16 (livelihood zone): There are twenty-nine main type of livelihood zones[11] in Afghanistan, as defined by USAID (see Table 1 for full reference). For this research, it is hypothesized that farm households which livelihood is associated with livestock and similar (in non-arid conditions) [12] are less prone to cultivate opium poppy.

[11] After selecting only one livelihood zone per district (the most prominent in coverage inside each district). Those livelihood zones are: Kabul and Logar Irrigated Zone, East-Central Orchard and Agriculture Zone, East-Central Vineyard, Cereal and Horitculture Zone, East-Central Mountainous Agro-Pastoral Zone, Eastern Mixed Agriculture and Forest Zone, Eastern Intensive Irrigated Agriculture Zone, Eastern Semi-Arid Agriculture Zone, Eastern Agro-Pastoral and Forest Zone, Eastern Cross-Border Trade and Labor Zone, Kunduz-Baghlan High Cereal Production Zone, Northeastern Highland Agro-Pastoral Zone, Southern Intensive Irrigated Vegetable and Orchard Zone, South-Central Mountain Wheat, Dried Fruit and Livestock Zone, Southeastern Zabul Rainfed Cereals and Orchard Zone, Eastern Deep-Well Irrigated Agriculture Zone, Southeastern High-Migration, Forest-Product and Livestock Zone, Takhar-Badakshan Mixed-Agriculture Zone, Northern Rainfed Mixed Farming Zone Northwest Agro-Pastoral Zone, Northern Intensive Irrigated Agriculture Zone, Amo River Irrigated Cereals and Oilseed Zone, West-Central Highland Agro-Pastoral Zone, Northern Kandahar Agriculture and Livestock Zone, Southern Semi-Arid Agro-Pastoral Zone, Helmand Intensive Irrigated Wheat and Cash Crop Zone, South-Central Mixed Farming Zone, Western Intensive Irrigated Agriculture Zone, Western Semi-Arid Agro-Pastoral Zone, Western and Southern Cross-Border Trade and Labor Zone.

[12] Livelihoods associated with livestock and similar include: East-Central Mountainous Agro-Pastoral Zone, Eastern Agro-Pastoral and Forest Zone, Northeastern Highland Agro-Pastoral Zone, South-Central Mountain Wheat, Dried Fruit and Livestock Zone, Southeastern High-Migration, Forest-Product and Livestock Zone, Northwest Agro-Pastoral Zone, West-Central Highland Agro-Pastoral Zone, Northern Kandahar Agriculture and Livestock Zone

H17 (multi-dimensional poverty): farmers who are the poorest lack access to reliable sources of (legal) income and opium poppy can became the best suitable option for them. As such high multi-dimensional poverty is likely associated with more opium poppy cultivation.

H18 (other socio-economic conditions and external incentives): lack of education inside the districts is likely associated with more opium poppy cultivation (measured with percentage of villages with availability of boy and girl school as proxy). Also, the availability of advanced money for opium poppy cultivation inside the district is expected to facilitate its cultivation. Finally, awareness campaigns against opium poppy cultivation is expected to be negatively correlated with opium poppy cultivation.

4) Analytical Representation of the Hypotheses

The analytical representation of the hypotheses is presented in Figure 1 below:

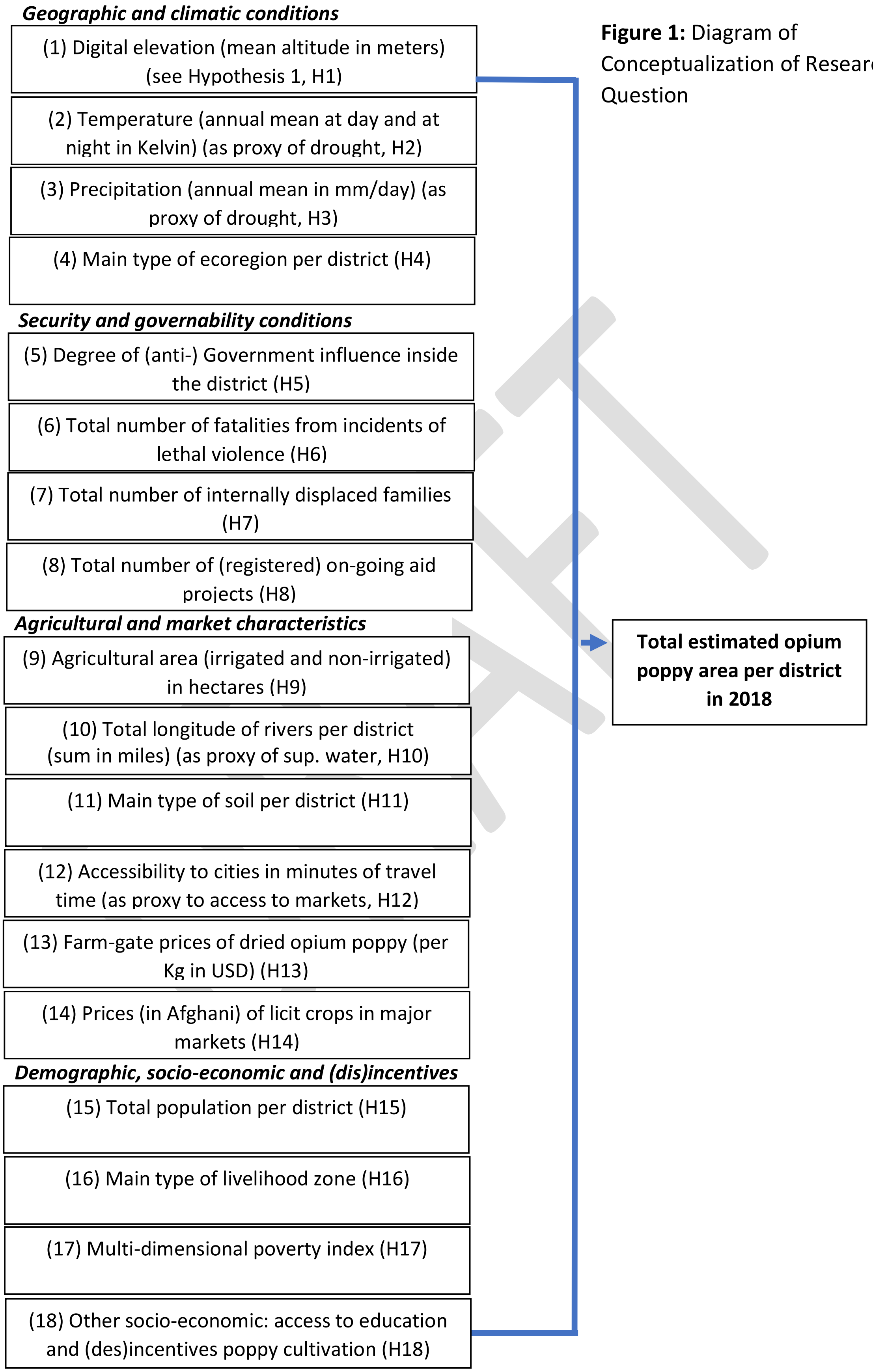


Figure 1: Diagram of Conceptualization of Research Question

5) Accessibility and Quality of Data

It was challenging to gather and systematize information for each the 18 main hypotheses (and 31 specific hypotheses) considered in this research (see Table 1 for the full list). Not only the access to good quality data about Afghanistan is limited, but also the format in which the data is available varies according to the source of information (from tabulated reports; satellite images; shapefiles, including points, lines and polygons; to datasets in excel or similar).

The unit of analysis in this research is the district to make possible to compare the multiple sources and types of information However, this unit of analysis in a country such as Afghanistan is not without complications. For example, the Afghan Geodesy and Cartography established 399 districts in 2012. Nevertheless, other sources use 407 districts, such as the U.S. Forces – Afghanistan. Also, some districts share names, and many districts have several names. This research uses the list of 399 districts from the Afghan Government.

Table 1 lists the source of data and format per hypothesis (explanatory variable) considered in this research. The data included in (1), (6), (8), (9), (10), (14), (15), (16) and (18) were mostly manually digitalized from reports, as automatic matching of names was mostly not possible due to duplication of names and misspellings. The data in (2), (3), (4), and (13) was downloaded from open source remote sensing images and further processed to obtain an average value per district (procedure not included in this report). The data in (5), (7), (11), (12), (17) and (19) were obtained in shapefiles (points, lines or polygons), which still required further processing to obtain an average of their corresponding values per district.

Table 1: Source of data and format per hypothesis (explanatory variable)

Type of data per district	Source	Year of the data	Format	Resolution
Dependent variable				
(1) Total estimated opium poppy area per district	UNODC Reports[13]	2017-2018	Tabulated from reports	-
Explanatory variables				
Geographic and climatic conditions				
(2) Digital elevation (mean altitude in meters)	The Shuttle Radar Topography Mission (Joint Project by NASA, National Geospatial-Intelligence Agency, and the	2000	Raster	Pixel size: 0.0003, -0.0003

[13] The results at district level are approximations, as the remote sensing assessment conducted by UNODC is designed to obtain estimates of opium poppy at province level. Specifically, for district level estimations all cells are used which have the majority of agricultural area in that district. That means that in certain cases, agricultural area and poppy cultivation is accounted for in a neighboring district and not within the district where cultivation occurred. This is, however, in most cases set off by those cells, where the contrary is the case. (see UNODC. 2018. "Afghanistan Opium Survey 2018. Cultivation and Production").

Type of data per district	Source	Year of the data	Format	Resolution
	German and Italian Space Agencies)			
(3) Temperature a) annual mean at day b) annual mean at night in Kelvin	MODIS (Collection 6, Land Surface Temperature Products)	2017-2018	Raster	Pixel size: 0.01, -0.01
(4) Precipitation (annual mean in mm/day)	USGS. Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS)	2017-2018	Raster	Pixel size: 0.05, -0.05
(5) Main type of ecoregion per district (10 ecoregions)	World Wildlife Fund (WWF). Global 200 (Terrestrial) Ecoregions.	2004	Polygon (shapefile)	-
Security and governability conditions				
(6) Degree of (anti-) Government influence inside the district[14]	Special Inspector General for Afghanistan Reconstruction (SIGAR) Reports	2017-2018	Tabulated from reports	-
(7) Most likely estimate of total number of fatalities from incidents of lethal violence	UCDP Georeferenced Event Dataset. Department of Peace and Conflict, Uppsala University	2017-2018	Point (Shapefile)	-
(8) Total number of internally displaced families: a) emigrated b) immigrated c) net migration to the district (b-c)	UN Office for the Coordination of Humanitarian Affairs. The Humanitarian Data Exchange.	2017-2018	Dataset (Excel or Similar)	-
(9) Total number of (registered) on-going aid projects[15]	International Aid Transparency Initiative	2018	Dataset (Excel or Similar)	-
Agricultural and market characteristics				
(10) Agricultural area in hectares:	FAO. 2016. The Islamic Republic of	2010	Tabulated from reports	-

[14] Where 1=Controlled by the Government of Afghanistan, 2=Influence of the Government of Afghanistan, 3=Contested, 4=Insurgent activity, 5=High insurgent activity

[15] Projects that started before 2018 and were still on-going in 2018

Type of data per district	Source	Year of the data	Format	Resolution
a) total b) irrigated c) rain-fed d) barren land	Afghanistan. Land Cover Atlas.			
(11) Total longitude of rivers per district (sum in miles)	USAID (derived from US Defense Mapping Agency, 1967-1988)	1990	Lines (shapefile)	-
(12) Main type of soil (18 types)	USDA. 2001. Afghan Soil Map	2001	Polygon (shapefile)	-
(13) Accessibility to cities (high density urban centers) in minutes of travel time (mean of the district)	Weiss et al. 2018. "A global map of travel time to cities to assess inequalities in accessibility in 2015". *Nature* **553**: 333–336.	2015	Raster	Pixel size: 0.0083, -0.0083 (as stated in the Info from Provider)
(14) Farm-gate prices of dried opium poppy (per Kg in USD)	UNODC's Opium Poppy Price Reports	2017-2018	Tabulated from reports	-
(15) Prices (in Afghani) in eight major markets of: a) wheat b) wheat flour c) rice	World Food Program. Afghanistan - Market Price Bulletins.	2017-2018	Tabulated from reports	-
Demographic, socio-economic and presence of (dis)incentives				
(16) Population per district a) total b) rural	Central Statistics Organization of Afghanistan[16]	2018	Tabulated from reports	-
(17) Main type of livelihood zone per district (29 zones where people share similar options for food and income and access to markets)	USAID's Famine Early Warning Systems Network	2011	Polygon (shapefile)	-
(18) Index of multi-dimensional	National Statistics and Information Authority and Oxford University.	2017	Tabulated from reports	-

[16] The population for the district of Sharak-e-Hayratan in Balkh province was missing from the official statistics. The information for this district was obtained from Special Inspector General for Afghanistan Reconstruction (SIGAR). 2018. Addendum to SIGAR's January 2018 Quarterly Report to the United States Congress.

Type of data per district	Source	Year of the data	Format	Resolution
poverty (per province)	2019. Multi-Dimensional Poverty Index 2016-2017. Report and Analysis.			
(19) Percentage of villages with availability education and (des)incentives for opium poppy cultivation inside the district: a) boy school b) girl school c) advanced money for opium poppy cultivation e) awareness campaigns against opium poppy cultivation	UNODC's Annual Socio-Economic Surveys (Internal Information)	2017	Dataset (Excel or similar with geo-referenced location)	-

6) Boundaries, Scale, and Other Concerns

The data corresponding to (9), (15) and (18) in Table 1 (see previous section) were originally reported at province / regional level instead of at district level. Therefore, all the districts inside the same province / region have the same values for these data. The difference in scale may have influenced the results obtained from this research.

In addition, the data corresponding to prices (14) and socio-economic characteristics (19) in Table 1 corresponded to specific locations (or a representative sample of villages), but for some districts there were not observations (due to the random sampling). To avoid having districts with not observations a kriging procedure was used to obtain values of these variables for all the country using the available geo-referenced village information (see Appendix A for an overall description of the method).

Also, for the main type of ecoregion (5), main type of soil (12), main type of livelihood zone (17), there were too many categories and it would have not been possible to include these data in the models without grouping some of these categories together. For example, desert type combined with other type of soil were grouped in the "desert category". The way the categories were grouped together could have influenced the results.[17]

[17] Ideally different sensitivity analysis would have been conducted, but due to time constraints they have not been performed (at this stage of the research).

7) Computational Methods, Implementation and (Goodness of) Fit

The procedure followed in the research, which was revised iterative to finally produce the expected output, included the following steps:

(a) Gather and systematize data from multiple sources corresponding to the hypotheses (explanatory variables) (as detailed in section 5 of this report)

(b) Use indicator kriging for extrapolating the values of variables for which there was not information for some districts (for example, percentage of villages with boy or girl schools inside the district) and simple kriging for prices. The detailed results per variable are included in the Appendix A of this report

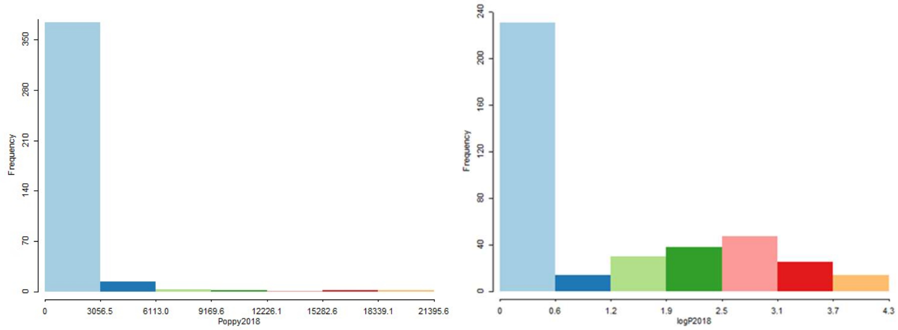
(c) Conduct descriptive statistics of the dependent variable (opium poppy cultivation area per district in 2018) and its correlation with all explanatory variables (included in Table 2 in this section). The corresponding visualizations for each explanatory variable are presented in Appendix B of this report.

(d) Conduct an analysis of the spatial autocorrelation of the dependent variable (see Table 3 in this section). The spatial weight matrix for which the highest Moran's I results was obtained for the dependent variable corresponded to the 4-nearest neighbour. Therefore, this matrix was used for all the spatial correlation models.

(e) The dependent variable was censored in zero (i.e., all opium poppy areas have positive values) with several zero values (i.e., district without opium poppy cultivation) (see its distribution in Figure 2 below). One of the most suitable models under these conditions is the Tobit model. Therefore, Tobit models without and with spatial autocorrelation were run. As the number of potential explanatory variables to be included in the models were in total 31, a stepwise approach was used to decide which ones to keep based on their significance and changes in the goodness of fit of the model (see Table 2 below)[18]. Normality assumption tests and goodness of fit results were used to decide if the Tobit models were actually appropriate (the results of the Tobit models and tests are shown in sub section 7.1 below).

[18] An alternative to the step-wise approach would have been to reduce the number of variables using principal component analysis techniques, which provide a combination of those variables. However, it is more difficult to interpret such combinations; and therefore, (at this stage of the research) only individual variables have been included as explanatory variables.

Figure 2: Distribution of opium poppy areas per district in 2018



(a) Without transformation (b) Log transformed

(f) After testing, all the Tobit models (with and without spatial autocorrelation correction) violated the normality assumption (of the error term), making them inappropriate. As such, the dependent variable was transformed to binary (1=presence of opium poppy inside the district, 0=no presence) for allowing running probit models with and without spatial correlation instead (The results of the probit models are included in sub section 7.2 below).

(g) The goodness of fit results of the probit suggest that the models were appropriate. The predicted values of the probit model with spatial correlation (created mostly with data from 2017) was used to create a map to check if the results resemble the situation actually observed in 2018 (see Figures from 3 to 5 in section 8 of this report).

Table 2: Results of the correlations between dependent and explanatory variables and order in which the explanatory variables were included in the stepwise Tobit models

Type of data per district	Variable name	Standardized correlation results[19]		Order of inclusion in the stepwise tobit regression	Included in the tobit regression
		Opium Poppy Area in 2018	Log of Opium Poppy Area in 2018		
Previous opium poppy cultivation in the district					
(1) Total estimated opium poppy area per district in 2017	Poppy2017	0.925**	0.579**	1	Yes (sig)
Geographic and climatic conditions					
(2) Digital elevation (mean altitude in meters)	Elev_mean	-0.200**	-0.241**	8	Yes (sig, pR2)
(3) Temperature a) annual mean at day b) annual mean at night in Kelvin	T2017DMean T2017NMean	0.289** 0.261**	0.288** 0.329**	6 4	No (NS) Yes (sig, pR2)
(4) Precipitation (annual mean in mm/day)	PP2017mean	-0.269**	-0.190**	15	No (NS)

[19] The appendix A includes the graphs and the non-standardized results for each variable

Type of data per district	Variable name	Standardized correlation results[19]		Order of inclusion in the stepwise tobit regression	Included in the tobit regression
		Opium Poppy Area in 2018	Log of Opium Poppy Area in 2018		
(5) Main type of ecoregion per district (10 ecoregions)	DesertEco	0.235**	0.131**	20	No (NS)
Security and governability conditions					
(6) Degree of (anti-) Government influence inside the district[20]	SIGAR2017	0.248**	0.224**	10	No (NS)
(7) Most likely estimate of total number of fatalities from incidents of lethal violence	Conf2017	0.315**	0.309**	5	No (NS)
(8) Total number of internally displaced families: a) emigrated b) immigrated c) net migration to the district (b-c)	 2017Emigra 2017Immigr 2017NetMig	 NS NS NS	 NS NS NS	 (a) (a) (a)	 No (NS) No (NS) No (NS)
(9) Total number of (registered) on-going aid projects[21]	N_IATI	NS	NS	(a)	No (NS)
Agricultural and market characteristics					
(10) Agricultural area in hectares: a) total b) irrigated c) rain-fed d) barren land	 TotalAgrA Irrig_land rainf_land barre_land	 0.157** 0.400** NS 0.334**	 0.224** 0.202** 0.137** 0.220**	 11 14 19 12	 Yes (sig, pR2) Yes (sig, pR2) Omitted No (NS)
(11) Total longitude of rivers per district (sum in miles)	River_L	NS	0.145**	17	No (NS)
(12) Main type of soil per district (18 types)	RockySoil	-0.157**	-0.188**	16	No (NS)
(13) Accessibility to cities (high density urban	Accessmean	NS	0.141**	18	Yes (sig, pR2)

[20] Where 1=Controlled by the Government of Afghanistan, 2=Influence of the Government of Afghanistan, 3=Contested, 4=Insurgent activity, 5=High insurgent activity

[21] Projects that started before 2018 and were still on-going in 2018

Type of data per district	Variable name	Standardized correlation results[19]		Order of inclusion in the stepwise tobit regression	Included in the tobit regression
		Opium Poppy Area in 2018	Log of Opium Poppy Area in 2018		
centers) in minutes of travel time (mean of the district)					
(14) Farm-gate prices of dried opium poppy (per Kg in USD)	Price17	0.140**	0.211**	13	No (NS)
(15) Prices (in Afghani) in eight major markets of: a) wheat b) wheat flour c) cooking oil	 Wheat_17 WheatF_17 COil_17	 NS NS NS	 NS NS NS	 (a) (a) (a)	 No (NS) No (NS) No (NS)
Demographic, socio-economic and presence of (dis)incentives					
(16) Population per district a) total b) rural	 pop_t_both pop_r_both	 NS 0.268**	 NS 0.276**	 (a) 7	 No (NS) Yes (sig,pR2)
(17) Main type of livelihood zone per district (29 zones where people share similar options for food and income and access to markets)	PastorLL	NS	NS	(a)	No (NS)
(18) Index of multi-dimensional poverty (per province)	Poverty.	0.190**	0.372**	3	Yes (sig,pR2)
(19) Percentage of villages with availability education and (des)incentives for opium poppy cultivation: a) boy school b) girl school c) advanced money for opium poppy cultivation e) awareness campaigns against opium poppy cultivation	 BoySchK17 GirlSchK17 AdvMonK17 AwareK17	 NS -0.294** 0.465** -0.190**	 NS -0.240** 0.477** NS	 (a) 9 2 (a)	 No (NS) No (NS) Yes (sig, pR2) Yes (sig, pR2)

**Significant at 0.05 level
NS=Not significant
(a) Included in the stepwise regression for evaluation (after the other variables were included)
Yes (sig, pR2)= yes, to be included (significant at 0.05 level and improvement in pseudo R2)
No (NS)=Not included (no significant and no improvement in pseudo R2)

Table 3: Moran's I results for the dependent variable using different weight matrices

Spatial weight matrix	**Moran's I**
Queen and Rook's contiguity	
Queen's contiguity, order 1	0.571***
Queen's contiguity, order 2	0.314***
Rook's contiguity, order 1	0.580***
Rook's contiguity, order 2	0.331***
K-nearest neighbours	
3-nearest neighbours	0.544***
4-nearest neighbours	**0.604****
5-nearest neighbours	0.571***
6-nearest neighbours	0.510***
7-nearest neighbours	0.504***
8-nearest neighbours	0.472***
Distance	
General distance, 200 km	0.281***
Inverse distance, 200 km, power 1	0.281***
Inverse distance, 200 km, power 2	0.281***
General distance, 300 km	0.173***
Inverse distance, 300 km, power 1	0.173***
Inverse distance, 300 km, power 2	0.173***
General distance, 400 km	0.104***
Inverse distance, 400 km, power 1	0.104***
Inverse distance, 400 km, power 2	0.104***
General distance, 500 km	0.057***
Inverse distance, 500 km, power 1	0.057***
Inverse distance, 500 km, power 2	0.057***

*** p=0.001 (with 999 permutations)

7.1 Tobit Model Results[22]

The Tobit model is relevant when the dependent variable is a mixture of observations with zero and positive values only (such as in the case of opium poppy cultivation). In that case, ordinary least square (OLS) regression will not yield consistent parameter estimates because the

[22] The Tobit models and tests were run in STATA 16.

dependent variable is censored (at zero, not negative values). The zero values can be interpreted as left-censored variable (y) that equals zero when y<L (where L is constant threshold). In particular, if the underlying data generating process is non-normal, the Tobit estimators are inconsistent.[23]

The results of the standard Tobit model (truncated at 0 and for which the dependent variable corresponded to the log of opium poppy areas per district in 2018) and with spatial autocorrelation correction (*i.e.,* spatial lag and spatial error cross sectional regressions), as well as the results of their corresponding normality assumption tests are presented below.

a) Tobit model (without spatial autocorrelation correction)

The selected explanatory variables of the standard Tobit model were statistically significant at least at 0.05 level. However, the McFadden's pseudo R-squared was not very high (0.14). Although it is not equivalent to a R-squared in a OLS regression, higher values are preferred because the pseudo R-squared is associated with the explained variability of the model, shows if there is an improvement from the null model to fitted model, and represents the square of the correlation.[24]

```
Fitting full model:

Iteration 0:   log likelihood = -737.48493
Iteration 1:   log likelihood = -636.76501
Iteration 2:   log likelihood = -615.48252
Iteration 3:   log likelihood = -611.57974
Iteration 4:   log likelihood = -611.48289
Iteration 5:   log likelihood = -611.48262
Iteration 6:   log likelihood = -611.48262

Tobit regression                                Number of obs     =        398
                                                   Uncensored     =        178
Limits: lower = 0                                  Left-censored  =        220
        upper = +inf                               Right-censored =          0

                                                LR chi2(10)       =     214.69
                                                Prob > chi2       =     0.0000
Log likelihood = -611.48262                     Pseudo R2         =     0.1493
```

logP2018	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
poppy2017	.000418	.0000917	4.56	0.000	.0002377	.0005984
advmonk17	5.365796	1.381012	3.89	0.000	2.650593	8.080999
poverty	12.83717	3.008536	4.27	0.000	6.922095	18.75224
tn2017mean	.0100678	.0022132	4.55	0.000	.0057164	.0144192
pop_r_both	.000023	7.36e-06	3.12	0.002	8.52e-06	.0000375
elev_mean	.0018282	.0007668	2.38	0.018	.0003205	.0033358
totalagra	.0000372	.0000127	2.92	0.004	.0000122	.0000622
irrig_land	-.0000574	.0000237	-2.42	0.016	-.0001041	-.0000108
accessmean	.005583	.0018794	2.97	0.003	.001888	.0092781
awarek17	1.327292	.7760159	1.71	0.088	-.1984302	2.853014
_cons	-152.1429	32.51258	-4.68	0.000	-216.0658	-88.22004
var(e.logP2018)	15.24887	1.828105			12.04682	19.30203

[23] For more details about the Tobit model, assumptions, and its specification see Cameron and Triveli. 2010. "Microeconometrics Using STATA. Revised Edition." STATA Press.

[24] UCLA. What are pseudo R2?. Available at: https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-what-are-pseudo-r-squareds/ (Accessed on 03 May 2020)

Normality test of the Tobit model (normality assumption violated)

Two tests were run for evaluating the normality assumption of the Tobit model: (1) a conditional moment test for assessing the null hypothesis that the disturbances in the Tobit model have a normal distribution and (2) a LM-statistics for testing the Tobit specification against the alternative of a model that is non-linear in the regressors and contains an error term that can be heteroskedastic and non-normally distributed.

The results of both normality tests are provided below:

The conditional moment test rejected the null hypothesis of normality with probability 0.0075.

```
Conditional moment test against the null of normal errors

    CM      Prob > chi2
  9.7741     0.00754
```

Similarly, the LM-statistics rejected the null hypothesis of normality (201.38 is larger than the critical values ranging from 3.4 to 10.4)

```
LM test of Tobit specification
                  Bootstrap critical values
   lm        %10       %5        %1
  201.38   3.38706  5.6181231 10.387964
```

b) Spatial Lag Tobit Model

The results of the Spatial Lag Tobit model, which assumes that the dependent variable is autocorrelated (i.e., opium poppy cultivation at one locations is affected by the opium poppy cultivation at the nearby locations), suggested that this specification was better than the Tobit model (the sigma was statistically significant at 0.01 level). Also, the Rho was statistically significant at 0.01 level indicating that (a general) Tobit model is preferred than an OLS model. The spatial weight matrix used for the Spatial Lag Tobit model was the 4-nearest neighbours (see results of the spatial autocorrelation of the dependent variable in Table 3).

```
 Sample Size       =         399
 Wald Test         =    238.9658   |   P-Value > Chi2(13)      =      0.0000
 F-Test            =     18.3820   |   P-Value > F(13 , 386)   =      0.0000
(Buse 1973) R2     =      0.3830   |   Raw Moments R2          =      0.1060
(Buse 1973) R2 Adj =      0.3638   |   Raw Moments R2 Adj      =      0.0782
 Root MSE (Sigma)  =      3.7472   |   Log Likelihood Function =   -326.6349
------------------------------------------------------------------------------
- R2h= 0.3830   R2h Adj= 0.3638  F-Test =   18.38 P-Value > F(13 , 386)0.0000
- R2v= 0.1585   R2v Adj= 0.1324  F-Test =    5.58 P-Value > F(13 , 386)0.0000
```

logP2018	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
logP2018						
Poppy2017	.0002814	.0000404	6.96	0.000	.0002021	.0003606
TN2017mean	.0006471	.0005989	1.08	0.280	-.0005267	.0018209
pop_r_both	2.90e-06	3.65e-06	0.79	0.427	-4.25e-06	.0000101
Irrig_land	-.0000275	.0000102	-2.70	0.007	-.0000475	-7.56e-06
barre_land	-2.54e-06	1.36e-06	-1.86	0.063	-5.21e-06	1.35e-07
total_land	1.22e-06	9.60e-07	1.27	0.205	-6.65e-07	3.10e-06
N_IATI	.0135256	.012338	1.10	0.273	-.0106564	.0377075
PP2017mean	-.3309941	.3033034	-1.09	0.275	-.9254578	.2634697
River_L	-9.25e-07	9.55e-07	-0.97	0.333	-2.80e-06	9.47e-07
Conf2017	.0007921	.0007077	1.12	0.263	-.000595	.0021792
SIGAR2017	.2874653	.1181945	2.43	0.015	.0558083	.5191224
Price17	.0056877	.0028544	1.99	0.046	.0000931	.0112823
Wheat_17	.1102134	.0569526	1.94	0.053	-.0014116	.2218385
_cons	-8.681269	8.821376	-0.98	0.325	-25.97085	8.608311
/Rho	.1865232	.0624454	2.99	0.003	.0641326	.3089139
/Sigma	1.467614	.0741181	19.80	0.000	1.322346	1.612883

```
 LR Test SAR vs. OLS (Rho=0):      8.9221   P-Value > Chi2(1)   0.0028
 Acceptable Range for Rho:        -1.9139   <  Rho  < 1.0000
------------------------------------------------------------------------------
=================================================================================
* Model Selection Diagnostic Criteria
=================================================================================
- Log Likelihood Function                   LLF            =   -326.6349
------------------------------------------------------------------------------
- Akaike Information Criterion              (1974) AIC     =     14.5713
- Akaike Information Criterion              (1973) Log AIC =      2.6791
------------------------------------------------------------------------------
- Schwarz Criterion                         (1978) SC      =     16.7604
- Schwarz Criterion                         (1978) Log SC  =      2.8190
------------------------------------------------------------------------------
```

Normality test of the Spatial Lag Tobit Model (normality assumption violated)

All the normality tests (Jarque-Bera, White, Doornik-Hansen, Pagan-Vella and Chesher-Irish) indicated that the errors are not normally distributed with probability 0.000.

```
==============================================================================
* Non Normality Tests
==============================================================================
 Ho: Normality - Ha: Non Normality
------------------------------------------------------------------------------
*** Non Normality Tests:
- Jarque-Bera LM Test                 =  36.8475     P-Value > Chi2(2) 0.0000
- White IM Test                       =  61.6715     P-Value > Chi2(2) 0.0000
- Doornik-Hansen LM Test              = 113.7713     P-Value > Chi2(2) 0.0000
```

```
LM Test - Ho: No Skewness
- LM Test                    =   28.5877   P-Value > Chi2(1)   0.0000

LM test - Ho: No Kurtosis
- LM Test                    =   45.1782   P-Value > Chi2(1)   0.0000

LM Test - Ho: Normality (No Kurtosis, No Skewness)
- Pagan-Vella LM Test        =   45.2771   P-Value > Chi2(2)   0.0000
- Chesher-Irish LM Test      =  389.3523   P-Value > Chi2(2)   0.0000
------------------------------------------------------------------------------
```

c) Spatial Error Tobit Model

The results of the Spatial Error Tobit model suggested that this specification was better than the standard Tobit model (the lambda or coefficient of the spatially correlated errors was statistically significant at 0.01 level)[25]. Also, the Sigma was statistically significant at 0.01 level.

However, the spatial lag model performed better than the spatial error model, as indicated by their lower values of Akaike Information Criterion (14.6 versus 16.9) and Schwarz Criterion (16.8 versus 19.5). This suggests that the spatial autocorrelation correction was needed at the level of the dependent variable (but not at the level of explanatory variables).

```
 Sample Size       =         399
 Wald Test         =    216.3449   |   P-Value > Chi2(13)      =      0.0000
 F-Test            =     16.6419   |   P-Value > F(13 , 386)   =      0.0000
(Buse 1973) R2     =      0.3598   |   Raw Moments R2          =     -0.0378
(Buse 1973) R2 Adj =      0.3399   |   Raw Moments R2 Adj      =     -0.0700
 Root MSE (Sigma)  =      4.0373   |   Log Likelihood Function =   -324.1333
------------------------------------------------------------------------------
- R2h= 0.3598   R2h Adj= 0.3399  F-Test =   16.64 P-Value > F(13 , 386)0.0000
- R2v= 0.1231   R2v Adj= 0.0958  F-Test =    4.16 P-Value > F(13 , 386)0.0000
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
logP2018						
Poppy2017	.0002784	.0000415	6.71	0.000	.0001971	.0003598
TN2017mean	.0000301	.0001628	0.18	0.853	-.000289	.0003491
pop_r_both	5.99e-06	3.83e-06	1.56	0.118	-1.52e-06	.0000135
Irrig_land	-.0000271	.0000105	-2.58	0.010	-.0000478	-6.53e-06
barre_land	-2.23e-06	1.44e-06	-1.55	0.121	-5.05e-06	5.88e-07
total_land	9.22e-07	9.77e-07	0.94	0.345	-9.92e-07	2.84e-06
N_IATI	.0082506	.0118658	0.70	0.487	-.015006	.0315071
PP2017mean	-.5508026	.3225058	-1.71	0.088	-1.182902	.0812972
River_L	-6.00e-07	1.01e-06	-0.60	0.550	-2.57e-06	1.37e-06
Conf2017	.0006445	.0006923	0.93	0.352	-.0007123	.0020014
SIGAR2017	.3219687	.1245418	2.59	0.010	.0778713	.5660661
Price17	.0066949	.0035287	1.90	0.058	-.0002212	.013611
Wheat_17	.1302472	.0647799	2.01	0.044	.0032809	.2572134
_cons	.521775	1.690993	0.31	0.758	-2.79251	3.83606
/Lambda	.313108	.070914	4.42	0.000	.1741192	.4520968
/Sigma	1.427401	.0735476	19.41	0.000	1.283251	1.571552

```
 LR Test SEM vs. OLS (Lambda=0):  19.4951   P-Value > Chi2(1)   0.0000
 Acceptable Range for Lambda:     -1.9139 < Lambda < 1.0000
------------------------------------------------------------------------------
```


[25] The spatial lag model relates the explanatory variables and the dependent variable as in a standard Tobit regression model, except that the dependent variable is auto- regressed on spatially lagged dependent variables. In contrast the spatial error model accounts for spatial dependence by an error term and an associated spatially lagged error term (Chi and Zhu. 2020. “Spatial Regression Models for the Social Sciences”. SAGE Publications)

```
==============================================================================
* Model Selection Diagnostic Criteria
==============================================================================
- Log Likelihood Function                  LLF            =   -324.1333
---------------------------------------------------------------------------
- Akaike Information Criterion             (1974) AIC     =     16.9150
- Akaike Information Criterion             (1973) Log AIC =      2.8282
---------------------------------------------------------------------------
- Schwarz Criterion                        (1978) SC      =     19.4561
- Schwarz Criterion                        (1978) Log SC  =      2.9682
```

Normality test of the Spatial Error Tobit Model (normality assumption violated)

As in the case of the spatial lag model, all the normality tests (Jarque-Bera, White, Doornik-Hansen, Pagan-Vella and Chesher-Irish) indicated that the errors were not normally distributed with probability 0.000.

```
==============================================================================
* Non Normality Tests
==============================================================================
 Ho: Normality - Ha: Non Normality
------------------------------------------------------------------------------
*** Non Normality Tests:
- Jarque-Bera LM Test              =  43.7095     P-Value > Chi2(2) 0.0000
- White IM Test                    =  55.5597     P-Value > Chi2(2) 0.0000
- Doornik-Hansen LM Test           = 156.2222     P-Value > Chi2(2) 0.0000


   LM Test - Ho: No Skewness
   - LM Test                   =   28.5877   P-Value > Chi2(1)   0.0000

   LM test - Ho: No Kurtosis
   - LM Test                   =   45.1782   P-Value > Chi2(1)   0.0000

   LM Test - Ho: Normality (No Kurtosis, No Skewness)
   - Pagan-Vella LM Test       =   45.2771   P-Value > Chi2(2)   0.0000
   - Chesher-Irish LM Test     =  389.3523   P-Value > Chi2(2)   0.0000
  ----------------------------------------------------------------------------
```


7.2 Probit Model Results[26]

The probit is a standard model of a mutually exclusive binary or dichotomous dependent variable. This kind of model focuses on the determinants of the probability *p* of the occurrence of one outcome rather than an alternative outcome that occurs with a probability *1-p*.[27] [28]

a) Probit model (without spatial autocorrelation correction)

The probit model showed a good performance with several statistically significant explanatory variables (at 0.05 level) and a high pseudo r-squared (0.67).

[26] The probit model was run in STATA 16 and the probit models with spatial autocorrelation correction in R.

[27] For more details about the Tobit model, assumptions, and its specification see Cameron and Triveli. 2010. "Microeconometrics Using STATA. Revised Edition." STATA Press.

[28] In comparison with the Tobit model, it is de facto impossible to test for normality in a probit model. The residual that should be normally distributed is the difference between the unobserved latent variable and the predicted values. For this model, the normality assumption governs the functional form relating the explanatory variables with the probability (see Wooldridge.2002. "Econometric Analysis of Cross Section and Panel Data" MIT Press.)

```
Iteration 0:   log likelihood = -274.24418
Iteration 1:   log likelihood = -190.10298
Iteration 2:   log likelihood = -167.08018
Iteration 3:   log likelihood = -136.05774
Iteration 4:   log likelihood = -116.23376
Iteration 5:   log likelihood = -97.057326
Iteration 6:   log likelihood = -89.124286
Iteration 7:   log likelihood =  -88.84152
Iteration 8:   log likelihood = -88.840756
Iteration 9:   log likelihood = -88.840756

Probit regression                               Number of obs     =        399
                                                LR chi2(11)       =     370.81
                                                Prob > chi2       =     0.0000
Log likelihood = -88.840756                     Pseudo R2         =     0.6761

------------------------------------------------------------------------------
 P2018_Dummy |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   Poppy2017 |   .0377668   .0054988     6.87   0.000     .0269893    .0485442
  Accessmean |   .0025353   .0008781     2.89   0.004     .0008143    .0042564
  pop_r_both |   5.84e-06   3.31e-06     1.76   0.078    -6.52e-07    .0000123
   TotalAgrA |   .0000169   6.41e-06     2.64   0.008     4.33e-06    .0000295
      N_IATI |  -.0136892   .0117187    -1.17   0.243    -.0366574    .0092789
   AdvMonK17 |    2.51947   .9439136     2.67   0.008     .6694333    4.369507
    PastorLL |  -.5616947   .3571781    -1.57   0.116    -1.261751    .1383616
  TN2017mean |   .0014555   .0006047     2.41   0.016     .0002704    .0026406
    Wheat_17 |  -.1525791   .0708454    -2.15   0.031    -.2914335   -.0137247
   RiceHQ_17 |   .1049283   .0347708     3.02   0.003     .0367789    .1730778
   DesertEco |  -.6890912    .365837    -1.88   0.060    -1.406119     .027936
       _cons |  -27.51208   9.795051    -2.81   0.005    -46.71003   -8.314136
------------------------------------------------------------------------------
```

b) Spatial Lag Probit Model[29]

The results of the Bayesian estimation of the Spatial Lag Probit model are shown below. The spatial correlation parameter (rho) was significant at 0.01 level suggesting that the specification is adequate. The spatial weight matrix used was the 4-nearest neighbours (see results of the spatial autocorrelation of the dependent variable in Table 3).

```
--------MCMC spatial autoregressive probit--------
Execution time  =  8.320 secs

N draws         =   1000, N omit (burn-in)=    100
N observations  =    399, K covariates    =     12
# of 0 Y values =    221, # of 1 Y values =    178
Min rho         = -1.000, Max rho         =  1.000
--------------------------------------------------

              Estimate   Std. Dev    p-level t-value Pr(>|z|)
(Intercept) -1.982e+01  6.864e+00  4.000e-03  -2.888 0.004086 **
Poppy2017    2.221e-03  6.488e-04  0.000e+00   3.423 0.000683 ***
Accessmean   1.563e-03  7.113e-04  1.000e-02   2.197 0.028565 *
pop_r_both   6.447e-06  2.764e-06  8.000e-03   2.333 0.020150 *
TotalAgrA    9.756e-06  4.760e-06  7.000e-03   2.050 0.041037 *
N_IATI      -7.779e-03  7.462e-03  1.520e-01  -1.042 0.297837
AdvMonK17    3.085e+00  8.012e-01  0.000e+00   3.851 0.000137 ***
PastorLL    -1.009e-01  2.175e-01  3.310e-01  -0.464 0.642949
TN2017mean   1.037e-03  4.323e-04  8.000e-03   2.399 0.016917 *
Wheat_17    -8.850e-02  5.804e-02  6.000e-02  -1.525 0.128068
RiceHQ_17    7.289e-02  2.535e-02  1.000e-03   2.875 0.004252 **
DesertEco   -7.515e-01  3.058e-01  1.200e-02  -2.457 0.014426 *
rho          2.468e-01  8.469e-02  0.000e+00   2.914 0.003770 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

[29] The spatial error probit model could not run properly, and therefore, the results are not included in this report. Also, as a reference, the spatial error Tobit model was not preferred in comparison to the spatial lag Tobit model.

```
--------Marginal Effects--------

$Direct
           marginal.effect standard.error    z.ratio
Poppy2017     3.502179e-04   1.158040e-04  3.0242308
Accessmean    2.794489e-04   1.405577e-04  1.9881445
pop_r_both    1.309925e-06   5.741638e-07  2.2814487
TotalAgrA     1.889329e-06   9.482190e-07  1.9925026
N_IATI       -1.993261e-03   1.878461e-03 -1.0611137
AdvMonK17     5.836276e-01   1.367038e-01  4.2692856
PastorLL     -1.007654e-02   4.922874e-02 -0.2046882
TN2017mean    2.008938e-04   8.774068e-05  2.2896307
Wheat_17     -1.597988e-02   1.011532e-02 -1.5797706
RiceHQ_17     1.304852e-02   4.727018e-03  2.7604119
DesertEco    -1.532806e-01   5.827731e-02 -2.6301935

$Indirect
           marginal.effect standard.error    z.ratio
Poppy2017     1.345796e-04   3.407377e-05  3.9496544
Accessmean    1.170509e-04   8.028948e-05  1.4578607
pop_r_both    5.923637e-07   4.204508e-07  1.4088777
TotalAgrA     8.451672e-07   6.602355e-07  1.2800995
N_IATI       -9.996345e-04   1.223684e-03 -0.8169058
AdvMonK17     2.477934e-01   1.077970e-01  2.2987048
PastorLL     -4.133509e-03   2.486140e-02 -0.1662621
TN2017mean    8.760442e-05   5.994906e-05  1.4613144
Wheat_17     -6.489759e-03   4.773647e-03 -1.3594970
RiceHQ_17     5.425310e-03   2.675116e-03  2.0280650
DesertEco    -6.935063e-02   4.935961e-02 -1.4050076

$Total
           marginal.effect standard.error    z.ratio
Poppy2017     4.847975e-04   1.298567e-04  3.7333276
Accessmean    3.964998e-04   2.001818e-04  1.9806989
pop_r_both    1.902289e-06   9.132405e-07  2.0830100
TotalAgrA     2.734496e-06   1.476022e-06  1.8526122
N_IATI       -2.992895e-03   2.972803e-03 -1.0067588
AdvMonK17     8.314210e-01   1.913582e-01  4.3448413
PastorLL     -1.421005e-02   7.229642e-02 -0.1965526
TN2017mean    2.884982e-04   1.317098e-04  2.1904085
Wheat_17     -2.246964e-02   1.404324e-02 -1.6000320
RiceHQ_17     1.847383e-02   6.404099e-03  2.8846876
DesertEco    -2.226312e-01   9.717509e-02 -2.2910318
```

8) Validation

Areas in dark red show districts with high predicted opium poppy cultivation calculated mostly with data from 2017 (or before) using the spatial lag probit model (Figure 3); predicted presence of opium poppy from the spatial lag probit model (Figure 4); and actual presence of opium poppy cultivation in 2018 from UNODC's annual monitoring report (Figure 5).

In general, the spatial lag probit model correctly predicted the presence or absence of opium poppy correctly in 342 or 86 percent of the districts; while the number of false positives was 41 (or 10 percent of the districts); and the number of false negatives was 16 (or 4 percent of the districts). Therefore, the accuracy of the spatial lag model was of 86 percent, which for socio-economic models is good or relatively high.[30]

[30] For example, socio-economic models were categorized as of high accuracy when they correctly predicted 85% of the cases (see Ren et al. 2019. Predicting socio-economic levels of urban regions via offline and online indicators. Plos One 14:7)

Figure 3: Map of the predicted values of the spatial lag probit model (combined odd ratios)

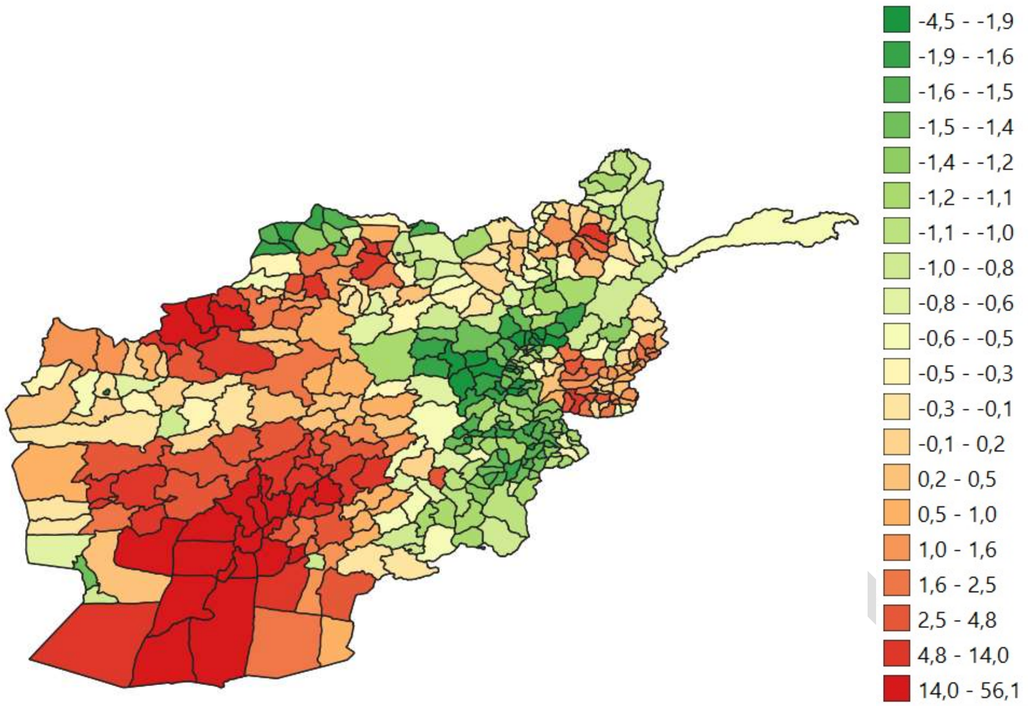


Figure 4: Map of the predicted presence or absence of opium poppy cultivation of the spatial lag probit model

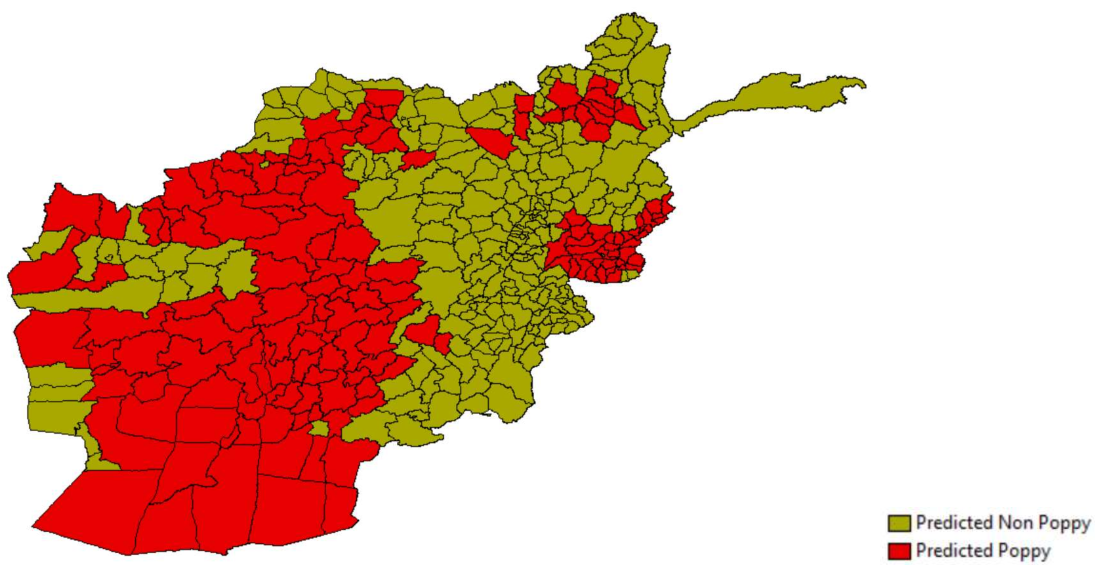
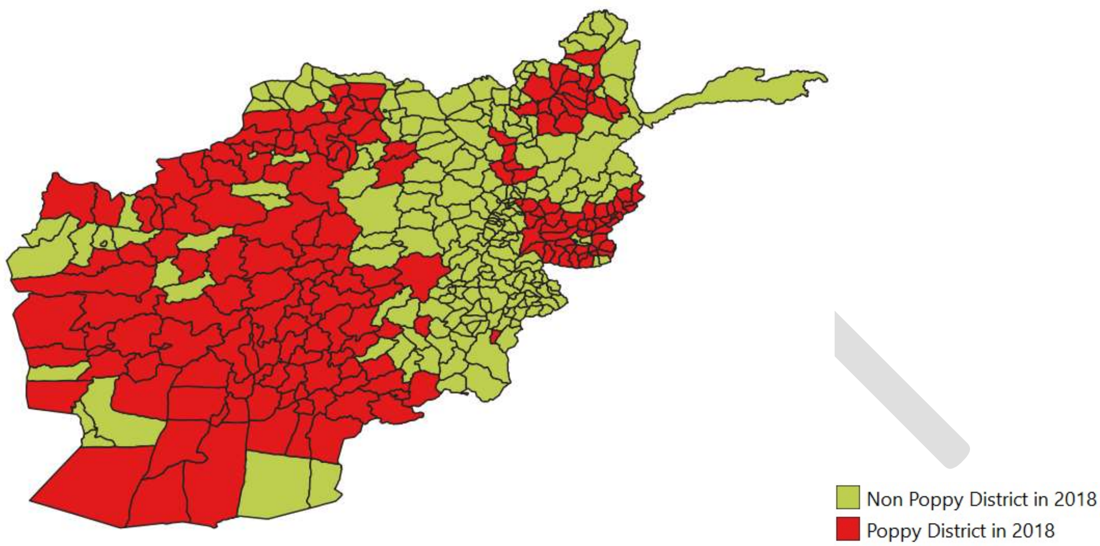


Figure 5: Map of the presence or absence of opium poppy cultivation per district in 2018 (from UNODC's Annual Monitoring Report)



9) Interpretation of Results and Future Activities /Research

The accuracy of the spatial lag model was of 86 percent and the results suggest that the probability of the presence of opium poppy cultivation inside a district was statistically significantly associated with the following (at 0.05 level) [31]:

-average accessibility of the district in minutes of travel time (10 additional minute that takes to access the district would increase the probability of the presence of opium poppy cultivation in 4 percent)

-size of the rural population inside the district (100 additional persons would increase the probability in 0.2 percent)

-availability of advanced money inside the district in the previous year (it would increase the probability in 83 percent)

-average temperature at night inside the district (increments of one Kelvin would increase the probability in 0.3 percent)

-market price of staples/rice (one additional Afghani would increase the probability in 2 percent)

-total agricultural area inside the district (100 additional hectares of agricultural land would increase the probability in 0.3 percent)

-land coverage being mostly desert inside the district (it would decrease the probability in 22 percent)

[31] The magnitudes reported correspond to the total marginal effect of the spatial lag probit model (see section 7.2.b), which are the sum of the direct effect and the indirect effects. The direct marginal effects are the expected average change across all observations for the dependent variable in a particular region due to increase of one unit for a specific explanatory variable in this region. The indirect marginal effects are the "spill-over effects" or the changes in the dependent variable of a particular region from a one unit increase in an explanatory variable in another region (Golgher and Voss. 2016. "How to Interpret the Coefficients of Spatial Models: Spillovers, Direct and Indirect Effects". Spatial Demography 4: 175-205)

-total opium poppy cultivation inside the district during the previous year (100 additional hectares would increase the probability in 5 percent)

Interestingly, the probability of the presence of opium poppy cultivation in a district was not statistically significantly associated with:
-farm-gate prices of opium poppy
-degree of anti-government influence inside the district
-number of fatalities from incidents of lethal violence
-total number of internally displaced families
-index of multi-dimensional poverty (measured per province)
-average precipitation inside the district (which is directly associated with droughts and floods)

In summary, the results suggest that opium poppy cultivation is more likely to occur in districts with more availability of agricultural land (and not having deserts as main type of eco-region), more people living in rural areas, more difficult accessibility (*i.e.,* it takes more time to reach them), where opium poppy cultivation occurred the previous year and there were external incentives for opium poppy cultivation such as advanced money, higher market prices of staples (i.e., rice) and higher average temperature at night (which could be associated with the lack of crop losses due to freezing temperatures at night). Security and multi-dimensional poverty factors do not seem to play a significant role in the presence or absence of opium poppy inside districts.

However, the results so far do not take into account the extension of opium poppy cultivation inside each district (as the normality assumption was violated when running Tobit models incorporating this aspect, which lead to inconsistent estimates). Also, this model considers mostly data from 2017 in an attempt to predict opium poppy cultivation in 2018. There was a severe drought in 2018 which could have distorted the predictions that could be non-longer considered valid for "normal" years. Future work will extent the current results by (a) running other models similar to Tobit for accounting for the extent of opium poppy cultivation and for which the normality assumption may not be violated (such as Censored or Cragg models), and (b) incorporating data from previous years (before 2017) and integrating more explicitly the temporal component in the regressions.

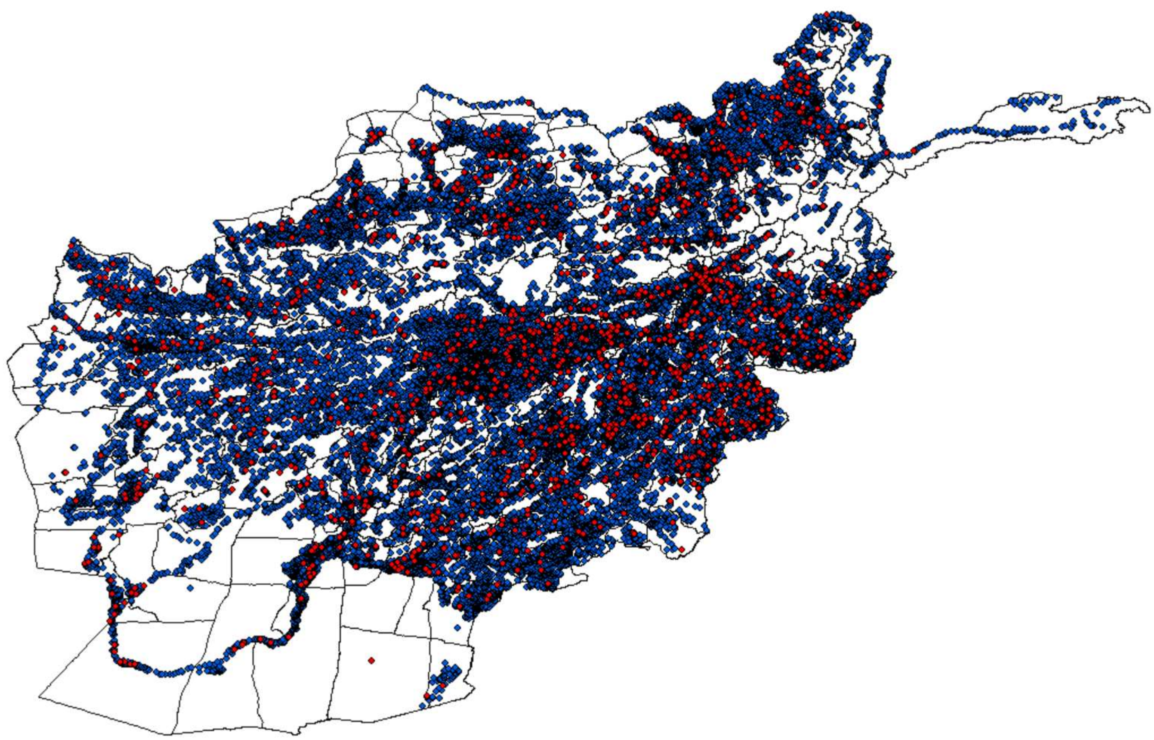
Appendices

A. Kriging results

Main procedure per socio-economic variable

The socio-economic values from the sample of villages were extrapolated (using kriging) to all the villages (in the sampling frame).

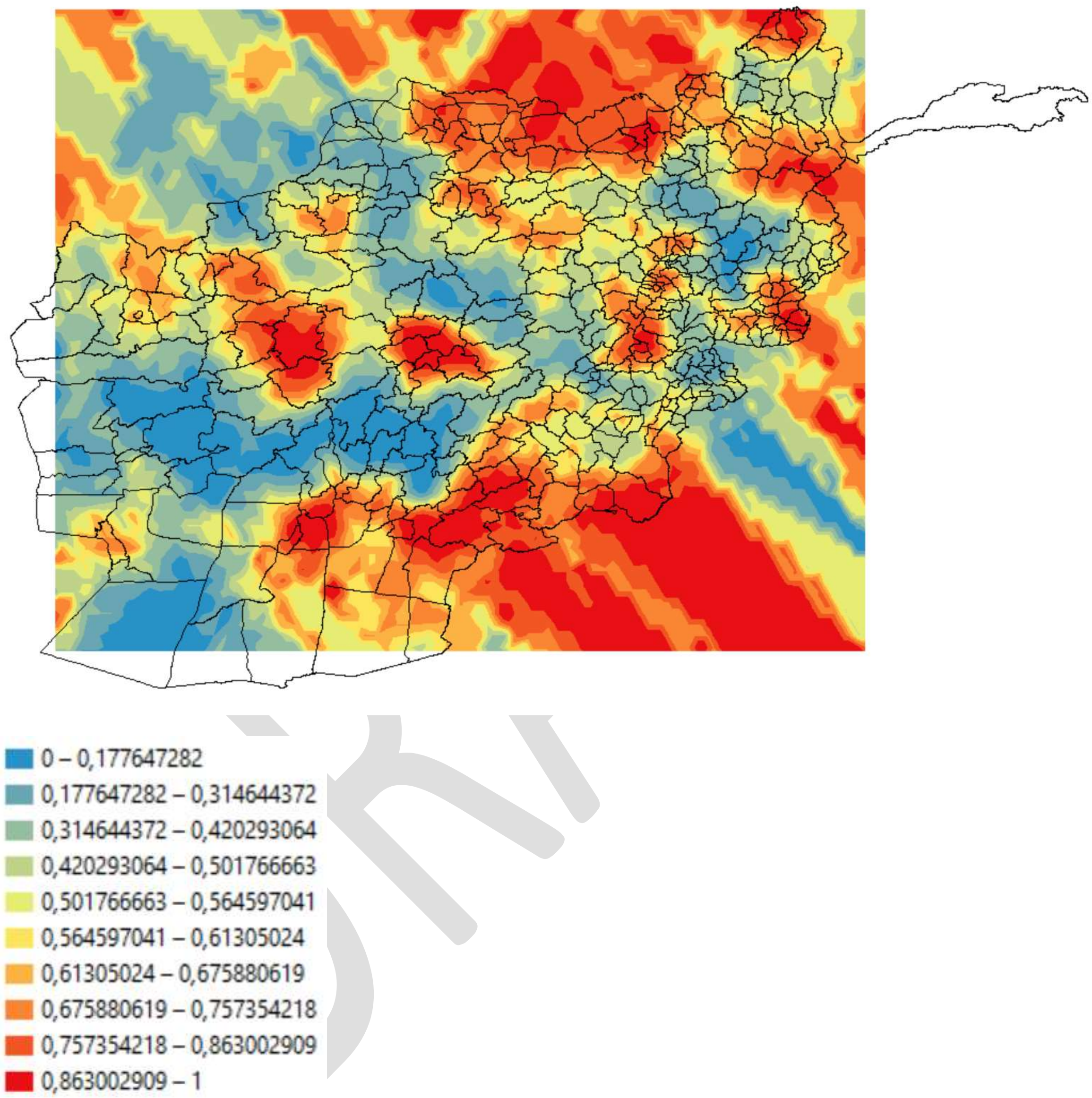
Figure 1: Location of the sample of villages in red (2017) and all villages in sampling frame in blue

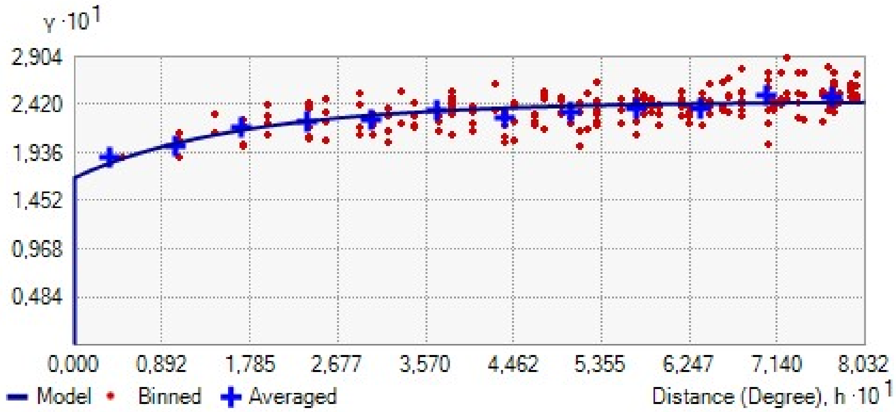


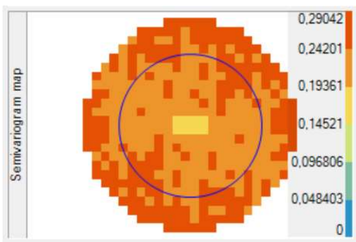
For doing so, kriging values were obtained for all the surface per variable, and then the corresponding kriging value of each village was extracted. The value assigned to each district corresponded to the average of the values of all the villages (of the village frame) inside the district.

Results for boy schools

Figure 2: Kriging results for boy schools





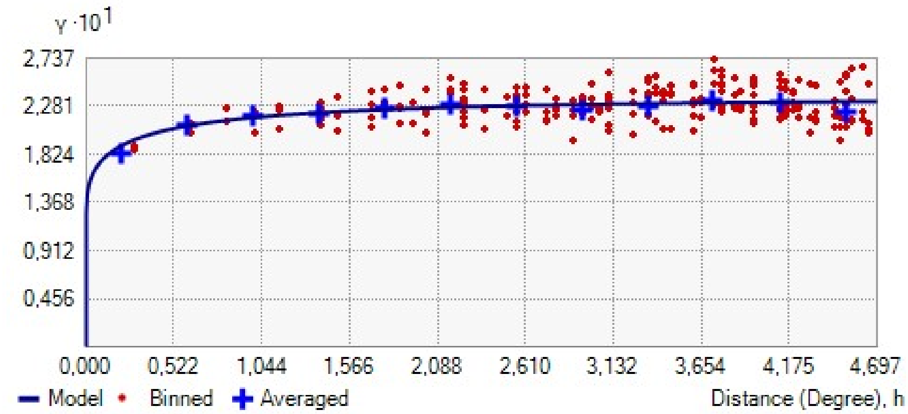


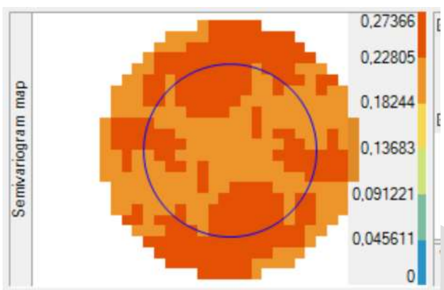
Prediction Errors	
Samples	1377 of 1377
Mean	-0,001136101
Root-Mean-Square	0,4629481
Mean Standardized	-0,002400829
Root-Mean-Square Standardized	0,9741343
Average Standard Error	0,4760282

Input datasets

⊟ **Dataset**	C:\Users\jgaryi\Desktop\Geostatistics\Project\AllDataMap\Socio-econ(2009-2017)\SE2017
Type	Feature Class
Data field 1	Boy_school
Records	1377
⊟ **Method**	**Kriging**
Type	Indicator
Output type	Probability
Threshold (Exceed)	0
⊟ **Dataset #**	1
⊟ **Searching neighborhood**	Standard
Neighbors to include	5
Include at least	2
Sector type	Four and 45 degree
Major semiaxis	0.535465577987
Minor semiaxis	0.535465577987
Angle	0
⊟ **Variogram**	Semivariogram
Number of lags	12
Lag size	0.066933197248
Nugget	0.168487491888
Measurement error %	0
⊟ **Model type**	Stable
Parameter	0.9734375
Range	0.535465577987
Anisotropy	No
Partial sill	0.076804004881

Results for girl schools



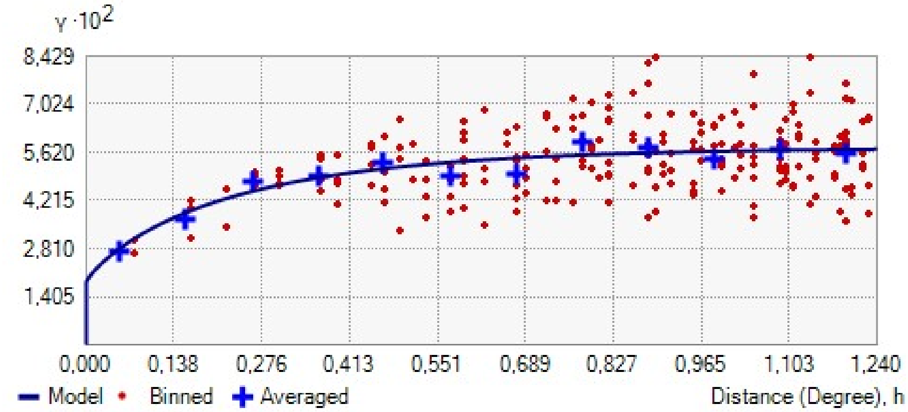


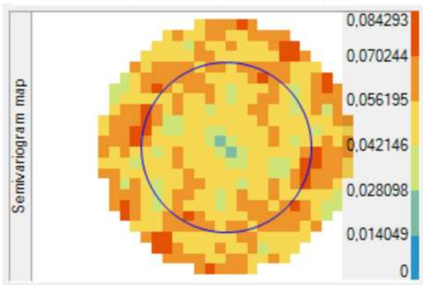
Prediction Errors	
Samples	1377 of 1377
Mean	-0,0005370128
Root-Mean-Square	0,4241599
Mean Standardized	-0,001450404
Root-Mean-Square Standardized	0,9603289
Average Standard Error	0,4423079

Input datasets

⊟ **Dataset**	C:\Users\jgaryi\Desktop\Geostatistics\Project\AllDataMap\Socio-econ(2009-2017)\SE2017
Type	Feature Class
Data field 1	Girl_schoo
Records	1377
⊟ **Method**	**Kriging**
Type	Indicator
Output type	Probability
Threshold (Exceed)	0
⊟ **Dataset #**	1
⊟ **Searching neighborhood**	Standard
Neighbors to include	5
Include at least	2
Sector type	Four and 45 degree
Major semiaxis	3.131601926874
Minor semiaxis	3.131601926874
Angle	0
⊟ **Variogram**	Semivariogram
Number of lags	12
Lag size	0.391450240859
Nugget	0.129091374328
Measurement error %	0
⊟ **Model type**	Stable
Parameter	0.4583984375
Range	3.131601926874
Anisotropy	No
Partial sill	0.106562815644

Results for advanced money



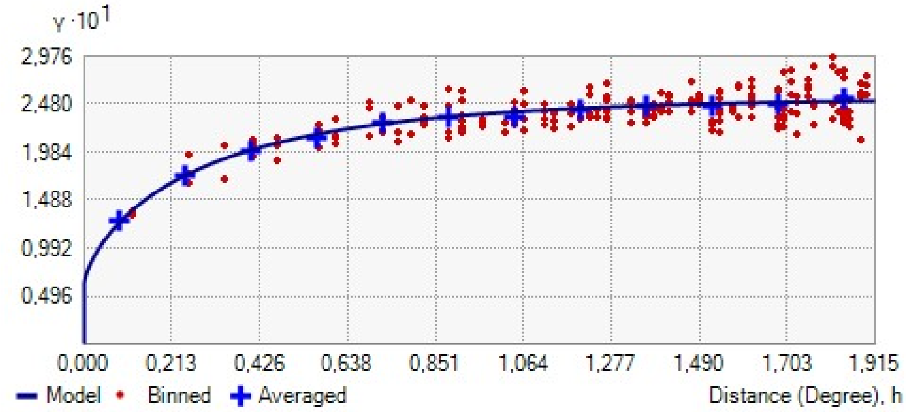


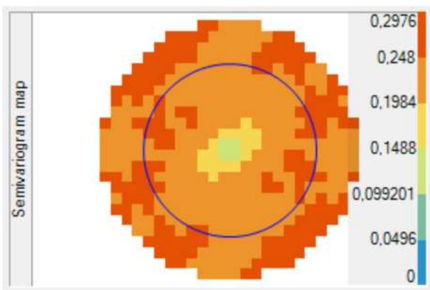
Prediction Errors	
Samples	1377 of 1377
Mean	-0,0002528871
Root-Mean-Square	0,2052665
Mean Standardized	-0,0001978592
Root-Mean-Square Standardized	1,056931
Average Standard Error	0,1950061

Input datasets

⊟ **Dataset**	C:\Users\jgaryi\Desktop\Geostatistics\Project\AllDataMap\SE_Kriging_2017 \SE_2017
Type	Feature Class
Data field 1	AdvMoney
Records	1377
⊟ **Method**	**Kriging**
Type	Indicator
Output type	Probability
Threshold (Exceed)	0
⊟ **Dataset #**	1
⊟ **Searching neighborhood**	Standard
Neighbors to include	5
Include at least	2
Sector type	Four and 45 degree
Major semiaxis	0.826887670952
Minor semiaxis	0.826887670952
Angle	0
⊟ **Variogram**	Semivariogram
Number of lags	12
Lag size	0.103360958869
Nugget	0.018496962342
Measurement error %	0
⊟ **Model type**	Stable
Parameter	0.8556640625
Range	0.826887670952
Anisotropy	No
Partial sill	0.039232287582

Results for Awareness Campaigns





Prediction Errors	
Samples	1377 of 1377
Mean	0,002186489
Root-Mean-Square	0,3646574
Mean Standardized	0,003088183
Root-Mean-Square Standardized	0,9584005
Average Standard Error	0,3849178

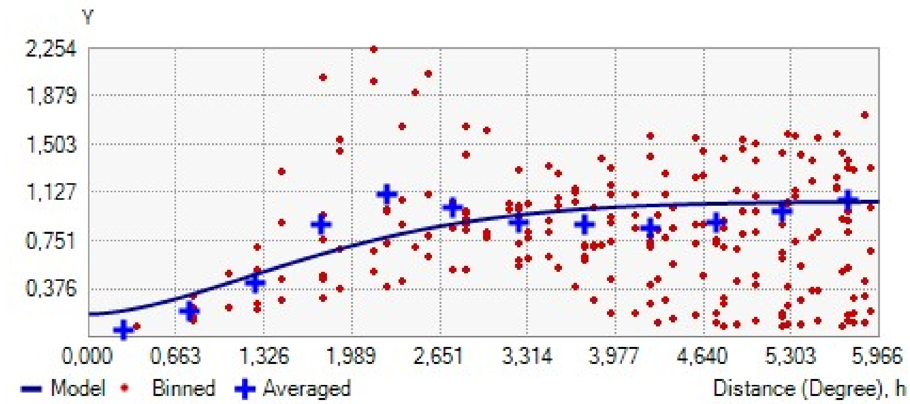
Input datasets

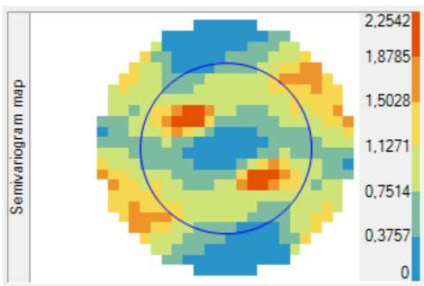
⊟ **Dataset**
C:\Users\jgaryi\Desktop\Geostatistics\Project\AllDataMap\SE_Kriging_2017\SE_2017

Type	Feature Class
Data field 1	AwareCpg
Records	1377

⊟ **Method**	**Kriging**
Type	Indicator
Output type	Probability
Threshold (Exceed)	0
⊟ **Dataset #**	1
⊟ **Searching neighborhood**	Standard
Neighbors to include	5
Include at least	2
Sector type	Four and 45 degree
Major semiaxis	1.276913490766
Minor semiaxis	1.276913490766
Angle	0
⊟ **Variogram**	Semivariogram
Number of lags	12
Lag size	0.159614186346
Nugget	0.063996648233
Measurement error %	0
⊟ **Model type**	Stable
Parameter	0.751953125
Range	1.276913490766
Anisotropy	No
Partial sill	0.190633742212

Price of opium poppy in 2017





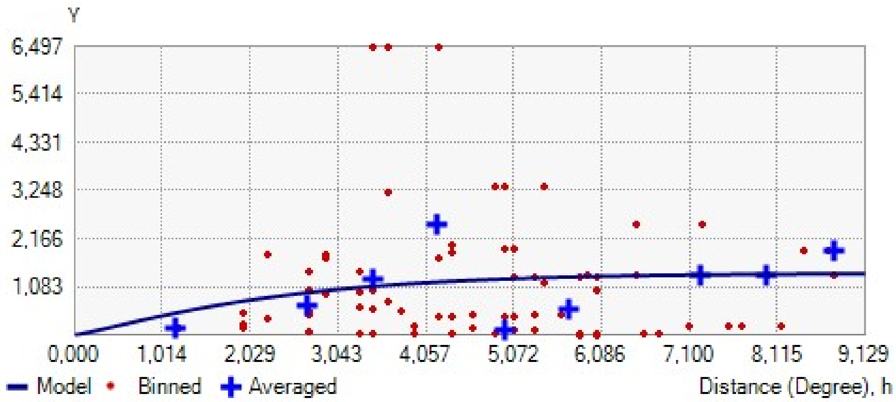
Regression function	
Prediction Errors	
Samples	258 of 258
Mean	1,478736
Root-Mean-Square	12,68626
Mean Standardized	0,0989226
Root-Mean-Square Standardized	0,5768768
Average Standard Error	23,00879

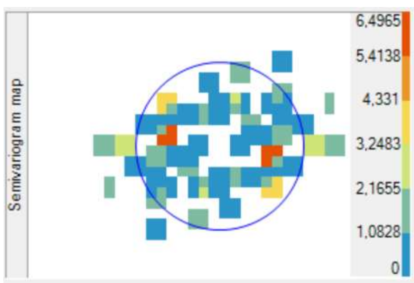
Input datasets

⊟ **Dataset**	C:\Users\jgaryi\Desktop\SelectedPr18
Type	Feature Class
Data field 1	Pr18
Records	258

⊟ **Method**	**Kriging**
Type	Simple
Output type	Prediction
⊟ **Dataset #**	1
Trend type	None
⊟ **Transformation**	Normal Score Transformation
Approximation	DensitySkew
Kernels	8
BaseDistribution	Empirical
⊟ **Searching neighborhood**	Standard
Neighbors to include	5
Include at least	2
Sector type	Four and 45 degree
Major semiaxis	3.977032326573
Minor semiaxis	3.977032326573
Angle	0
⊟ **Variogram**	Semivariogram
Number of lags	12
Lag size	0.497129040822
Nugget	0.177033302617
Measurement error %	100
⊟ **Model type**	Stable
Parameter	1.6765625
Range	3.977032326573
Anisotropy	No
Partial sill	0.879340064603

Price of wheat 2017





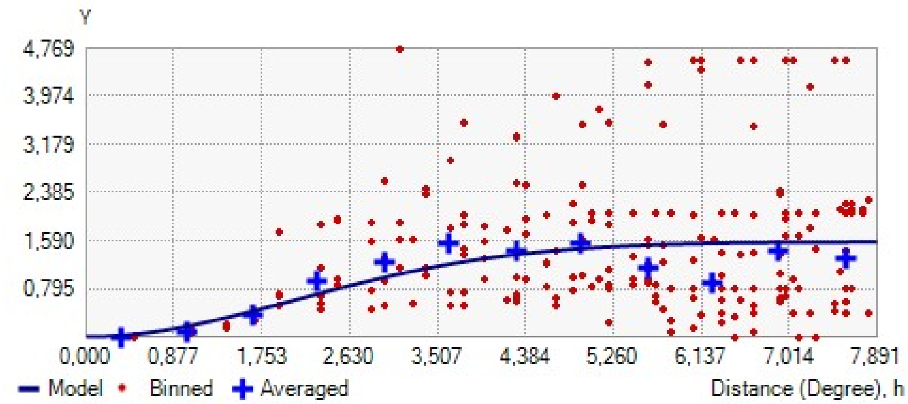
Prediction Errors	
Samples	8 of 8
Mean	0,3698807
Root-Mean-Square	2,858309
Mean Standardized	0,09863792
Root-Mean-Square Standardized	0,9650951
Average Standard Error	2,888203

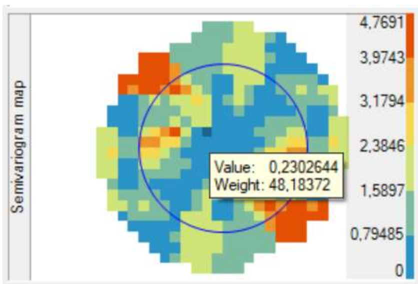
Input datasets

⊟ **Dataset**	C:\Users\jgaryi\Desktop\SelPrWheat17
Type	Feature Class
Data field 1	PrWheat17
Records	8

⊟ **Method**	**Kriging**
Type	Simple
Output type	Prediction
⊟ **Dataset #**	1
Trend type	None
⊟ **Transformation**	Normal Score Transformation
Approximation	DensitySkew
Kernels	1
BaseDistribution	Gamma
⊟ **Searching neighborhood**	Standard
Neighbors to include	5
Include at least	2
Sector type	Four and 45 degree
Major semiaxis	6.086105026709
Minor semiaxis	6.086105026709
Angle	0
⊟ **Variogram**	Semivariogram
Number of lags	12
Lag size	0.760763128339
Nugget	0
Measurement error %	100
⊟ **Model type**	Stable
Parameter	1.166796875
Range	6.086105026709
Anisotropy	No
Partial sill	1.387655393651

Price of wheat flour in 2017





Prediction Errors	
Samples	136 of 136
Mean	0,002801229
Root-Mean-Square	0,2645837
Mean Standardized	-0,002228792
Root-Mean-Square Standardized	2,483183
Average Standard Error	0,1679969

Input datasets

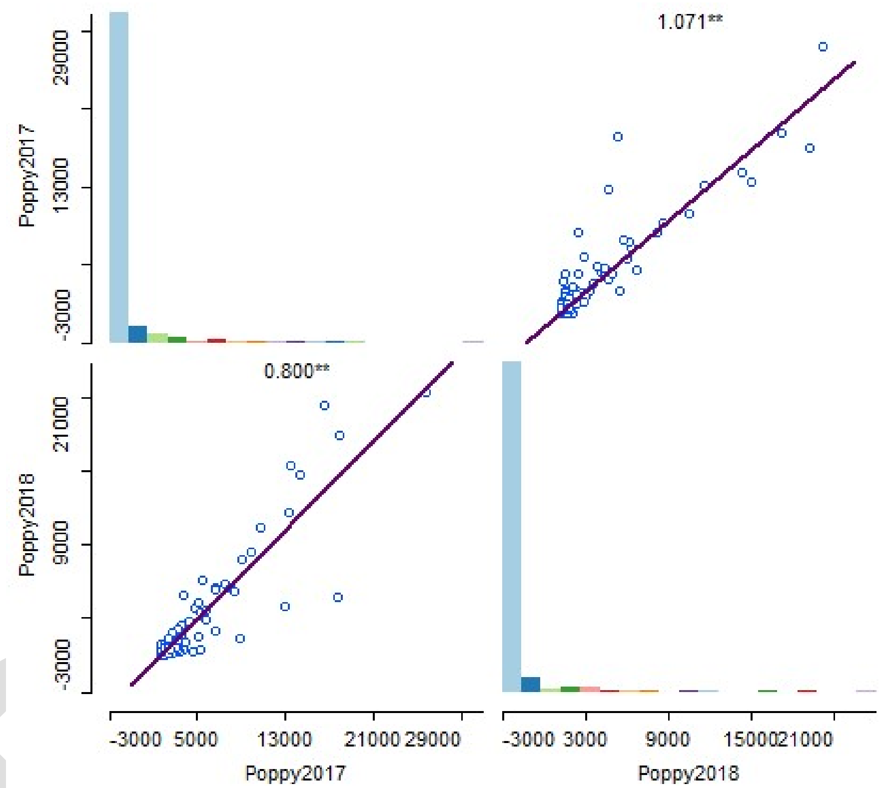
⊟ **Dataset**	C:\Users\jgaryi\Desktop\SelLegal
Type	Feature Class
Data field 1	Pr_W17
Records	136

⊟ **Method**	**Kriging**
Type	Simple
Output type	Prediction
⊟ **Dataset #**	1
Trend type	None
⊟ **Transformation**	Normal Score Transformation
Approximation	DensitySkew
Kernels	4
BaseDistribution	Gamma
⊟ **Searching neighborhood**	Standard
Neighbors to include	5
Include at least	2
Sector type	Four and 45 degree
Major semiaxis	5.260472978732
Minor semiaxis	5.260472978732
Angle	0
⊟ **Variogram**	Semivariogram
Number of lags	12
Lag size	0.657559122342
Nugget	0
Measurement error %	100
⊟ **Model type**	Stable
Parameter	1.95078125
Range	5.260472978732
Anisotropy	No
Partial sill	1.569629098438

B. Exploratory analysis per variable (non-standardized correlations between opium poppy in 2018 and potential explanatory variables)

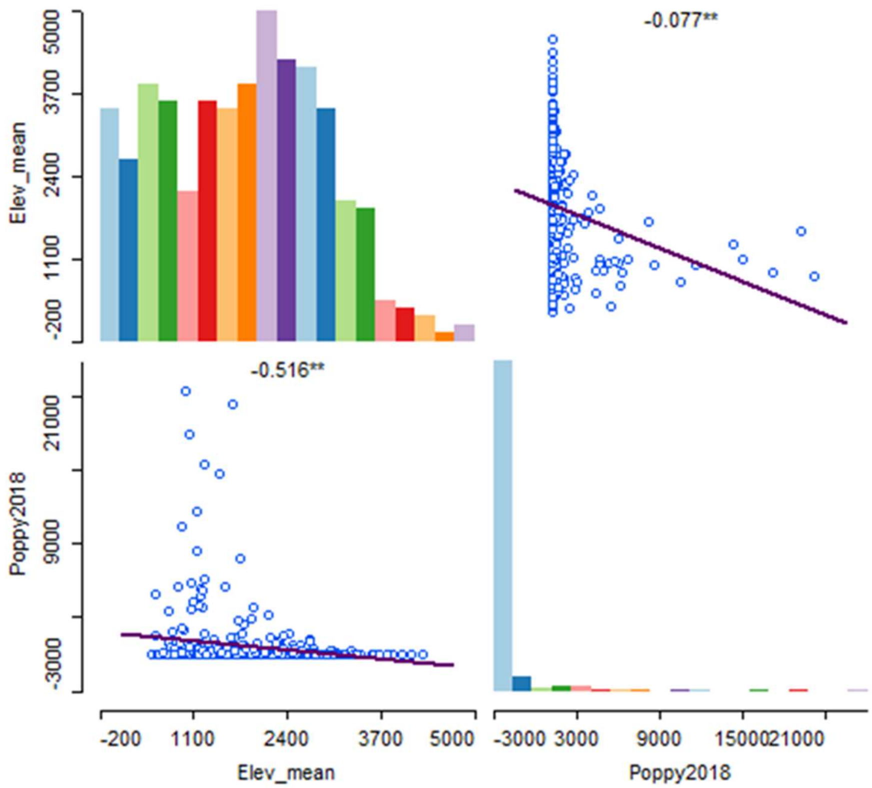
Scatter plot matrix of previous opium poppy area (2017) and opium poppy area in hectares in 2018

Significant positive correlation



Scatter plot matrix of mean altitude in meters of the district and opium poppy area in hectares in 2018

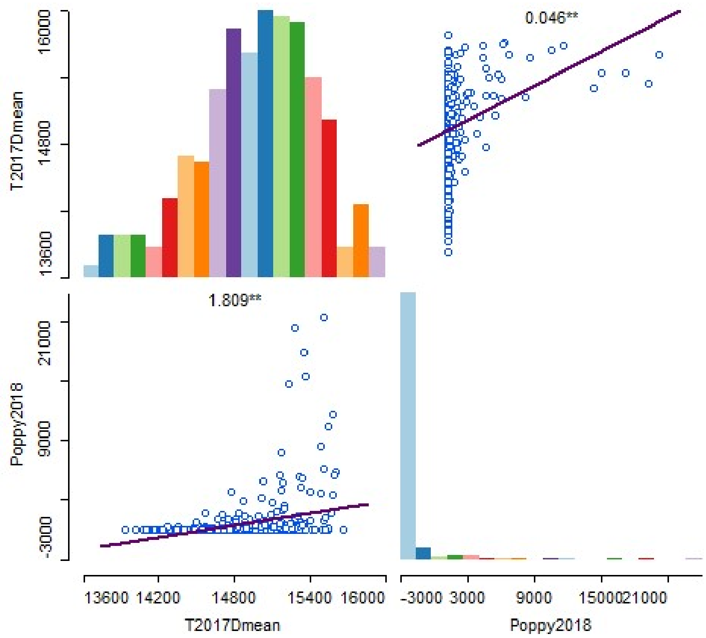
Significant negative correlation



Scatter plot matrix of mean temperature (in Kelvin) of the district in 2017 and opium poppy areas in hectares in 2018

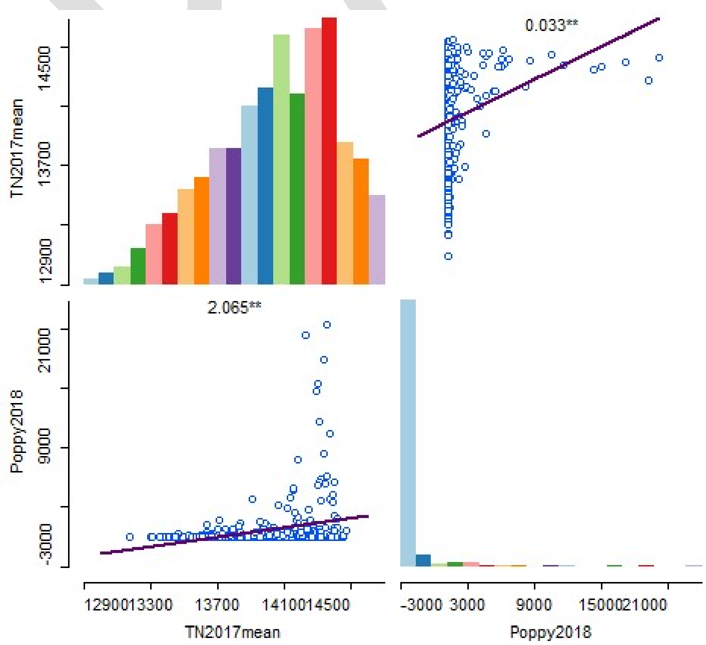
During day

Significant positive correlation



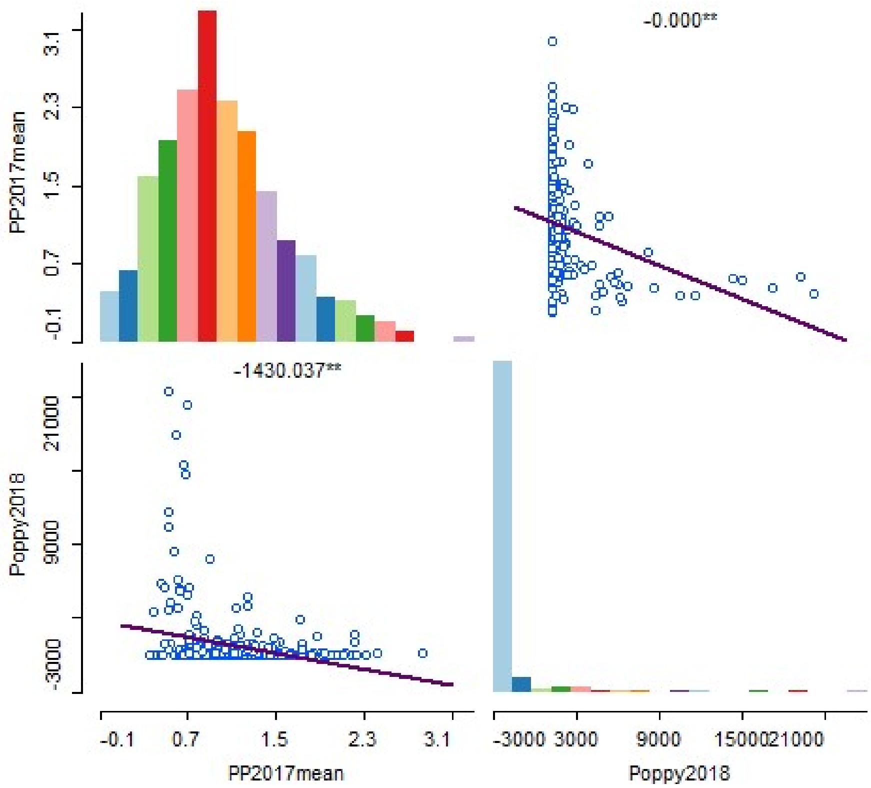
During night

Significant positive correlation



Scatter plot matrix of mean precipitation (mm/day) of the district in 2017[32] and opium poppy areas in hectares in 2018

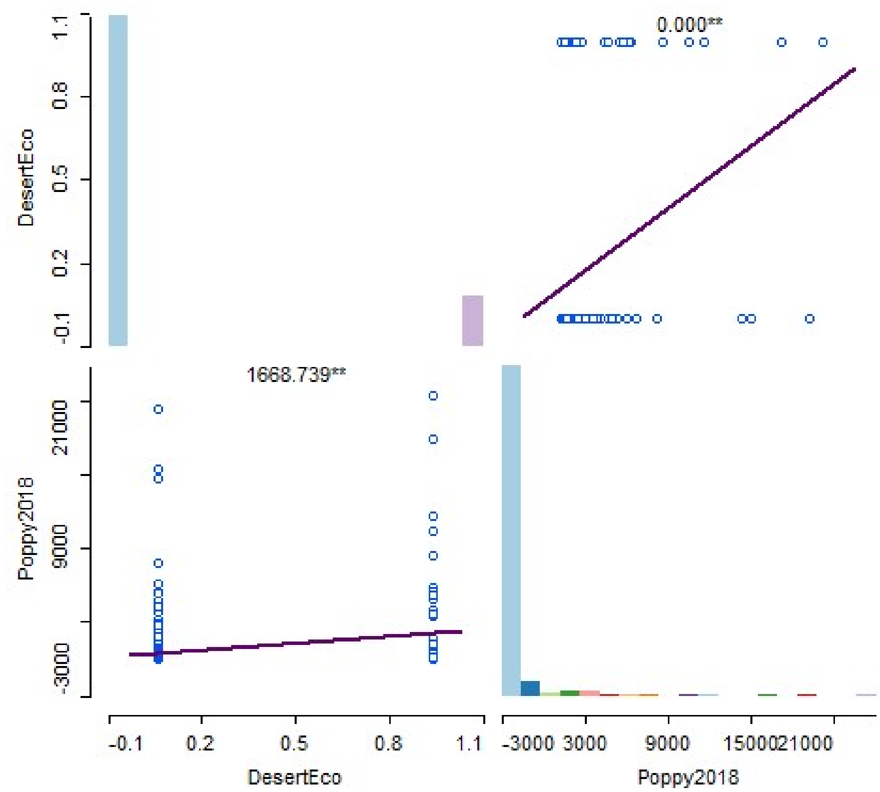
Significant negative correlation



[32] In addition, the correlation of the difference in precipitation in 2018 and 2017 (as a proxy of the drought that same districts experienced in 2018) versus opium poppy areas in 2018 was tested and it was not significant (not shown here).

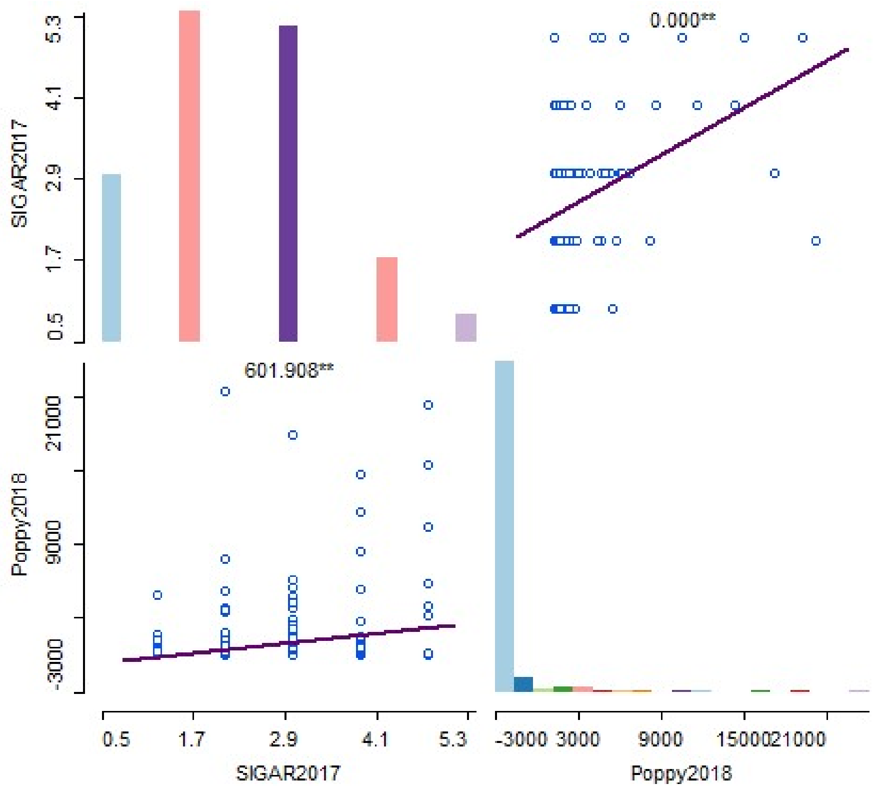
Scatter plot matrix of desert as main type of ecoregion in the district and opium poppy areas in hectares in 2018

Significant positive correlation



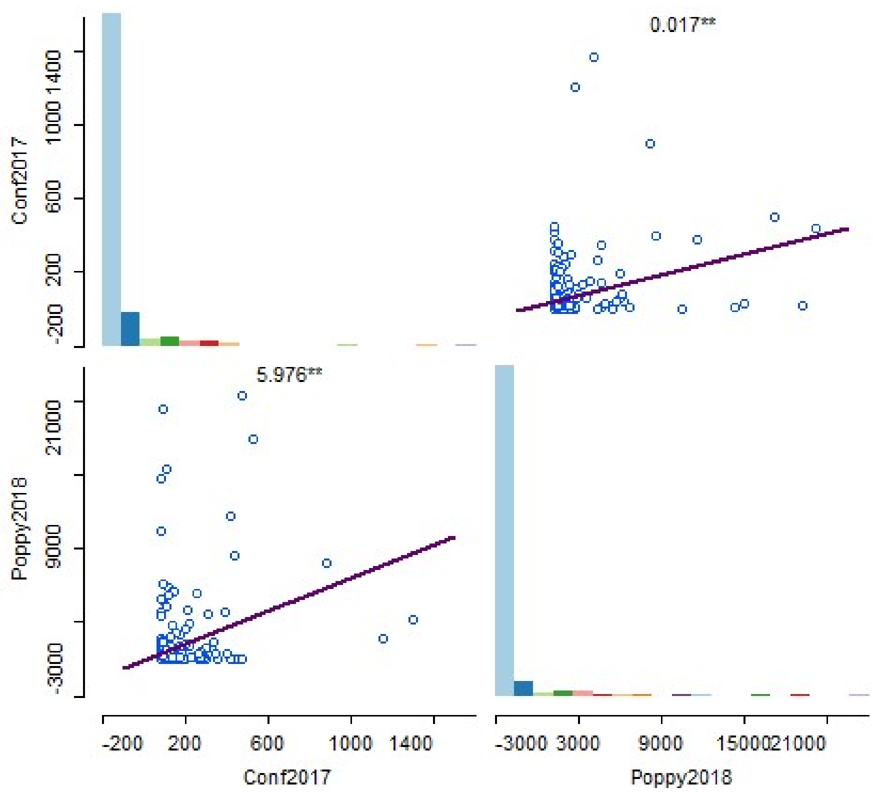
Scatter plot matrix of degree of anti-Government control (scale from 1=none to 5=very high) of the district in 2017 and opium poppy areas in hectares in 2018

Significant positive correlation



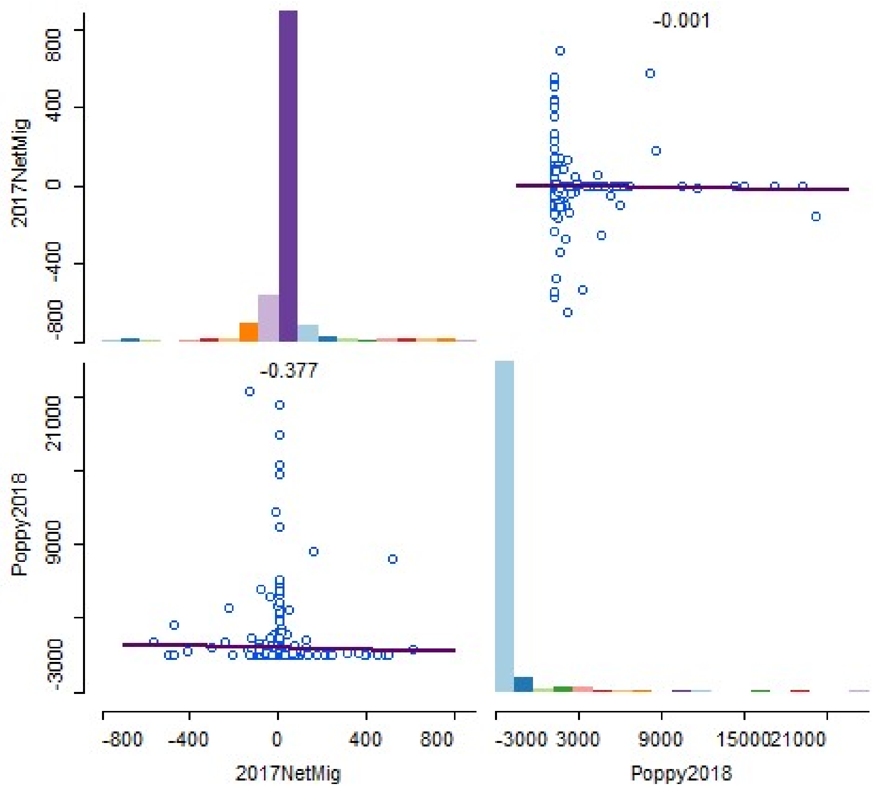
Scatter plot matrix of total number of fatalities from incidents of lethal violence in the district in 2017 and opium poppy areas in hectares in 2018

Significant positive correlation



Scatter plot matrix of net number of internally displaced families from the district in 2017 and opium poppy areas in hectares in 2018[33]

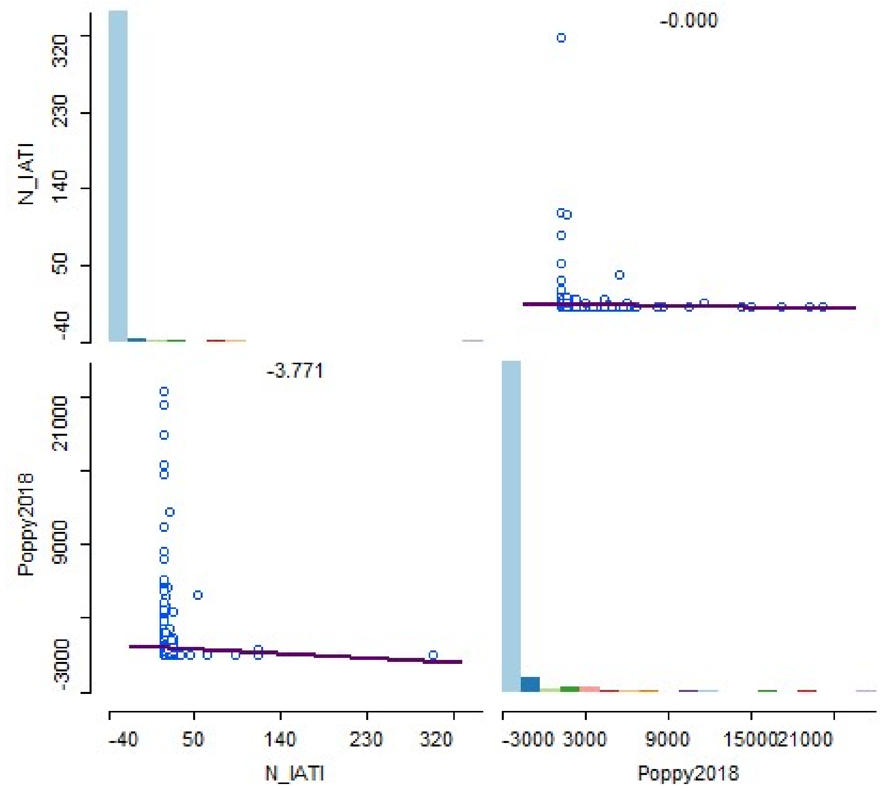
No significant correlation



[33] The correlations between number of families that emigrated and immigrated to the district and opium poppy cultivation in 2018 were not significant (not shown here).

Scatter plot matrix of total number of on-going projects inside the district[34] and opium poppy areas in hectares in 2018

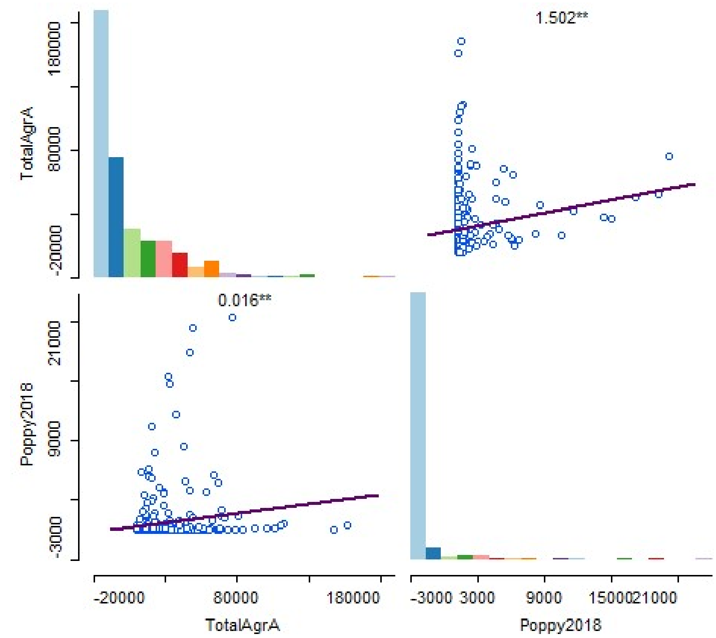
No significant correlation



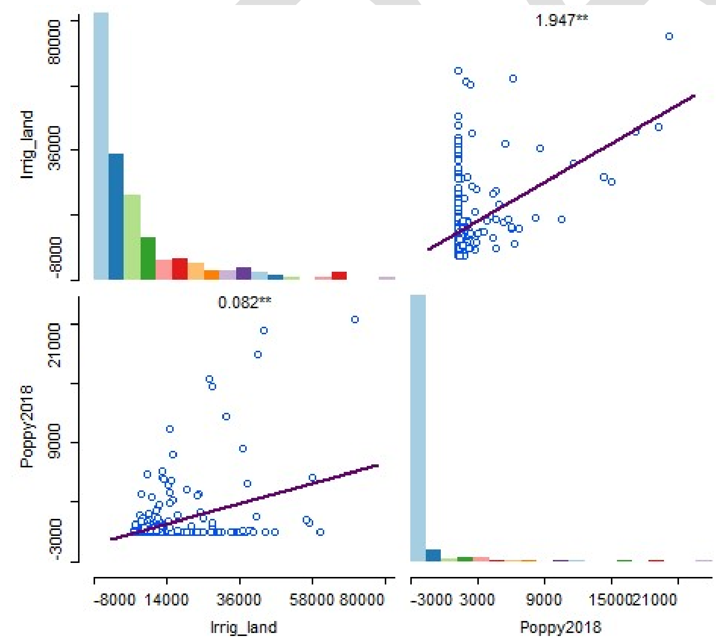
[34] Projects that started before 2018 and were still on-going in 2018

Scatter plot matrix of agricultural areas in hectares[35] and opium poppy areas in hectares in 2018

Total agricultural areas (significant positive correlation)



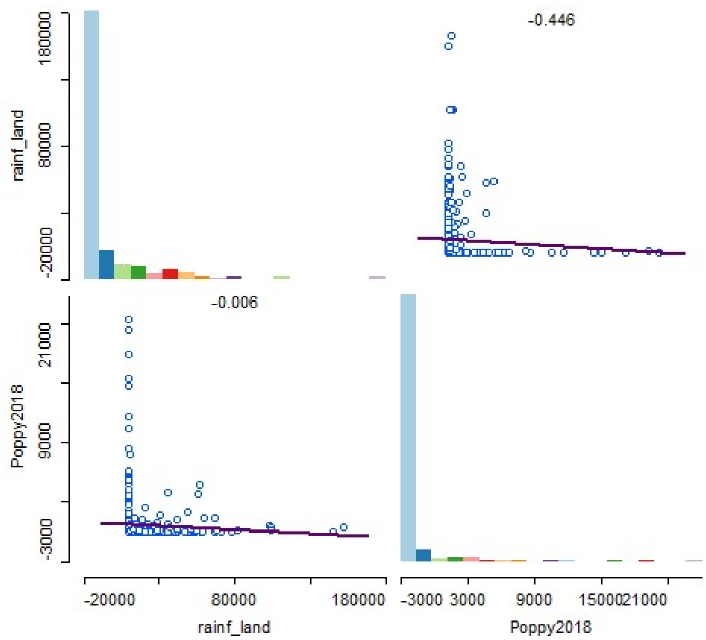
Irrigated agricultural areas (significant positive correlation)



[35] Total agricultural areas (irrigated and rain-fed) and barren land as measured by FAO in 2010 (more recent information not available)

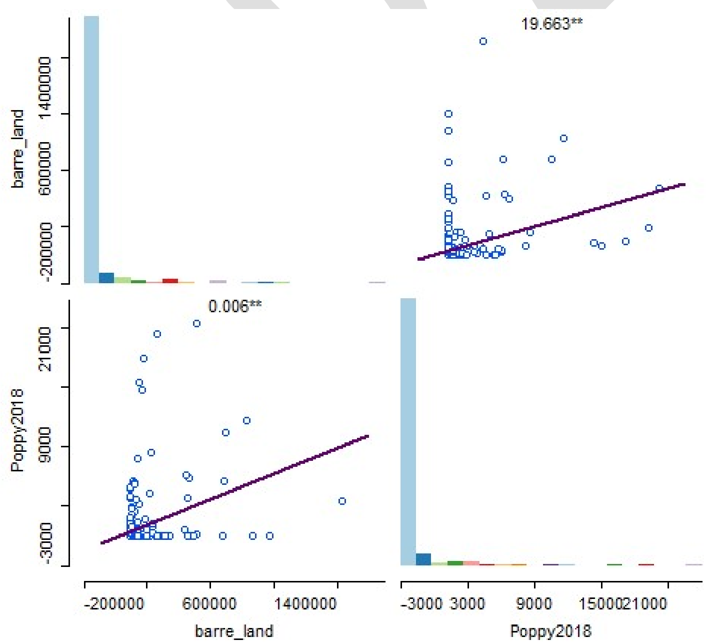
Rain-fed agricultural areas

No correlation



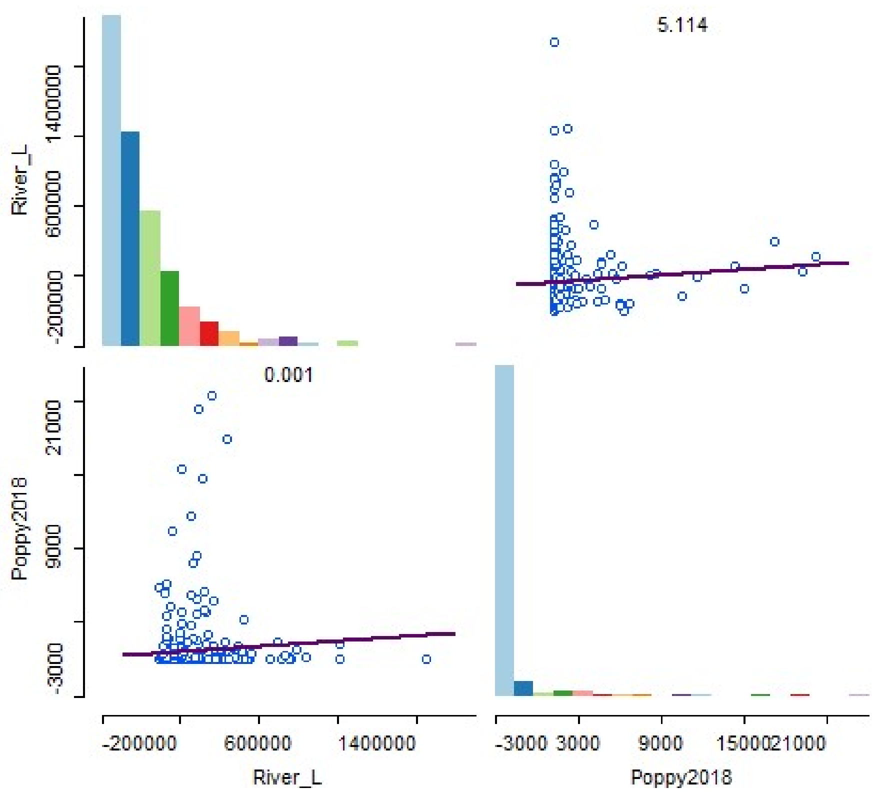
Barren land

Significant positive correlation



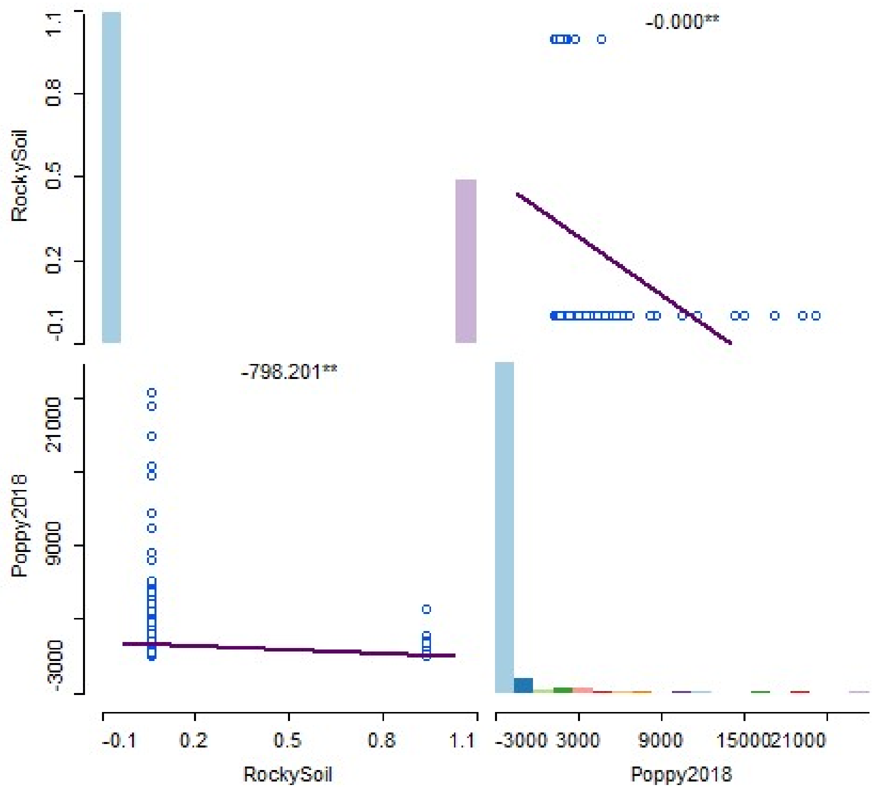
Scatter plot matrix of total longitude of rivers (in miles) inside the district and opium poppy areas in hectares in 2018

No correlation



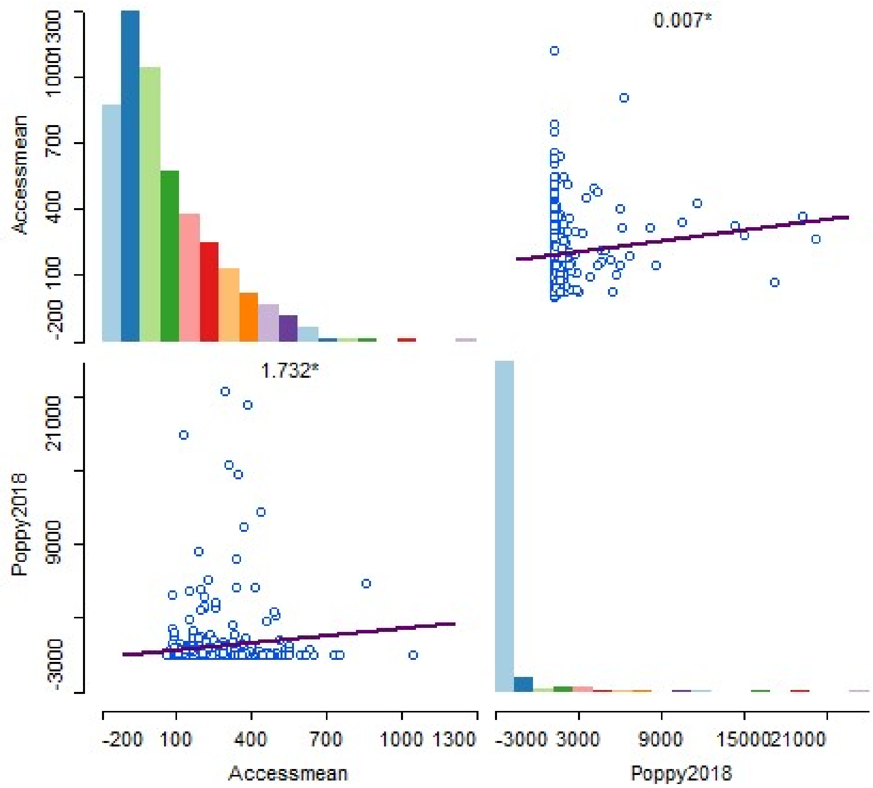
Scatter plot matrix of rocky soil as main type of soil inside the district and opium poppy areas in hectares in 2018

Significant negative correlation



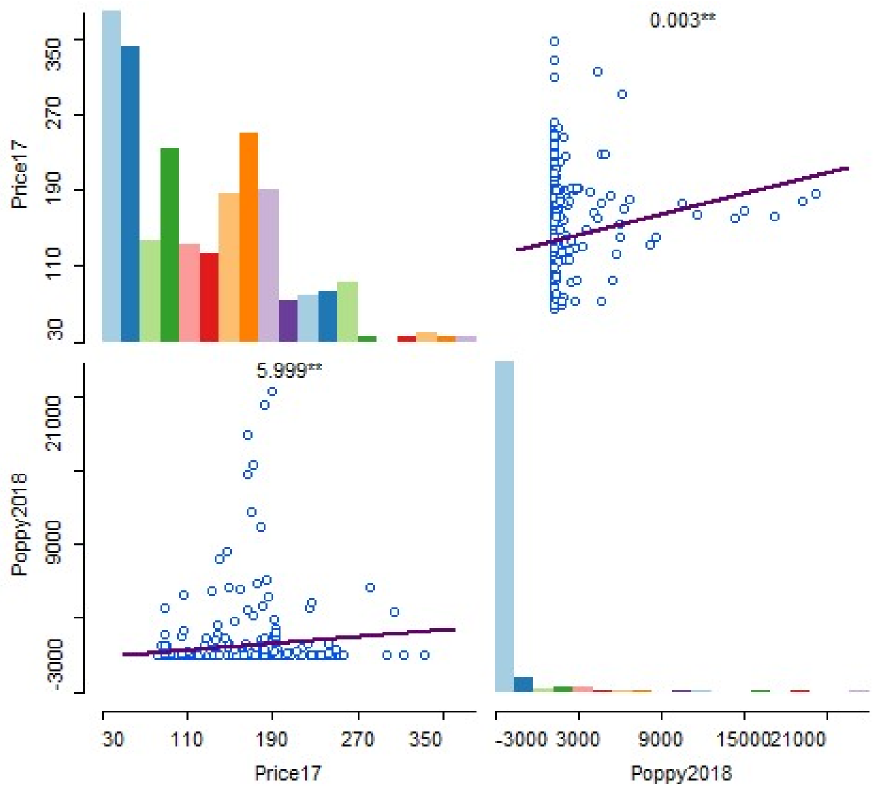
Scatter plot matrix of mean accessibility to cities (in minutes of travel time) and opium poppy areas in hectares in 2018

Significant positive correlation



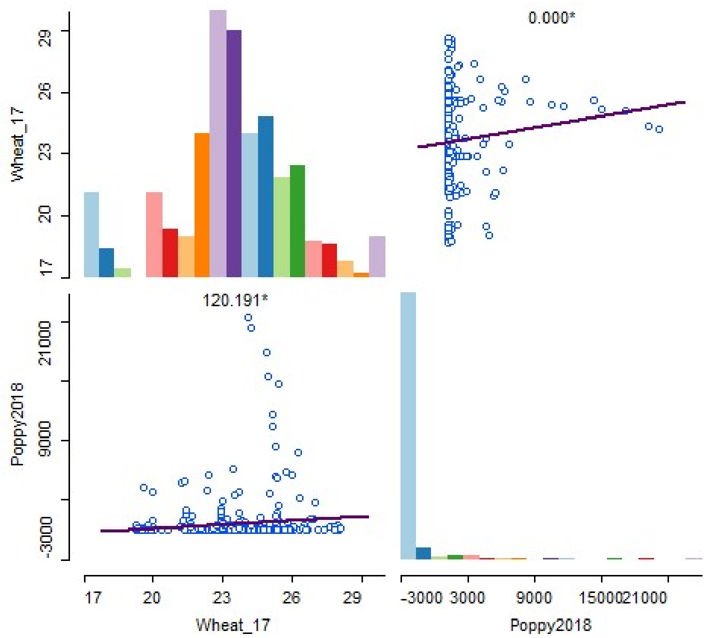
Scatter plot matrix of farm-gate prices of dried opium poppy (Kg/USD) in 2017 and opium poppy areas in hectares in 2018

Significant positive correlation

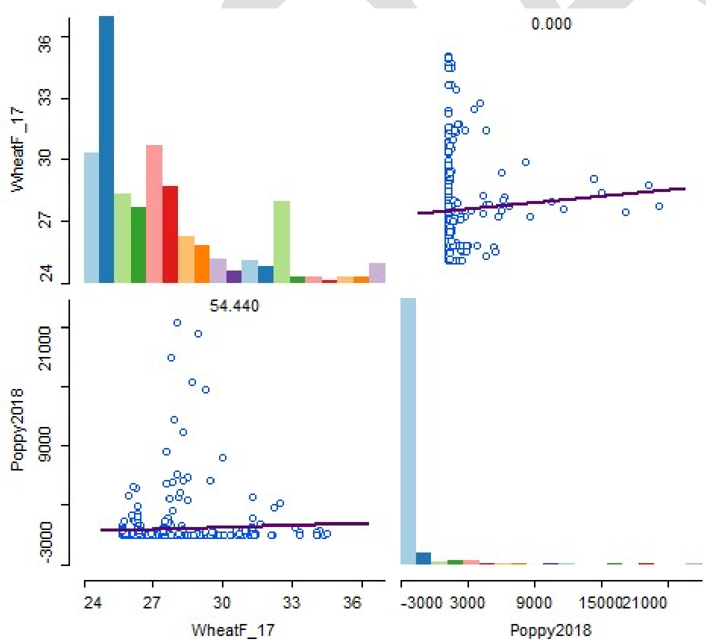


Scatter plot matrix of market prices of legal crops (in Afghani) in 2017 and opium poppy areas in hectares in 2018

Wheat (Significant positive correlation)

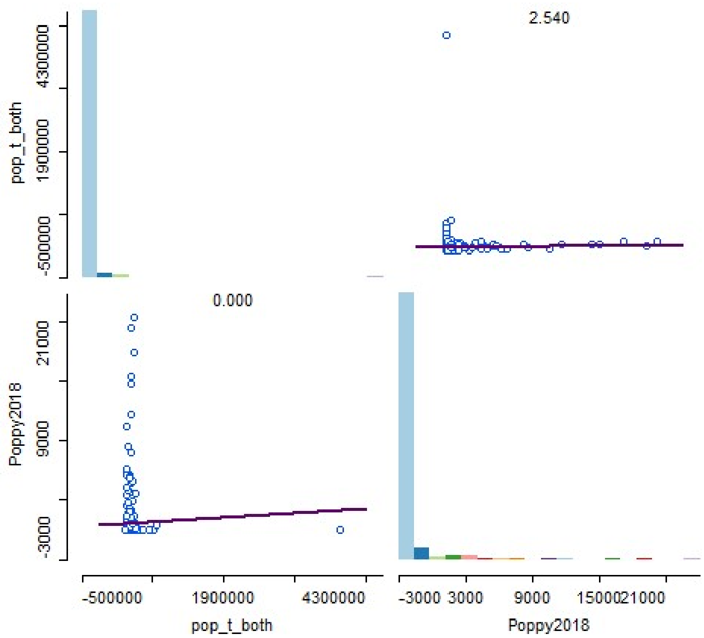


Wheat flour (No significant correlation)

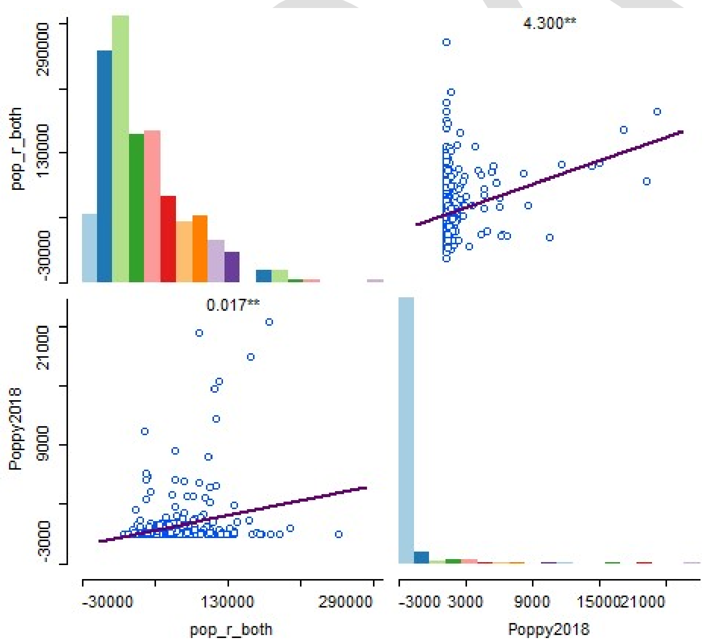


Scatter plot matrix of total population and rural population per district and opium poppy areas in hectares in 2018

Total population (no significant correlation)

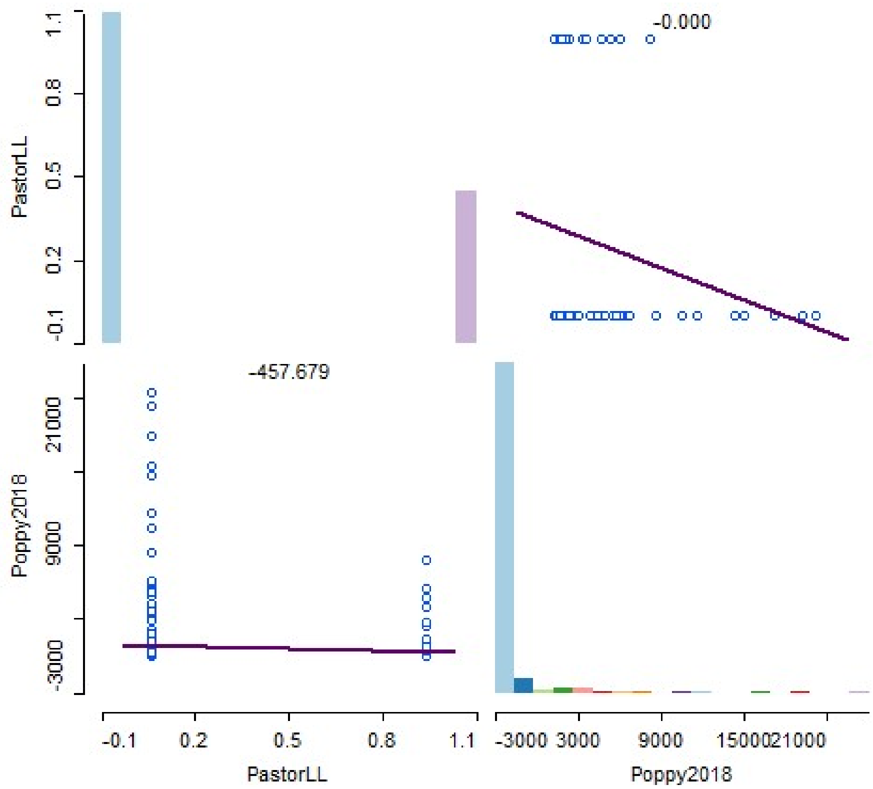


Rural population (significant positive correlation)



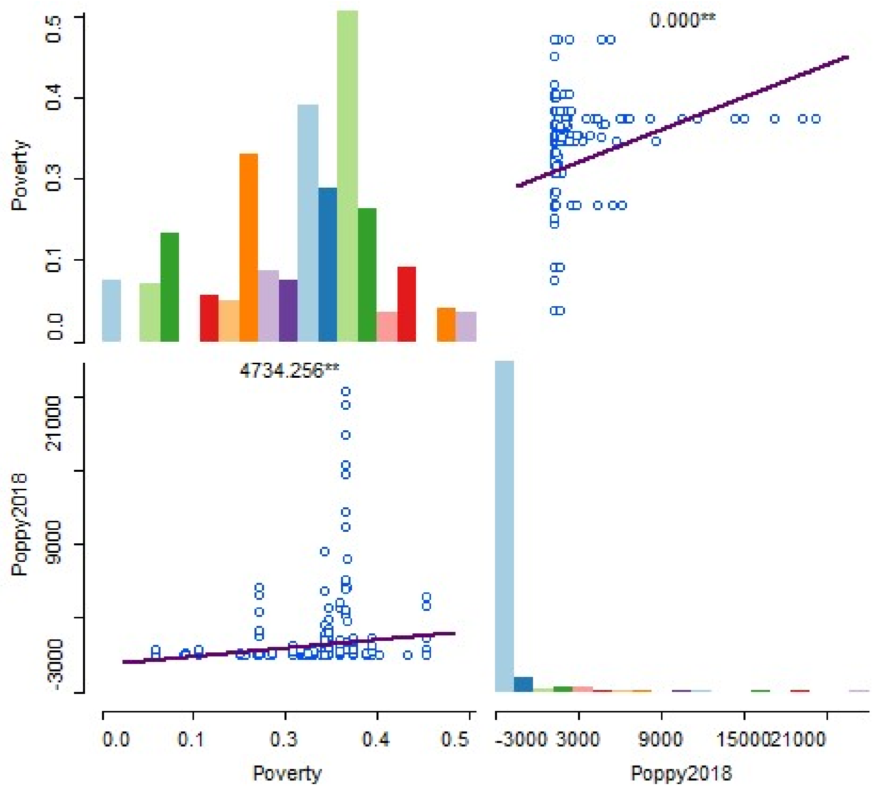
Scatter plot matrix of agro-pastoralism as main type of livelihood inside the district and opium poppy areas in hectares in 2018

No significant correlation



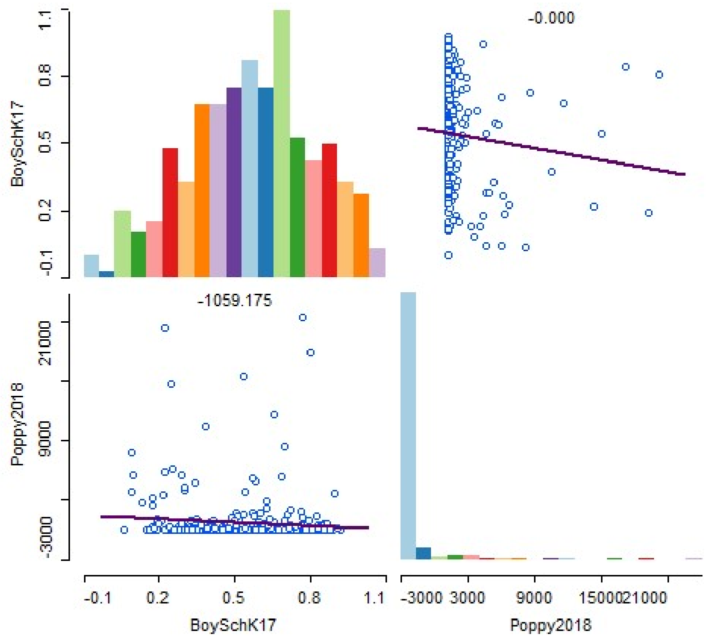
Scatter plot matrix of index of multi-dimensional poverty in 2017 and opium poppy areas in hectares in 2018

Significant positive correlation

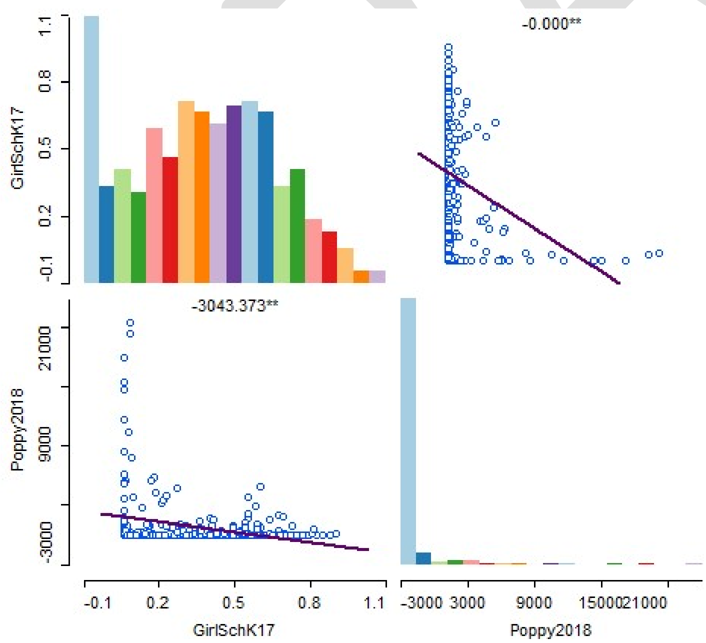


Scatter plot matrix of percentage of villages in 2017 and opium poppy areas in hectares in 2018 for the following characteristics:

Boy school (no significant correlation

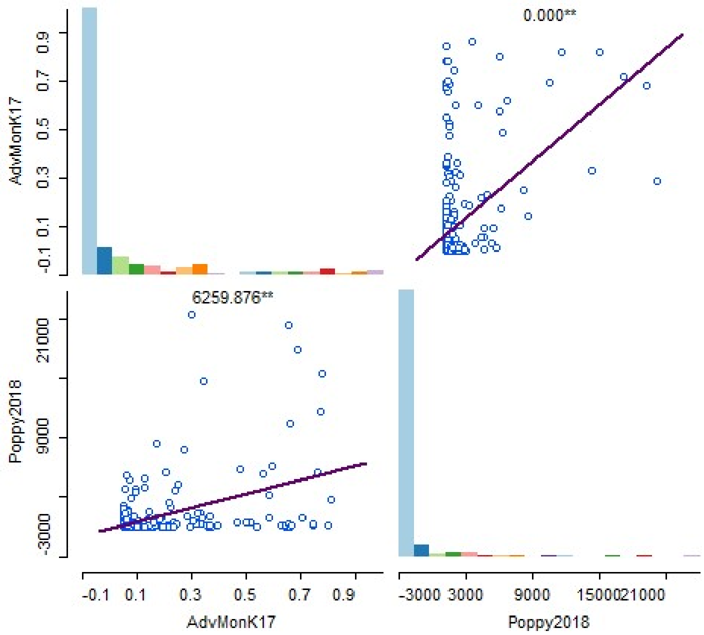


Girl school (significant negative correlation



Advanced money for opium poppy cultivation

Significant positive correlation



Awareness campaigns against opium poppy cultivation

Significant negative correlation