Fall 2016

Modeling Syrian Internally Displaced Person Movements: A Case Study of Conflict, Travel, Accessibility, and Resource Availability

Ethan Harrison
Rappahannock Community College

Modeling Syrian Internally Displaced Person Movements: A Case Study of Conflict, Travel Accessibility, and Resource Availability

Ethan C. Harrison

Rappahannock Community College and The College of William and Mary

Contributors: Sarah Harmon

Modeling Internally Displaced Person Movements

Abstract

This study explores the “push and pull” factors that influence the migration of Internally Displaced Persons (IDPs) in the context of the Syrian crisis. “Internal displacement” occurs when a conflict or natural disaster displaces people from their homes. The distinction between refugees or asylum seekers and IDPs is simple. IDP’s are those who have been displaced but are still in their home country, generally in the same area that the conflict or disaster occurred. Refugees and asylum seekers are displaced individuals who leave their home country, seeking protection elsewhere.

The research for this project was implemented using ArcGIS and R (a statistical software) in the form of models, matrices, and other quantitative based analysis. The data used for this study includes survey data, conflict event data, and temporal data on weather patterns and topography to calculate access to roads and rivers. The findings indicated higher levels of spatial correlation than initially expected between displaced populations and conflict intensity (defined as the number of conflict points within a given area). Additionally, the findings of this project were concurrent with other studies findings that displaced populations generally migrate to areas of high infrastructure (see Csala, 2016 and Alix-Garcia, et.al., 2013).

Modeling Internally Displaced Person Movements

Introduction

In 2015, worldwide displacement reached the highest level ever recorded in history with over 40,000 people becoming refugees or internally displaced daily. According to the Center for Global Development ("Refugees, Displacement and Development: What Should the World Do?", 2016), the average time an individual remains displaced today is between ten and twenty years, and almost all displaced persons rely on the benefits of developed countries for their sources of income. While the living conditions of refugees and IDPs vary and are specific to context-specific social, economic, political, and attitudinal factors, refugees and IDPs have been consistently categorized as the poorest populations in need of humanitarian assistance (De Bruijm, 2009). The problem of human displacement is rapidly growing in scale and in permanence, requiring the immediate attention of the global aid community.

Since 2011, the single-greatest driver of human displacement has been the Syrian civil war and its ongoing related conflicts (UNHCR News, 2016). Traditionally, ethnic conflict exacerbates internal displacement of individuals from access to economic and social safety nets and traffics these people towards better economic prospects and security elsewhere (Akee et al., 2010). Recent models exploring patterns of displacement and IDP/refugee movements rely on the official reports from UN organizations, and official figures provided by the Syrian government that disburse aid locally. While this study attempts to answer many of the same questions, our analytical approach to study how refugees move differs from the majority of studies that have been completed on the migratory movements of IDPs thus far.

From its inception, this project aimed to discover the greatest influencers (or "push and pull" factors) of human displacement, in an effort to build toward a model that can accurately predict population pools of displaced persons. By researching individual and independent factors that are likely to drive displacement such as conflict or infrastructure, we are able to discern the most likely pools and destinations of displaced persons. Thus, rather than focusing on static reports or hearsay, our research takes a more quantitative approach, and thus, allows for a predictive analysis of migratory flows.

Literature Review

In the four-and-a-half years since the start of the Syrian civil war, an estimated 200,000 people have been killed due to armed conflict and violence (Yourish, Lai, Watkins, 2015). However, the vast majority of deaths emerge from indirect consequences rather than direct violence. In 2015 alone, 600,000 Syrian displaced persons died from other preventable causes following displacement (Yourish et al., 2015). According to The UN Office for The Coordination of Humanitarian Affairs (UNOCHA) Financial Tracking System, as well as the multilateral contributions made to the crisis, over 5 billion USD in foreign assistance was allocated for displaced persons in Syria through 2015. Only 4 billion of those funds have reached IDPs (Burns, 2016). From this, we conclude that many shortcomings in assistance for Syrian IDPs are tied to aid delivery rather than resourcing, demanding new insights to improve targeting and efficiency of relief and assistance.

Although a large contributor to the shortfall in aid delivery is likely a result of the loss of resources allocated through the state due to corruption and bureaucracy, a persisting knowledge gap is also at play. Organizations granted access to deliver aid locally are at a significant disadvantage, not knowing where resources are most needed due to the difficulties associated with tracking moving people. Historically, organizations have relied on information pieced together from media reports, surveys, and actors on the ground. This information is unreliable at best, and damaging at worst. A frequent recruitment tactic used by ISIS in the Syrian Conflict has been to seize aid convoys, using the supplies as leverage for recruitment and as a stipend for the families of insurgents (Abi-Habib, 2014 and Mansfield, 2016). The complexities of these conflicts significantly limits the ability of humanitarian actors to reach refugee and displaced populations and deliver aid effectively. Understanding how these groups move, and the ideology behind these movements, is imperative for humanitarian organizations and policy makers to meet the needs of these groups, and to avoid the violence, poverty, and human rights violations that so often emerge in these contexts.

Previously, the flight patterns of displaced persons were studied by compiling records on individual stories through the use of social media reports or through mass movement groupings derived through computational analysis (Groen, 2016). In a report published in 2016 on Syrians displaced in Lebanon, the authors modeled the Syrian crisis using both of these methods. The findings of this study indicated refugees prioritize safety, and concentrations are highly correlated with areas of high population values (Csala, 2016). This result is supported by the widely used migration models of Gravity, Lee, and Spatial-Interaction stating that migration between origin and destination should be directly proportional to the product of population at origin and destination, and inversely proportional to the distance between them (Zipf, 1946). Additionally, the migration and interaction between two places is dependent upon the increased importance of one or both of the locations to the other (Zipf, 1946). A study published in 2012 conceptualized an ideal data model for migration in the form of an "Entity-Relationship diagram" (Hristoski & K Sotiroski, 2012). With a methodology quite similar to the one used here, the authors enumerated a number of factors that influence migration and integrated them in the proposed entity-relationship diagram. While in theory the E-R diagram has potential, finding a dataset with the granularity that the authors required is no small task, and when it comes to conflict related-migration, nigh impossible. To summarize, while the complexities of internal displacement are becoming highly policy-relevant, surprisingly little quantitative research is available on the subject. Despite the potential of previous research, early efforts to model internal displacement have come with mixed success. As we enter the sixth year of the Syrian conflict, a conflict the United Nations termed "the biggest humanitarian crisis of our time" (UNHCR, 2016), internal displacement, and the science behind it, is increasingly relevant. .

Data and Methods

The stock data components (e.g. those acquired from a third party) used for this project are referenced in the table below. They include the Uppsala Conflict Data Program's Conflict Events data, UNOCHA's spatial boundaries on Syria at the "adm3" or sub-district level, and The Multi-Sectoral Integrated Needs Assessment (MSNA) survey data on the sub-districts of Syria for 2014.

Data Source	**Data Details**
Multi-Sectoral Needs Assessment (UNOCHA)	Completed beginning of June-October 2014 by UNOCHA agencies and operators in 156 Syrian sub-districts. The assessment was in the form of a census-like survey in which 150+ interviewers were trained and deployed within assigned subdistricts. The result was a tabular dataset containing data on medical infrastructure, housing, humanitarian assistance, water and other infrastructural resources, and most importantly, accurate IDP population information.
Syrian Satellite Data (UNOCHA)	Includes Syrian shapefiles with spatial boundaries for the country's subdistricts. Spatially-referenced covariates were

	extracted by using these boundaries via computer resources at the College of William and Mary. These covariates included weather and topographical data, as well as infrastructure-related data such as population statistics and the concentration of night-time lights in a given area.
Uppsala Conflict Data Program (UCDP)	Includes over 9,000 geocoded conflict events for Syria 2011-2015. The UCDP ordinarily releases event-based and georeferenced datasets on organized violence that are disaggregated spatially and temporally to the level of individual incidents of fatal violence. Each conflict event normally comes with the date of the event, location with coordinates, actors participating in the event, estimates of fatalities, as well as variable that denote the certainty with which these data are known.

Due to the number of independent data sources used in this study, a significant portion of this project involved cleaning and standardizing these datasets. The satellite covariates obtained from William and Mary were merged with the MSNA data to produce a final dataset appropriate to the original research question. However, given the high number of variables this dataset contained, for early analysis the dataset was restricted to the thirty key variables believed to be the biggest influencers of IDP movements based on previous desk research. Similarly, after web-scraping and geo-referencing the UCDP conflict data, the resulting tabular data-set contained nine-thousand conflict events with detailed information on actors, type of violence, and dates for further analysis. This data was then imported into a GIS software where each individual conflict point was projected over UNOCHA's Syrian boundaries. Research had indicated a strong relationship between conflict (and its respective intensity), and displacement that was highly logical from a qualitative perspective. With this scenario in mind, we developed a metric to accurately quantify conflict intensity defined as the number of conflict points within a given sub-district. Additionally, a Pearson correlation matrix was constructed from the MSNA-Covariate dataset, indicating relative strength or correlation between every variable. The correlation matrix provided a clearer picture of potential relationships that were viable from a modeling standpoint.

The relationships drawn from the correlation matrix were further explored in the form of linear regressions. Using IDP population statistics from the MSNA data, the percentage of IDP's in each sub-district was calculated. This metric was than used as the constant or independent variable in each linear regression. The reliability of these models were judged based on their adjusted r-squared value. The r-squared value provides a numeric value to the variance and explanatory power of a linear regression to the response variable. The adjusted r-squared value computes this variance while also taking into account the number of predictors included in the model, avoiding an inflated value that can occur with an increase in the number of predictor variables (See Figure 1 in Appendix A for the full call).

After an iterative modeling process, the final predictive model includes point count (the previously discussed conflict intensity metric), distance to roads, distance to rivers, and night time lights. Additionally, this model was incorporated into a classification and regression tree that assigned probabilities to the most likely location of IDPs in relation to these predictors. This tree also quantifies and provides a visualization of the factors that correlate, and thus influence, IDP populations the most (see figure 2 in Appendix A, page). Finally, this model is used to predict the probable location of an individual IDP throughout Syria.

Results

The early findings from this research confirmed many of the theories suggested in desk research. The results from the linear regressions and models answered the original research question directly. The final linear model included population statistics, distance to roads, distance to rivers, nighttime lights, and point count. Based on the methodology by which the research was conducted, these are the greatest influencers of displaced migration flows. The presence of the variable point count, indicates that conflict is in fact one of the greatest drivers or push factors of displacement (See figure 3 in Appendix A for a better illustration of this relationship). Additionally, the other predictive factors that were identified all share a common theme. Population statistics, the distance to roads and rivers, and the presence of night-time lights all can be viewed as proxies to infrastructure, indicating that the majority of internally displaced persons pool in areas of high infrastructure (See figures 4, 5, and 6 in Appendix A).

Conclusion

The findings from this project resulted in a few key observations. First, that conflict does in fact play a role in the movements of displaced populations, a theory that has been voiced by academics but with little quantitative evidence. A study on land use in Darfur during its civil war indicated similar phenomena to the ones indicated in this research, finding that ethnic conflict results in the majority of displaced persons fleeing conflict and seeking protection in larger urban areas and denser population centers. Additionally, the findings in this study indicate that displaced persons migrate towards areas of higher infrastructure rather than living in rural camps. This contradicts the common misconception that displaced Syrians are hiding by the borders in tents. In 2016, UNOCHA estimated that there were 6.6 million IDPs in Syria, of these, they estimated a mere 108,000 resided in refugee camps (United Nations High Commissioner for Refugees, 2016), leaving an open question of where and how the remaining displaced persons relocate and survive.

While the datasets that were used and created for this research have many potential applications in regards to tracking migration, there are limitations. The lack of data on IDP movements from one distinct point to another, known as “flow data”, limits the current applicability of this research. Despite this fact, the potential applications and future causal analysis based on this research and data are numerous. By focusing on what these groups move toward, rather than how they move, many of the previously-discussed issues related to conflict and displacement can be avoided. Predicting where displacement populations are likely to congregate enables humanitarian actors to meet the people most in need with greater speed and accuracy. Further analysis of the accuracy of aid targeting to displaced populations specific to Syria could be pursued based on locational aid information and the model's predicted locations of IDPs. Finally, if actual flow data ever does become available, it would enable the development of a model that could accurately predict flight routes and destinations based on real time conflict events. This would allow policymakers to tailor their responses to known destinations, and allow humanitarian organizations to avoid the complications that so often plague conflict and disaster situations, the potential of which, can not be underestimated.

References

ABI-HABIB, Maria (2014, November 17). Islamic State Poaches International Aid for Syrians. Retrieved December 09, 2016, from http://www.wsj.com/articles/islamic-state-poaches-international-aid-for-syrians-1416158609

Akee, R. K., Basu, A. K., Chau, N. H., & Khamis, M. (2010). "Chapter 28 Ethnic Fragmentation, Conflict, Displaced Persons and Human Trafficking: An Empirical Analysis. Frontiers of Economics and Globalization Migration and Culture", 691-716. doi:10.1108/s1574-8715(2010)0000008034

Alix-Garcia, Jennifer, Anne Bartlett, and David Saah. (2013). The Landscape of Conflict: IDPs, aid, and land use change in Darfur. Journal of Economic Geography, 13 (4) : 589-617.

Burns, Molly S,. (2016). “Economic Consequences of Conflict, Displacement and Humanitarian Aid: Implications for Syria.” International Development, Community and Environment (IDCE), no. 69 (2016). Retrieved 15 July, 2016 from http://commons.clarku.edu/cgi/viewcontent.cgi?article=1036&context=idce_masters_papers.

Center for global development, . (2016,). “Refugees, Displacement and Development: What Should the World Do”. Retrieved 9 December 2016, from http:www/cgdev/org/blog/refugees-displacement-and-development-what-should-world-do.

Csala, Denes. (2016) “Motives to Flee: Modeling Syrian Refugee Crisis.” Academia, 2016. Accessed July 1, 2016. http://www.academia.edu/21947416/Motives_to_flee_modeling_Syrian_reufgee_crisis

De Bruijm, Bart. (2016). “Human Development Reports”. Retrieved 9 December 2016, from http://hdr.undp.org/en/content/living-conditions-and-well-being-refugees.

Groen, D. (2016). Simulating Refugee Movements: Where would You Go? Procedia Computer Science, 80, 2251-2255. doi:10.1016/j.procs.2016.05.400

Hristoski & K Sotiroski. (2012) Conceptual data modelling for modern human migration. Management Information System 7, 3-12.

Karen Yourish, Rebecca Lai, and Derek Watkins. (September 15, 2015). Death in Syria. [Weblog]. Retrieved 9 December 2016, from http://www.nytimes.com/interactive/2015/09/14/world/middleeast/syria-war-deaths.html.

Katie Mansfield. (2016, 21 January 2016). Has taxpayer funded British aid meant for Syria ended up in hands of ISIS militants?. [Weblog]. Retrieved 9 December 2016, from http://www.express.co.uk/news/world/636723/taxpayer-funded-British-aid-stolen-Syria-ISIS

United Nations High Commissioner for Refugees. (2016, no-date). Syria conflict at 5 years: the biggest refugee and displacement crisis of our time demands a huge surge in solidarity. [Weblog]. Retrieved 9 December 2016, from http://www.unhcr.org/en-us/news/press/2016/3/56e6e3249/syria-conflict-5-years-biggest-refugee-displacement-crisis-time-demands.html

United Nations High Commissioner for Refugees. (2016, no-date). Worldwide displacement hits all-time high as war and persecution increase. [Weblog]. Retrieved 9 December 2016, from http://www.unhcr.org/en-us/news/latest/2015/6/558193896/worldwide-displacement-hits-all-time-high-war-persecution-increase.html.

United Nations High Commissioner for Refugees. (2016, no-date). Syria conflict at 5 years: the biggest refugee and displacement crisis of our time demands a huge surge in solidarity. [Weblog]. Retrieved 9 December 2016, from http://www.unhcr.org/en-us/news/press/2016/3/56e6e3249/syria-conflict-5-years-biggest-refugee-displacement-crisis-time-demands.html

UNSC. (2016) Implementation of Security Council Resolutions. Report. United Nations Security Council, 2016. 1-17.

Zipf, George Kingsley. (1946). THE P1 P2D HYPOTHESIS: ON THE INTERCITY MOVEMENT OF PERSONS. (1946). American Sociological Review, 11(6), 677-686.

Data Sets

Uppsala Conflict Data

Program (Date of retrieval: 16/09/15) UCDP Conflict Encyclopedia:

www.ucdp.uu.se, Uppsala University

United Nations Office

for the Coordination of Humanitarian Affairs. (2016). Syria Multi-Sectoral

Needs Assessment (MSNA) - Final report [TABULAR]:

https://www.humanitarianresponse.info

United Nations Cartographic Section (UNCS) and United

Nations Office for the Coordination of Humanitarian Affairs (UNOCHA). (2016).

Syrian Arab Republic - Administrative Boundaries, Populated Places:

https://data.humdata.org/dataset/syrian-arab-republic-administrative-boundaries-populated-places

Appendix A (Figures, Maps, and Graphs)

Figure 1.0

Final Linear Model

```
Call:
lm(formula = idp.percent ~ PNTCNT + Dist.to.Roads + Dist.to.Rivers +
    Night.Time.Lights, data = dta.conflict)

Residuals:
      Min        1Q    Median        3Q       Max
-0.010743 -0.005822 -0.003779  0.000087  0.075879

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        2.958e-03  1.767e-03   1.674   0.0969 .
PNTCNT            -3.754e-07  6.422e-06  -0.058   0.9535
Dist.to.Roads      1.811e-06  1.810e-07  10.009  < 2e-16 ***
Dist.to.Rivers    -1.096e-06  1.231e-07  -8.901 9.37e-15 ***
Night.Time.Lights  1.981e-04  2.067e-04   0.959   0.3398
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01257 on 115 degrees of freedom
  (164 observations deleted due to missingness)
Multiple R-squared:  0.5113,  Adjusted R-squared:  0.4943
F-statistic: 30.08 on 4 and 115 DF,  p-value: < 2.2e-16
```

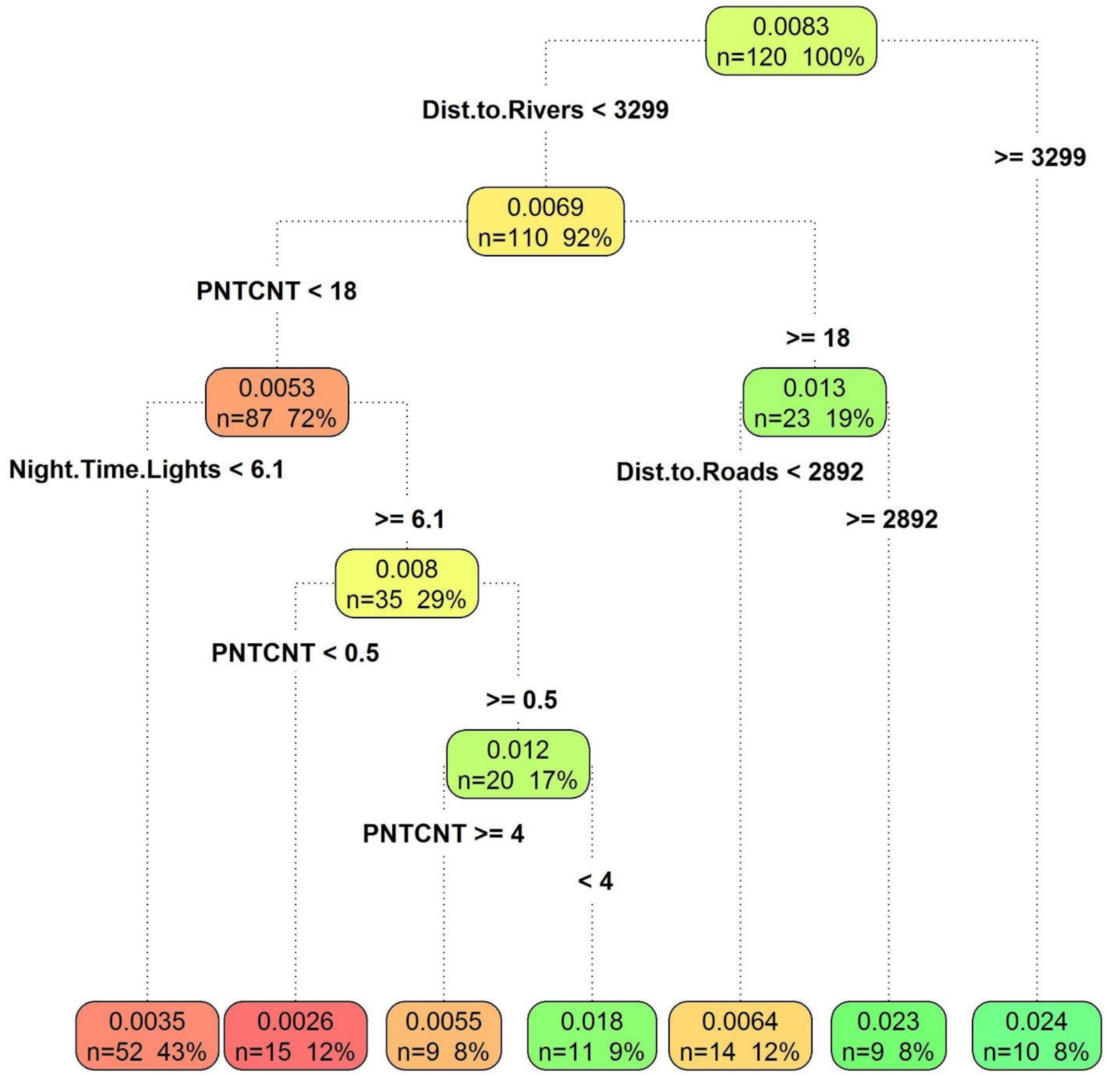


Figure 2.0: Machine Learning Algorithm plotted as a decision tree. The percentage in colored boxes is the percent of the total number of IDPs in Syria where N equals one observation which equals a sub district in Syria. The decision making tree can be viewed by reading from top to bottom, IE in the first divergence, 92% of all Syrian IDPs are within X Kilometers of a river. Continuing down the left side of the tree, 72% of all IDPs reside in an area with a point count of less than 18 (or less than 18 conflict events).

Conflict and Displacement Density

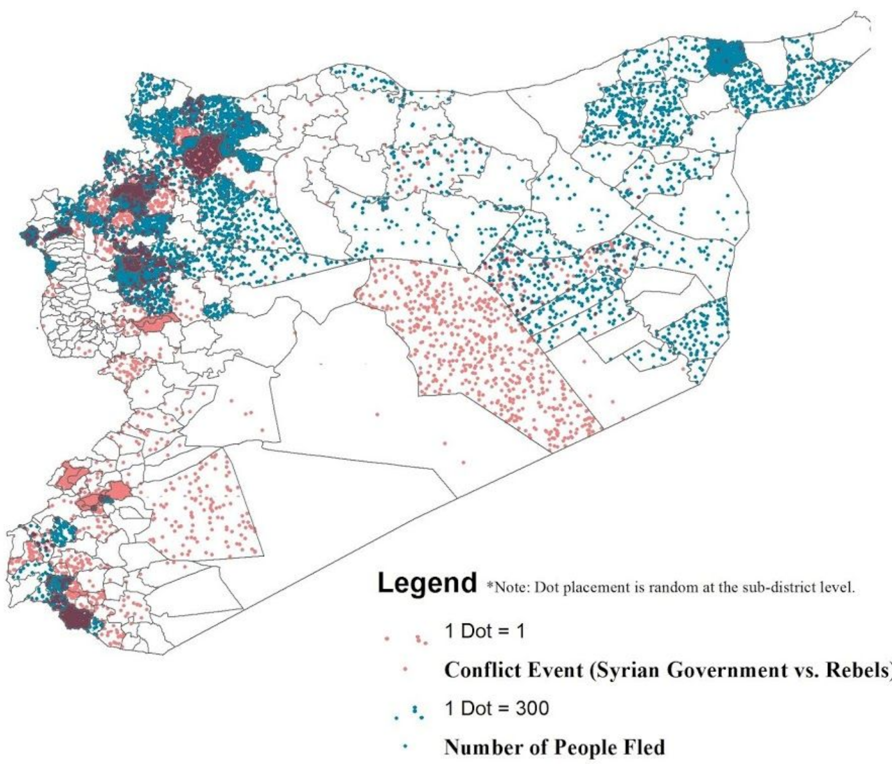


Figure 3.0

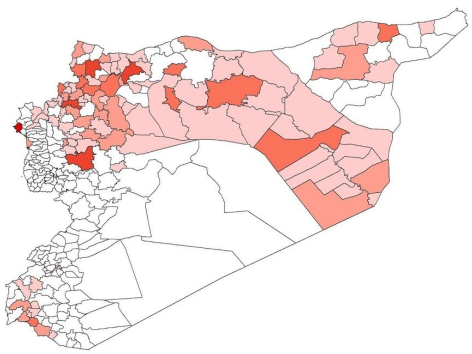
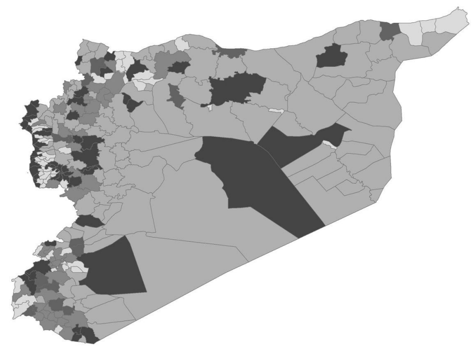
Historical IDP Pools Across Syrian Sub-districts	***Predicted IDP Pools Across Syrian Sub-districts***
	 *Figure 4.0*

Figure 5.0: Heat map of the density of Night-Time Lights in Syria. This variable can be viewed as a strong indicator of infrastructure

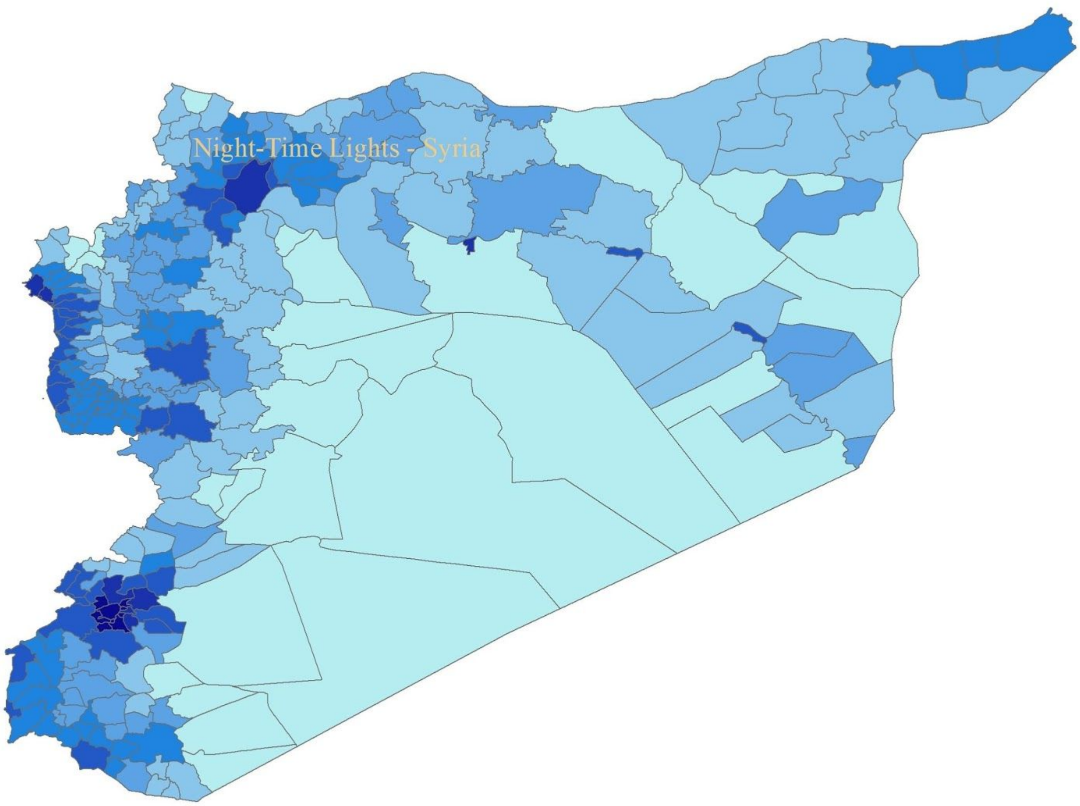
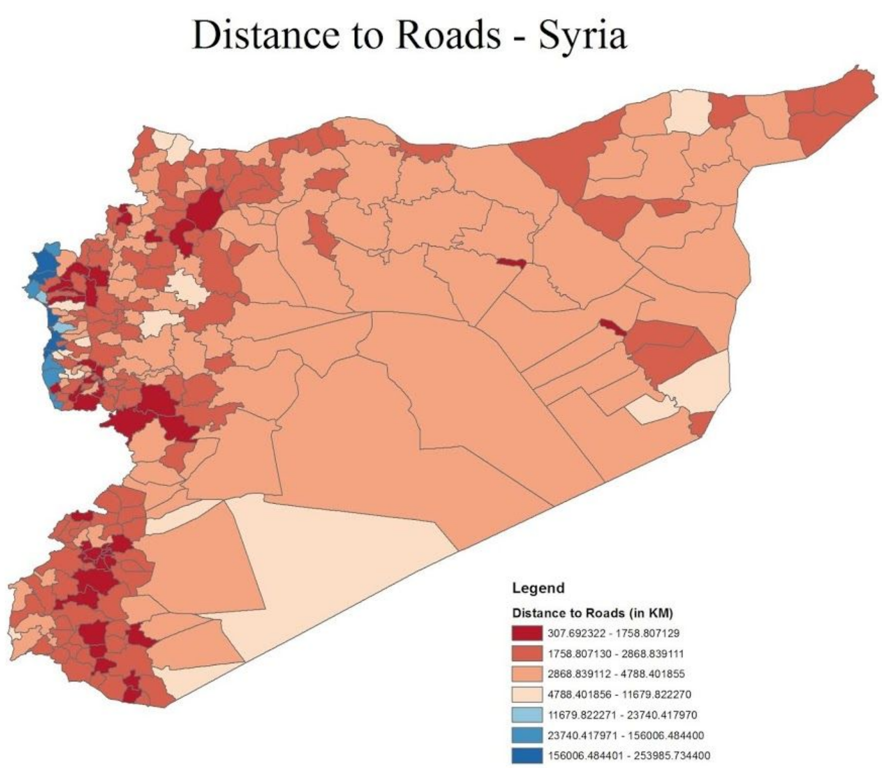


Figure 6.0


Distance to Roads - Syria
Legend
Distance to Roads (in KM)
307.692322 - 1758.807129
1758.807130 - 2868.839111
2868.839112 - 4788.401855
4788.401856 - 11679.822270
11679.822271 - 23740.417970
23740.417971 - 156006.484400
156006.484401 - 253985.734400