Geospatial analysis of displacement in Afghanistan, a case study

Anais Dahmani-Scuitti, Walker Kosmidou-Bradley, Johanna Lee Belanger, Erwin Knippenberg
*(*Poverty & Equity Global Practice, The World Bank)

Contents

1. Introduction

1.1. Executive Summary

A geospatial analysis of displacement in Afghanistan, using Nighttime Light data

The objective of this work is to demonstrate whether geospatial data can be used to proxy for important displacement flows in Afghanistan, where survey data is hard to collect. Understanding how to better proxy for displacement flows in Afghanistan is very timely, as the recent spike in displacement flows following the COVID-19 crisis is expected to be exacerbated by a developing drought. The International Organization for Migration (IOM) estimates that more than 12,800,000 individuals dependent upon agriculture and livestock could face food insecurity or displacement in the 25 priority provinces most likely to suffer from drought in 2021.[1] Drawing lessons from the 2018 drought is a necessity to augment our ability to better identify and predict displacement flows when data are scarce.

The project tests whether inflows of displaced populations into existing settlements can be detected using freely available remote sensing data, such as nighttime lights. This work was designed as a proof of concept and focuses on an area with a well-known and documented shock, the 2018 drought, that drove important displacement patterns. An estimated 120,000 people were displaced from rural areas of Badghis Province and congregated to the provincial capital, Qala-e Naw, to receive emergency humanitarian assistance; Along with Herat Province, Badghis experienced the largest displacement flows in the country in 2018.[2] To identify these patterns, the team combines displacement data from the IOM Displacement Tracking Matrix (DTM) with Nighttime Lights (NTL) data obtained from NOAA's VIIRS satellite.[3] While this exploratory study was regionally focused to the shock-driven displacement in Badghis, all data used are available country-wide and the methodology could be expanded. Further discussion of the data and methodology used can be found in Annex B.

Linear regressions show that cumulated displacement movement over 2018-2020 can be proxied by trends in NTL imagery. Settlements with higher net inflows of displaced persons between 2018 and 2020 have larger NTL growth than others. The findings hold when accounting for population in 2017 as a proxy for initial urbanization, and when looking at the net inflow of persons as a share of the settlement population, or when modifying the catchment areas around settlements GPS points from 1km to 2km and 5km radius. A Jenks decomposition of settlement based on their initial NTL levels in January 2018 (high or low) hints that findings differ across the two categories. Larger net inflows always correlated with larger NTL in settlements with initially large NTL levels, while the correlation is more complex in more rural settlement, as the positive correlation would not hold in extremely rural areas with very small past NTL levels.

While quadratic analyses yield more complex results, the regression suggest a concave relationship between NTL and displacement inflows, and a possibly concave growth of NTL with regards to previous month levels, provided that displacement flows are large enough (sufficient conditions). In other words, settlements that show the largest level of expansion are those with the lowest level of displacement

[1] https://mcusercontent.com/227c5b501abac2d6784c32232/files/7d592f4b-ea32-40bf-bc2a-5a9cb7bb4cfd/PotentialDroughtAreas_23FEB2021.pdf

inflows to start with. In addition, a sufficient condition for the NTL growth to be concave in lagged NTL is that Total Inflow be larger than 74,124 individuals.

In the quadratic analyses, there exist several cases whereby a marginal increase in the displacement inflow would raise the NTL growth, in settlements that are both net receivers (inbound) and senders (outbound). For example, a marginal increase in the net inflow of displaced person would raise the NTL levels in inbound settlements with already medium to high level of human activity (NTL) when the total inflow of displaced is not very large. However, it will always drop the NTL levels in settlements that serve as important inbound with a *Total Inflow larger* than 77,255 individuals over the 2018-2020 period. In settlements that serve as important outbound areas, a marginal reduction in the number of people who flee the settlement (i.e., increase in net inflow) leads to an increase in the NTL levels regardless of the past NTL levels. In settlements that serve as small outbound areas (Total Inflow negative and small in absolute value), a marginal decrease in outflows would increase the NTL if past NTL values are larger than 0.28.

Reverse engineering these results, the model uses NTL data to predict whether a settlement was a net receiver of displacement flows in 2018-2020 and found promising results. The paper uses a simple fixed effect model to predict total inflow of displaced persons and estimate whether a settlement has a positive or negative net inflow, i.e., whether it was an inbound area or served as outbound that people flee. We find that there is a significant and concave relationship between average NTL and the cumulated net inflow of displaced population over 2018-2020. The reverse model correctly classifies 63.2 percent of these settlements, a promising baseline of accuracy that could be improved upon with more sophisticated algorithms. We looked at how the NTL model performed compared to a similar analysis using random variable as main regression (over 500 iterations), and found that NTL is both more significant at estimating cumulative net inflow of displaced, and tend to better predict whether the settlement is a net receiver or sender. Over the 500 iterations, the average probability of accurately categorizing a settlement in a random model is at 59.4%, almost 2 standard-deviation lower than the 63.2% obtained using NTL. While this study focuses on the Badghis region where displacement flows have been particularly important, external validity needs to be tested, and these results make the case for a scale-up of such methodology to the whole Afghanistan. Findings are nevertheless promising, and additional research will be required to better understanding the factors intertwined with displacement flows NTL growth, e.g., such as the nature of economic activity.

This study provides a first basis to better understand the drivers of settlement growth and see whether population displacements can be proxied by variations in geospatial data, this work ought to be expanded to other areas. This paper contributes to the literature on geospatial data and displacement by combining remote sensing data and IOM administrative data in Afghanistan, a country where security and accessibility constraints hinder the capacity to collect ground-truthed data. Our study provides a deep dive into the Badghis province and focuses on settlement-level dynamics, the smallest administrative unit possible. Geospatial data offers a unique opportunity to track the evolution of indicators through time-series satellite images, in hard-to-reach areas (e.g., due to conflicts and natural disasters). With this work, we aim to complement the literature on displacement by providing an example of how to proxy for displacement flows in situations where data are inconsistent and hard to collect. The methodology can be easily replicated to other regions or the entire country. Results also provide a basis to better understand the drivers of displacement drive settlement growth, and explore how once could better predict population displacements based on variations in geospatial data (e.g., shifts in vegetation density due to drought). While this study focuses on the Badghis region where displacement flows have been particularly important, external validity needs to be tested, and those results would make the case for a scale-up of such methodology to the whole Afghanistan.

1.2. Context

The 2020 Global Report ranks Afghanistan among the top five countries with high levels of internal displacement due to conflict and violence and ranks it first with the highest number of IDPs due to natural disasters (IDMC 2020a, 11-12). As of 31 December 2019, a total of 3 million people were internally displaced as a result of the conflict, the highest in the history of Afghanistan (IDMC 2020a, 11). Since 2017, 350,000 – 480,000 people have been displaced annually due to ongoing conflict and violence (OCHA 2020a; OCHA 2021). In 2019, the Taliban controlled more territory of the country than ever since the US-led intervention; clashes between the army and the Taliban, most of which took place in the north, east, and northeast of the country, led to displacements of 461,000 people (IDMC 2020a, 49). In 2020, more than 413,000 people were displaced due to conflict, and a total of 32 out of 34 provinces reported some level of forced displacement (OCHA 2020b). In recent years, there have also been massive internal displacements due to natural disasters. For example, in 2018, the worst drought in decades triggered about 371 000 displacements, and, in 2019, floods and drought caused displacements of 117,000 people (OCHA 2020b, 49-50). As of 31 December 2019, the total number of internally displaced persons as a result of natural disasters reached 1.2 million (OCHA 2020b, 50).

Displaced persons are most visibly seen in border provinces, with a few areas bearing the brunt, such as Herat, Nangarhar and Kabul provinces (Figures 1.1 and 1.2)**.** According to a joint WB and UNHCR report in 2019, around one in three Afghan returnees had settled in Kabul and Nangarhar.[2] The Afghanistan Living Conditions Survey (ALCS) 2017 showed that Nangarhar province was hosting the highest share of displaced population (relative to host communities), as 30% of its total population was a displaced individual.[3] Herat province concentrates the largest share of IDPs hosted among all 34 provinces, with more than 900,000 IDPs arrived since 2012. Nangarhar is the most important province of arrival for returnees from abroad, with almost 525,000 international returnees being hosted in the province.

Badghis province was one of the most negatively affected by the 2018 drought, and the province records the second highest number of internally displaced population in that year across the country (behind Herat province). Between mid-2017 and mid-2018, Afghanistan recorded a deficit of 70 percent precipitation, and insufficient snowfall, which yet constitutes an important source of water for crops and irrigation (FEWSNET). Humanitarian actors identified that the critical situation Herat and Badghis (among the 12 worst affected provinces) had tremendous impact on the affected households, see 2018 reports by the Afghan Red Crescent Society (ARCS)[4], OCHA[5] and Oxfam.[6] These adverse impacts hence translated into large displacement flows. Out of the current IDPs located in the country, almost 293,000 individuals arrived in Herat in 2018, and more than 162,500 in Badghis, making it the second province hosting the largest number of 2018 IDPS. A FEWS report in April 2020 showed that the 2018 drought had long lasting

---

[2] *World Bank Group/UNHCR, Living conditions and settlement decisions of recent afghan returnees, Findings from a 2018 Phone Survey of Afghan Returnees and UNHCR data, June 2019*

[3] *"World Bank. 2017. Afghanistan Development Update, May 2017. Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/27550 License: CC BY 3.0 IGO."*

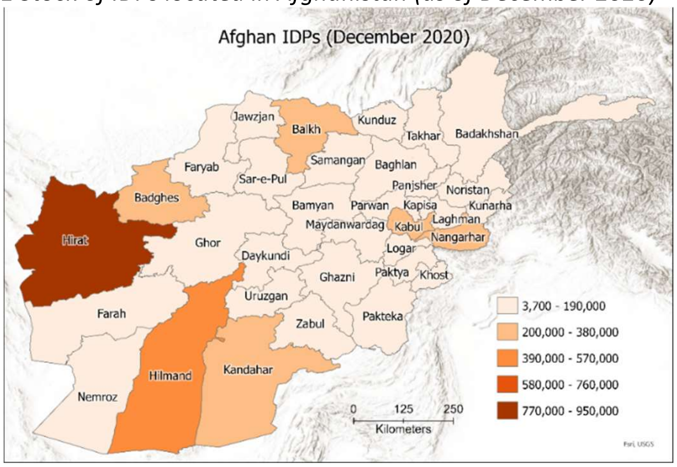
[4] ARCS Information Bulletin n° 1 https://www.ifrc.org/docs/Appeals/18/IBAFdr160518.pdf

[5] https://reliefweb.int/report/afghanistan/afghanistan-drought-response-situation-report-no-2-16-september-2018

[6] https://www.humanitarianresponse.info/sites/www.humanitarianresponse.info/files/assessments/drought_affected_in_hirat_and_badghis_provinces_oxfam_15_august_2018.pdf

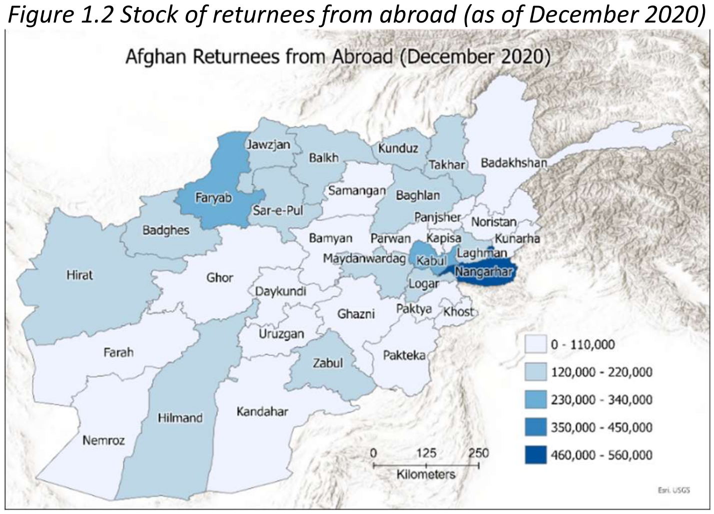
impact, especially on displaced households in urban areas who tend to rely on small business, non-agriculture wage labor, and low salary jobs, as well as remittances.[7]

Figure 1.1 Stock of IDPs located in Afghanistan (as of December 2020)



Source: WB staff estimation using IOM DTM 2020

Figure 1.2 Stock of returnees from abroad (as of December 2020)



Source: WB staff estimation using IOM DTM 2020

2. Literature review on geospatial data and displacement

Using remotely sensed data and satellite imagery in humanitarian situations, Quinn et al. (2018) showed that machine learning algorithms can return precise estimates of forcibly displaced people and geographic structures. They demonstrated how machine learning with satellite imagery could support humanitarian operations by providing objective assessments of the situation in the aftermath of a natural disaster or conflicts events, both situations highly prevalent in the Afghanistan context. [8] Results are

[7] See https://reliefweb.int/sites/reliefweb.int/files/resources/AFGHANISTAN_Food_Security_Outlook_10_2019.pdf

[8] Quinn, John A., et al. "Humanitarian applications of machine learning with remote-sensing data: review and case study in refugee settlement mapping." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376.2128 (2018): 20170363.

further improved when manual analysis of satellite and sensor date are used to support machine learning findings due to the high diversity of image characteristics which complicate the automated processes.

Our study directly complements a body of existing humanitarian work using Nighttime lights (NTL) imagery to monitor displacement crises. Li and Li (2014) explored the spatial and temporal patterns of night-time light in Syria, its international borders and surrounding regions,[9] finding a moderate correlation between NTL loss and numbers of IDPs of each province. Witmer and O'Loughlin (2011) analyzed fluctuations in the NTL levels of cities within the Caucasus region of Russia and Georgia between 1992 and 2009 period, in order to detect conflict-related events, including large flows of populations.[10] Their findings confirmed that large displacement movements were possible to detect through such satellite data.

More generally, there exists a large strand of literature that provide information on displacement related issues by evaluating satellite imagery, e.g., such as buildings and infrastructures. Our work on displacement flows and NTL also echoes Giada et al. (2003), which utilized satellite images from refugee settlements (including temporary accommodations such as tents) to estimate refugee population.[11] The identification of new refugee settlements has also been enabled by remote sensing (UNITAR 2011)[12], which can also provide detailed maps of settlements (Wang et al. 2015, Aravena Pelizari et al. 2018)[13], both topics extremely relevant for our analysis. Several reviews of previous research on remote sensing in conflict and human rights work were elaborated by Marx & Goward (2013)[14], Witmer (2015) and Quinn et al. (2018).[15]

Our study builds on recent progress made to assess displacement through geospatial proxy indicators, including both urban and natural measures. First, the Global Report on Internal Displacement (2018) from the Internal Displacement Monitoring Centre (IDMC) contributed to the global research on displacement by using indicators of housing destruction and flooding gathered through satellite and aerial imagery. In particular, the report pointed out that such methodology was especially effective in urban settings, whereby the large availability of images to track construction was used as a proxy for the duration of displacement. This hence relates directly to our study, which attempt to proxy displacement flows

---

[9] Li X, Li D. 2014 Can night-time light images play a role in evaluating the Syrian Crises? *Int. J. Remote Sens.* 35, 6648–6661. (doi:10.1080/01431161.2014.971469)

[10] Witmer FDW, O'Loughlin J. 2011 Detecting the effects of wars in the Caucasus regions of Russia and Georgia using radiometrically normalized DMSP-OLS nighttime lights imagery. *Gisci. Remote Sens.* 48, 478–500. (doi:10.2747/1548-1603.48.4.478)

[11] Giada S, DeGroeve T, Ehrlich D, Soille P. 2003 Information extraction from very high-resolution satellite imagery over Lukole refugee camp, Tanzania. *Int. J. Remote Sens.* 24, 4251–4266. (doi:10.1080/0143116021000035021)

[12] UNITAR. 2011 UNOSAT Brief 2011—satellite applications for human security. Geneva, Switzerland: United Nations Institute for Training and Research.

[13] See Wang S, So E, Smith P. 2015 Detecting tents to estimate displaced populations for postdisaster relief using high resolution satellite imagery. *Int. J. Appl. Earth Observ. Geoinf.* 36, 87–93. (doi:10.1016/j.jag.2014.11.013). See also Aravena Pelizari P, Sprohnle K, Geib C, Schoepfer E, Plank S, Taubenbock H. 2018 Multisensor feature fusion for very high spatial resolution built-up area extraction in temporary settlements. *Remote Sens. Environ.* 209, 793–807. (doi:10.1016/j.rse.2018.02.025)

[14] Marx A, Goward S. 2013 Remote sensing in human rights and international humanitarian law monitoring: concepts and methods. *Geographic. Rev.* 103, 100–111. (doi:10.1111/j.1931- 0846.2013.00188.x)

[15] Witmer FDW. 2015 Remote sensing of violent conflict: eyes from above. *Int. J. Remote Sens.* 36, 2326–2352. (doi:10.1080/01431161.2015.1035412)

through NTL and city expansion measures. The Global Report on Internal Displacement (2020) from the Internal Displacement Monitoring Centre (IDMC) is an additional example of the interaction between satellite data and displacement research. The IDMC augmented its 2019 interactive monitoring tool with satellite imagery to obtain figures on internally displaced people. The report for instance uses satellite imagery to provide information on displaced population due to housing destruction in several regions of Turkey. Human Right Watch (HRW) has also extensively used time-series satellite imagery to inform about housing destruction (and the subsequent vulnerabilities). Indeed, the IDMC 2020 report based their estimation of the total number of IDPs in Egypt on data from a Human Right Watch 2018 report, which analyzed satellite imagery of housing destruction (complemented by interviews with affecting families in North Sinai).[16] HRW also went over a time-series of satellite images (January 2013-May 2018) to identify and potentially assess the impact of several conflict and military events. [17] In a 2018 paper, HRW also documented how Rohingya villages were being bulldozed in Myanmar.[18] The high frequency of satellite images allowed to track the evolution of destruction over time.

This study contributes to the work produced by the humanitarian-development nexus to predict displacement flows using variation in satellite imagery-derived data (e.g., Normalized Difference Vegetation Index (NDVI), Standardized Precipitation-Evapotranspiration Index (SPEI), and NTL). It is closely related to the UNHCR Winter Cell work project implemented by UNHCR between October 2015 and June 2016, which aimed at forecasting migration flows with weather data, as harsh winter tends to impede movements across borders and within countries. The team collaborated with national migration agencies and several meteorological offices, and monitored social media to elaborate daily weather reports and provide information on possible effect on camps, border controls and transportation (e.g., trains and buses). For instance, the Winter Cell forecasted that the broad area within the eastern Balkan Peninsula, Turkey, the Eastern Mediterranean, and the Middle East, which was hosting millions of displaced populations in camps and settlements, would undergo a harsh 2015/2016 winter (heavy snowfalls and temperatures reaching -5° to -15°[19].A new strand of the economics literature has also been studying the relationship between climate change and internal migration, see Beine and Jeusette (2019)[20] or Cattaneo et al., (2019)[21]. The WB is currently publishing a working paper on the role of climate change on international migration from West Africa, using High Frequency Surveys from IOM.

Finally, our work relates to the larger use of remotely sensed data in the context of human mobility. Social media data have also been used to better understand population flows within a city. As can be found on the Migration Data Portal, an international project including countries in Asia, Europe and the USA handles Twitter data to visualize population distribution in several cities. The objective there is to

---

[16] https://www.hrw.org/news/2018/05/22/egypt-army-intensifies-sinai-home-demolitions
[17] https://www.hrw.org/sites/default/files/report_pdf/egypt0519_web3_0.pdf
[18] https://www.hrw.org/news/2018/02/23/burma-scores-rohingya-villages-bulldozed
[19] https://www.unhcr.org/innovation/migration-mitigation-and-maps-the-predictive-role-of-unhcrs-first-winter-cell/

[20] Beine, M. A. and Jeusette, L. (2019). A Meta-Analysis of the Literature on Climate Change and Migration. *IZA Working Paper No. 12639*.
[21] Cattaneo, C., Beine, M., Fröhlich, C. J., Kniveton, D., Martinez-Zarzoso, I., Mastrorillo, M., Millock, K., Piguet, E. and Schraven, B. (2019). Human Migration in the Era of Climate Change. *Review of Environmental Economics and Policy*, 13 (2), 189–206

generate maps through artificial intelligence algorithms to refine our understanding of the diverse flows between long-term residents and short-terms visitors (including tourism).[22] Another timely example can be found in the ongoing COVID-19 Mobility Analysis of Statistics Estonia aims at tracking human mobility under the government-mandated mobility restrictions in the country, by using mobile phone data.[23]

This paper contributes to the literature on geospatial data and displacement by combining remotely sensed data and IOM administrative data in Afghanistan, a country where solid data evidence can be hard to gather. Our study provides a deep dive on Badghis province, hard hit by the 2018 drought, by focusing on dynamics at the settlement level, the smallest administrative unit possible. Geospatial data offers a unique opportunity to track the evolution of indicators through time-series satellite images, in hard-to-reach areas (e.g., due to conflicts and natural disasters). This work capitalizes on administrative data from IOM, which provides an exhaustive overview of the overall flow of displaced population in the region, at the settlement levels. Combining both administrative data and satellite images returns a unique framework to identify interesting correlation between geospatial indicators and displacement flows. With this work, we aim to complement the literature on displacement by providing an example of how to proxy for displacement flows in situations where data is volatile and hard to collect.

3. The data

3.1. Displacement and NTL data

Through its original approach, this study exploits a unique set of data from IOM, the Displacement Tracking Matrix (DTM December 2020), to identify detailed displacement patterns in the Badghis region following the 2018-drought. The IOM data being used for this analysis corresponds to 10 rounds of Baseline Mobility Assessments (BMA), conducted between January 2018 and December 2020. The information is collected at the settlement level, through focus group interviews with Key Informants, and includes information on settlements' location (GPS) and inflows and outflows of displaced population. These include both Internally Displaced Persons (IDPs) coming from other settlements, and returnees, former IDPs returning to their origin settlements. The DTM data is a detailed and unique source of information on displaced patterns in Afghanistan. Compared to Afghanistan Income Expenditure and Labor Force Surveys, the IOM dataset directly focuses on settlements with mobile population, ensuring a full representability of displacement in the country. Compared to other humanitarian agencies such as OCHA, it is also the only providing information on IDPs, returnees from abroad and migrants, regardless of their legal status and reasons for displacement.[24] In addition, while OCHA supplies more frequent information on flows of internal displacement at the district level, the information cannot be disaggregated at the village level.

IOM did not collect data prior to the year 2018 in Badghis, suggesting that there were limited evidence of major displacement flows before then, hence justifying our focus on the 2018-2020 period.

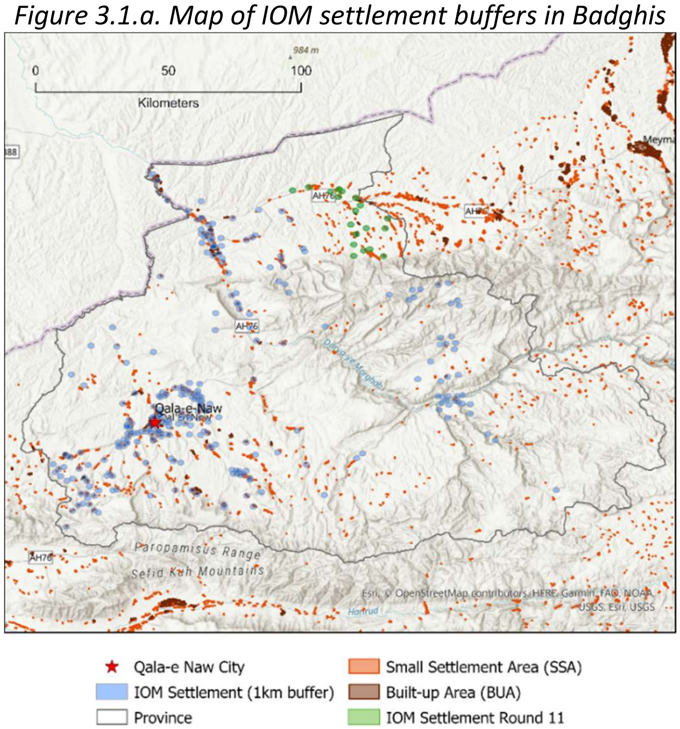
Limitations in terms of measurement errors are nevertheless important to keep in mind, and further make the case for better data on displacement in Afghanistan. Due to the absence of a clear definition

[22] https://migrationdataportal.org/data-innovation-58
[23] https://migrationdataportal.org/data-innovation/covid-19-mobility-analysis-statistics-estonia-collaborating-mobile-network
[24] OCHA data only focuses on conflict-induced IDPs, while UNHCR data reports numbers for documented returns.

on geospatial borders for IOM settlements, the possibility that settlements buffers could overlap cannot be excluded. This could bias the results if the weights of specific outcomes are artificially inflated by appearing multiple times in overlapping buffers, for what should actually be a unique settlement. In other words, in denser areas some of the buffers may overlap, hence the population summarized can be duplicated. Therefore, displacement numbers might also be overestimated by counting multiple times the same population. This however is not an important issue for our analysis, which focuses on the evolution of the displacement flows in settlements rather than absolute values. In addition, the survey relies on self-reported estimates, which can be associated with important measurement errors. There again, assuming that the bias remains constant throughout time, this issue is unlikely to affect dramatically the analysis.

The IOM Displacement Tracking Matrix (DTM December 2020) allowed the team to identify 294 settlements in Badghis. As of December 2020, there were 294 settlements in Badghis for which IOM has collected data on inflows and outflows of persons (Figure 3.1.a).[25] Summary statistics are provided in Table 3.0. Among these 294 settlements, 247 (84%) were surveyed 4 times during 2018 and 2020, 7 (2%) were surveyed 3 times, 20 settlements (7%) were surveyed twice, and 20 settlements (7%) recently emerged in the northern bound of the province and were surveyed between June and December 2020 by IOM (round 11 of data), see Figure 3.1.b. In total, we then have 1,069 observations (with multiple observations per settlements).

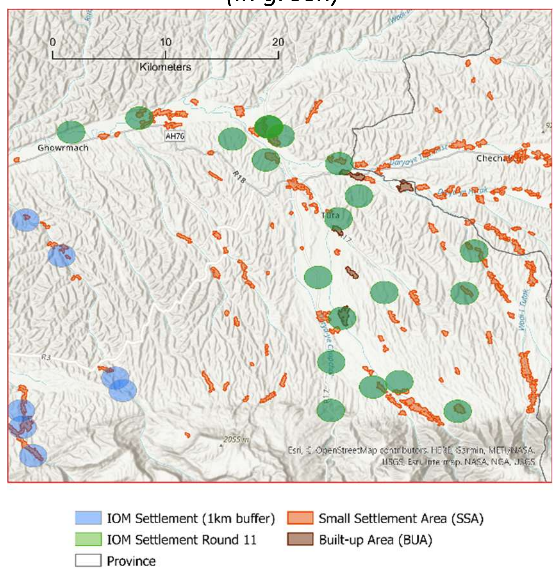
Figure 3.1.a. Map of IOM settlement buffers in Badghis



Source: WB staff' computation using IOM DTM 2020.

[25] Out of these 294 settlements, IOM collected the first flow data in 2018 for 248 settlements, while 7 settlements started being surveyed in 2019, and 19 were first surveyed before June 2020, and 20 surveyed in December 2020. This indicates that, prior to these dates, there were no significant inflows of displaced population.

Figure 3.1.b. *Zoom on IOM settlement that emerged between June and December 2020 (in green)*



Source: WB staff' computation using IOM DTM 2020.

Table 3.0. Summary statistics of displacement flows in Badghis settlements across the 2018-2020

Variable	Observations	Mean	Std. Dev.	Min	Max
IDPs arrivals	1,069	1048.3	5236.9	0	81545
IDPs returning	1,069	315.5	360.8	0	4116
IDPs fleeing	1,069	710.2	943.3	0	7530
Returnees from abroad	1,069	416.6	441.7	0	3397
Out migration	1,069	441.5	355.8	0	2552

Note: There are 294 settlements, among which 84% were surveyed 4 times over the 2018-2020 period, 2% surveyed 3 times, 20% surveyed twice and 20% surveyed once (in 2020). Hence, the total number of observations is 1,069.

Second, each settlement was associated with its average monthly NTL value around a 1km buffer, from January 2018 to January 2020. The imagery is obtained from the Day/night Band (DNB) of NOAA's VIIRS platform, known as Nighttime Light (NTL). This sensor has a ground resolution of approximately 750 meters by 750 meters. NTL from VIIRS, and its predecessor Defense Meteorological Satellite Program (DMSP), capture low-light emissions from earth. "These include sources that indicate aspects of human activity, like city lights, gas flares, fishing boats, and agricultural fires, while also capturing other nighttime lights phenomena such as auroras."[26] For additional data on VIIRS DNB and NTL in general please see Lights Every Night. This data is available for every month in the studied period (2018-2020), with the exception of seasonal issues with reflectance and the DNB Sensor (for June).[27]The team then created a buffer with 1km radius around the IOM settlements in Badghis and extracted the average monthly NTL value within these areas to capture the monthly evolution of nightlights from 2018 to 2020. An overview of the average NTL across IOM settlement for the entire 2020 year is displayed in Figure 3.2.b. Summary statistics are available in Table 3.3. The IOM data does not contain information on the actual settlements' sizes; hence we focus on the average NTL levels across a 1km-by-1km cell, centered around the GPS point

[26] Remotely-sensed data of Nighttime lights.
https://worldbank.github.io/OpenNightLights/tutorials/mod1_2_introduction_to_nighttime_light_data.html

[27] There are hence 9702 observation for NTL values, whereby 9702 = 2x11x294

collected by IOM. This relies on the assumption that IOM staff members recorded a central GPS point, around which NTL activity is likely to be representative of the urbanization dynamic. As a reference, it takes roughly 10 minutes for a person to walk 1km. While the absence of solid data on settlement boundaries is a caveat, robustness checks will be run when using a 2km and 5km radius.

Figure 3.2.a. Visualization of average NTL in January 2020 in Badghis

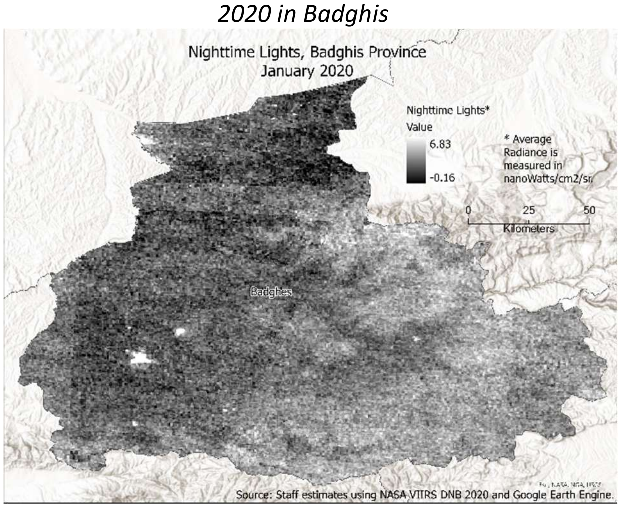
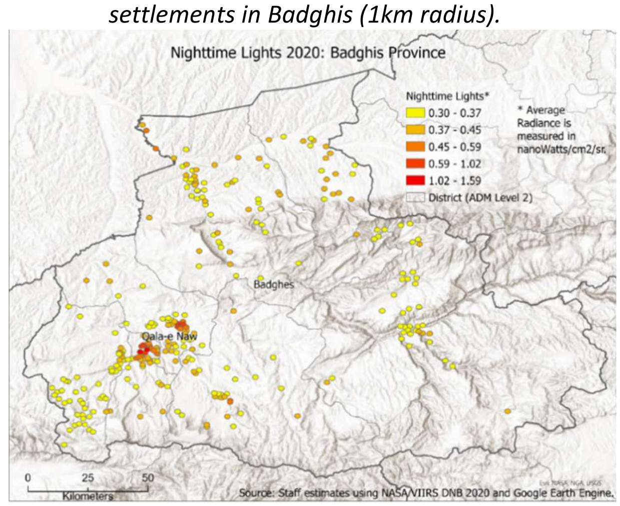


Figure 3.2.b Average NTL across the year 2020 in IOM settlements in Badghis (1km radius).



Note: WB staff computation. The mean NTL across all settlements in Badghis is 0.309, the standard deviation is 0.14, the minimum level is 0.03 and the maximum is 2.215, on a total of 9,702 cells.

Table 3.3 Summary statistics of average NTL across the Badghis region

Variable	*Mean*	*Std. Dev.*	*Min*	*Max*	*Observations*
Night-Time Light (NTL)	0.309	0.140	0.030	2.215	*9,702*

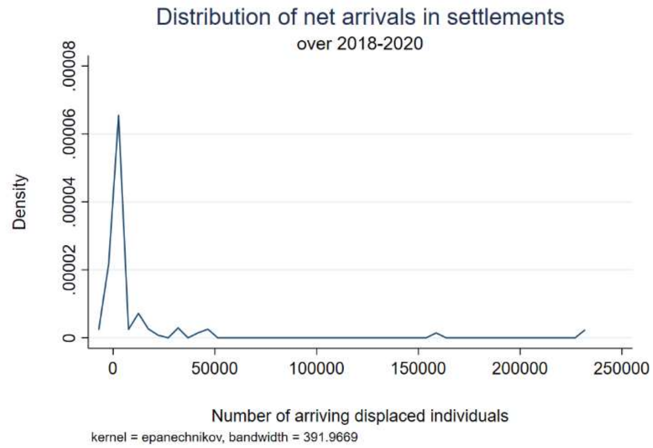
Source: WB staffs' computation.

3.2. Main regressor

To identify settlements receiving many displaced persons (IDPs, returnees and migrants), we look at the cumulative net inflow of persons in a settlement over December 2018, June 2019 and 2020 in the IOM data.[28] This is defined as the total inflows of persons, minus the total outflows. Let's label *Total Inflow* the variable that corresponds to the total net inflow of persons in a given settlement between December 2018 and June 2020 cumulated. Figure 3.3 shows the density of the *Total Inflow* variable across settlements in Badghis, and summary statistics of the variable can be observed in Table 3.4.

[28] Displacement data in the Badghis region was only collected by IOM in December 2018 , June 2019, and June 2020.

Figure 3.3 Distribution of the cumulative net inflow of persons (Total Inflow) *in Badghis*



Source: WB staffs' computations.

Table 3.4. Summary statistics for the cumulative net inflow in Badghis settlements

Variable	Nb settlements	Mean	Std. Dev.	Min	Max
$Total\ Inflow$	294	3254.905	18706.09	-6690	231486

Note: The net inflow of persons can be negative if the population that migrates out of the settlement (either internationally or within Afghanistan) is higher than the number of displaced persons settling in.

3.3. Controls

The correlation between displacement patterns and NTL growth could be affected by the initial urbanization level of the settlement, it is thus important to control for population estimates in 2017 (pre-drought levels). One could indeed be concerned that displacement flows to hosting areas are not random, as IDPs may prefer to settle in urbanized areas, where access to services is easier. One might be concerned that large flows of displaced persons tend to converge to large settlements, with specific NTL growth pattern, hence skewing our interpretation of the data. There is evidence that displaced Afghans flow to cities, where security and availability of services are perceived to be greater, see e.g. IDMC (2020)[29] or EASO (2020).[30] IPDs affected by conflict tend to flee rural areas for their regional centers (Samuel Hall/NRC/IDMC 2020, 20).[31] As the findings of the study confirm, IDPs assume that urban areas are safer and, in these areas, where employment opportunities, services, and humanitarian aid are more readily accessible, they would be more able to cope. Furthermore, IDPs seem to prefer to stay closer to their places of origin. Social ties could also be at stake, whereby migrants tend to congregate to places in which IDPs already settled, to capitalize on network effect. Thus, it is likely that large displacement flows converge to already large settlements with specific NTL growth pattern. Failing to account for initial settlements size (proxied by 2017 population levels) could skew our interpretation of the data. To prevent

[29] IDMC. 2020b. "A Different Kind of Pressure: The Cumulative Effects of Displacement and Return in Afghanistan".

[30] *ASO (European Asylum Support Office). 2020.* "Afghanistan Key Socio-Economic Indicators Focus on Kabul City, Mazar-e Sharif and Herat City".

[31] According to the study by Samuel Hall/NRC/IDMC conducted in 2017 in the provinces of Kabul, Herat, Kandahar, Kunduz, and Nangarhar, 92% of the respondents in the south-west of the country had moved to Kandahar city, 91% in the west to Herat city, and 76% in the east to Jalalabad (Samuel Hall/NRC/IDMC 2020, 20).

overweighting existing urban centers with high baseline NTL, the team therefore incorporated 2017 population data from the Government of Afghanistan (population originally from WorldPop, see Section 4.3).

In addition, we also attempt to proxy for the type of economic activity present in settlement, as the correlation between displacement flows and NTL growth might depend on the share of agricultural labor and urbanization levels. In the Round 10 of data (between June and December 2020), IOM started collecting data on the percentage of settlement income derived from main sectors. Out of the 294 Badghis settlement, 293 contains information on the proportion of agriculture and farming as a share of community income, as opposed to other activities such as trade, services, manufacturing, etc.). While it is unfortunate that this data does not exist for previous round, this caveat can be mitigated by controlling for the relative importance of agriculture in settlement in late 2020, which returns information on the general level of urbanity in settlement over the 2018-2020 period. We create the variable *Agric2020* that associates each settlement with its proportion of income obtained through agriculture (farming, crop production, etc.) and livestock (cattle, sheep, fish farming, etc.). For each of the settlements, summary statistics are presented in Table 3.5. On average, settlements derived 71% of their income from agricultural activities, with a minimum of 0% and a maximum of 100%.

Table 3.5.Statistics of the proportion of community income derived from agriculture in 2020 (Badghis).

Variable	Number of settlements	Mean	Std. Dev.	Min	Max
$Agric2020$	293	70.66	20.46	0	100

Source: WB staff's computation

4. Regression analysis

In this analysis, we regress the NTL growth on the total level of displacement in settlement, represented by the cumulative net inflow of displaced persons over the 2018-2020 period. The linear regression will return information on the general evolution of NTL, i.e., whether net inflow of displaced persons correlate with higher NTL growth. As will be seen in robustness checks, quadratic regressions can help identifying non-linear dynamics, i.e., whether displaced persons eventually leave the place they fled to ("boom town effect") or remain in the hosting location ("lock-in effects").

4.1. Benchmark NTL growth

This section explores the NTL growth rate observed in Badghis, i.e., the NTL evolution over time, before accounting for potential deviations from displacement patterns. In research, NTL patterns have shown that over time, radiance increases in-place as well as expands spatially. In other words, patterns indicate improvements in overall electrification, or an increase in other ambient light sources like fires, in a previously electrified areas, as well as increases or appearance of electrification in areas with little or no historic light emittance. These findings can be shown in Li et al. (2018), which used a global NTL time-series data (1992–2018) to show that the NTL time series experienced a trend of both increasing NTL and continuous spatial expansion for high luminance pixels, both in urban centers and fringe locations.[32] This indicates that the NTL growth rate for a given area will primarily be exhibited in an increase in overall average NTL for an already electrified area, or will indicate the presence of new electrification in areas

[32] Li, Xuecao, et al. "A harmonized global nighttime light dataset 1992–2018." Scientific data 7.1 (2020): 1-9.

with improved infrastructure or newly inhabited areas. While there is mostly growth and expansion of NTL over time, NTL levels do remain the same based on existing infrastructure and outdoor lighting in a given place, so continuity can be expected across years with some slight increases or decreases depending on actual electricity use as well as exogenous factors such as data collection changes and significant change in infrastructure (e.g., the changed presence of a military base or other large infrastructure).

As discussed earlier, NTL growth might be influenced by the original size of the settlement and general urbanization levels, the analysis includes a benchmark population estimate (2017) and a measure of the share of agricultural income at settlement level. Indeed, there are reasons to believe that the growth rate of a city depends on its initial size. Keeping in mind that the cumulative net inflow is based on the 2018 to 2020 flows of persons, the population levels of 2017 allow to proxy for initial urbanization levels and find the relevant benchmark to each settlement.

Table 4.1 displays the linear and quadratic regression of NTL levels on its past values, at the settlement level, controlling for district fixed effects. Let's define $NTL_{i,j,t}$ the average NTL in the 1km buffer around the settlement *i* (out of the 297 settlements) in district *j,* at date *t (*each month from Jan 2018 to Jan 2020). The following regressions looks at how the NTL levels naturally grow through time in settlements within Badghis province, accounting for the initial population in 2017, the share of community income derived from agricultural and livestock activities in late 2020, time seasonality (μ_t the time fixed effect) and δ_j the fixed effect for district *j*. Equation-1 assumes a linear evolution of NTL, while equation-2 is quadratic.

$$NTL_{i,j,t} = \beta_0 + \beta_1 NTL_{i,j,t-1} + \beta_2 Pop2017_i + \beta_3 Agric2020_i + \delta_j + \mu_t + u_{i,j,t} \qquad \text{(equation-1)}$$

$$NTL_{i,j,t} = \beta_0 + \beta_1 NTL_{i,j,t-1} + \beta_2 (NTL_{i,j,t-1})^2 + \beta_3 Pop2017_i + \beta_4 Agric2020_i + \delta_j + \mu_t + u_{i,j,t} \qquad \text{(equation-2)}$$

The NTL evolution is such that

$$\frac{dNTL_{i,j,t}}{dNTL_{i,j,t-1}} = \hat{\beta}_1 + 2\hat{\beta}_2 NTL_{i,j,t-1} \qquad \text{(equation-3)}$$

The regression analysis of the growth rate of NTL in Badghis settlement is positive and concave, as settlements with higher NTL will experience larger NTL in the next period. The linear regression suggest that the NTL growth follows a concave shape. On average, in an IOM settlement *i* within Badghis, the natural growth of NTL levels is such that a 1-unit increase in past NTL value ($NTL_{i,j,t-1}$) raise the current NTL if the past NTL value is smaller than 2.22, which is always true in the Badghis sample.[33]

Table 4.1 Linear and quadratic regression of NTL its lagged value and district fixed effects

	NTL levels	
VARIABLES	(1) Linear	(2) Quadratic
β_1; Lagged NTL	0.780***	1.194***
	(0.00738)	(0.0205)
β_2; (Lagged NTL)^2		-0.269***
		(0.0125)
Population 2017	1.13e-05***	9.03e-06***
	(7.75e-07)	(7.61e-07)

[33] Equation-3 is positive if and only if $NTL_{i,j,t-1}$> 1.194/0.534.

% agriculture in 2020	1.39e-05	3.97e-05
	(4.13e-05)	(4.02e-05)
Constant	0.0119*	-0.118***
	(0.00609)	(0.00846)
Observations	8,497	8,497
Number of settlements	293	293
District Fixed Effects	Yes	Yes
Time Fixed Effects	Yes	Yes

Standard errors in parenthesis
*** p<0.01, ** p<0.05, * p<0.1

4.2. Impact of displacements flows on NTL growth

Linear regressions

Additional population settling in the location are likely to impose deviation from the NTL growth, e.g., through settlements' expansion, growing services needs or the modification in the economic activity. High inflows of displaced population are associated with a growing need for basic electricity services such as indoor and street lighting. In addition, they could directly impact settlements' economic activity, e.g., by modifying the repartition of labor force throughout the different sectors, or through a direct increase in settlements' population. As NTL levels are traditionally used to monitor human activities, see e.g., Li et al. (2018), important influxes of persons are expected to impact the evolution of NTL levels. This note therefore does not assume a fixed growth rate of NTL, but rather compares growth rates across settlements and provides insights on why some settlements have higher growth rate than others - reflecting an additional inflow of migrants but also increased economic activity.

A linear regression at the settlement level allows to check for correlation between NTL growth and overall displacement flows between 2018 and 2020.[34] Let's define $NTL_{i,j,t}$ the average NTL in the 1km buffer around the settlement *i* (out of the 297 settlements) in district *j,* at date *t (*from Jan 2018 to Jan 2020). The following linear regression looks at how the NTL growth across time is affected by the continuous variable $Total\ Inflow$, representing the cumulative net inflow of persons in settlement over the 2018-2020 period (see Section 4.1 for descriptive statistics). The regression controls for δ_j the fixed effect for district *j* and μ_t the time fixed effect.

Defining X = $Total\ Inflow_i$, the regression is hence:

$$NTL_{i,t} = \alpha_0 + \alpha_1 NTL_{i,t-1} + \alpha_2 X_i + \alpha_3 X_i * NTL_{i,t-1} + \alpha_4 Pop2017_i + \alpha_5 Agric2020_i + \delta_j + \mu_t + u_{i,t} \quad \text{(equation-4)}$$

The NTL growth therefore follows the following equations

$$\frac{\mathrm{d}^2\,\mathrm{NT}_{\mathrm{i,t}}}{\mathrm{d\,NTL_{i,t-1}}\,\mathrm{d}X_i} = \alpha_3 \quad \text{(equation-5)}$$

$$\frac{\mathrm{d\,NTL_{i,t}}}{\mathrm{d\,X}} = \alpha_2 + \alpha_3 \mathrm{NTL_{i,t-1}} \quad \text{(equation-6)}$$

[34] The average NTL is not available for June 2020

Quadratic regressions

While the linear regression returned information on the general trend, the quadratic regression would allow to account for non-linear dynamics. For instance, a quadratic regression would identify any "boom town effect", whereby displaced persons eventually leave the place they fled to. Similarly, it would help identify "lock-in effects", observed when displaced population remains in the hosting location. Hence, the coefficients α_3 *and* α_5 of the quadratic regression are of particular interest, as they represent the correlation between high inflows of persons and the NTL growth (both linear and quadratic).

Defining X = $Total\ Inflow_i$, the regression is hence

$$NTL_{i,t} = \alpha_0 + \alpha_1 NTL_{i,t-1} + \alpha_2 X_i + \alpha_3 X_i * NTL_{i,t-1} + \alpha_4 (NTL_{i,t-1})^2 + \alpha_5 X_i * (NTL_{i,t-1})^2 + \alpha_6 (X_i)^2 + \alpha_7 (X_i)^2 * NTL_{i,t-1} + \alpha_8 (X_i)^2 * (NTL_{i,t-1})^2 + \alpha_6 Pop2017 + \mu_t + u_{i,t} \quad \text{(equation-7)}$$

One obtains that

$$\frac{\mathrm{d\,NTL_{i,t}}}{\mathrm{d}X_i} = \alpha_2 + \alpha_3 NTL_{i,t-1} + \alpha_5 NTL_{i,t-1}^2 + 2.\mathrm{X_i}(\alpha_6 + \alpha_7 NTL_{i,t-1} + \alpha_8 NTL_{i,t-1}^2) \quad \text{(equation-8)}$$

$$\frac{\mathrm{d^2\,NT\quad_{i,t}}}{\mathrm{d}X_i^2} = 2(\alpha_6 + \alpha_7 NTL_{i,t-1} + \alpha_8 NTL_{i,t-1}^2) \quad \text{(equation-9)}$$

Results

First, the linear regression suggests that settlements with higher net inflow of displaced population experience a larger NTL growth, regardless of whether the settlement is a net receiver or serves as outbound areas. As can be seen in column (1) of Table 4.5, coefficient β_3 is positive, i.e., equation-5 is positive and $\frac{\mathrm{d\,NTL_{i,t}}}{\mathrm{d\,NTL_{i,t-1}}}$ is increasing with the net inflow of displaced persons. On the other hand, settlements that served as outbound (in terms of cumulated flows during 2018-2020) are defined by a negative *Total Inflow*. Therefore, a positive β_3 signifies that, in outbound areas, a marginal increase in outflows (i.e., decrease in net inflows) drops the NTL levels.

Second, the linear regression shows that a marginal increase in the net inflow of displaced population raises the NTL level, if the settlement already recorded some level of human activity (i.e., the current NTL level is not too small). Column (1) of Table 4.5. shows that coefficient β_2 is negative, but β_3 positive, therefore the sign of equation-6 depends on the NTL level in past period. Settlements with higher net inflow of displaced persons are associated with lower NTL to start with, but with higher consecutive growth (positive interaction with lagged NTL). A marginal increase in the net inflow of displaced population increases the NTL levels if equation-6 is positive, that is, if $\mathrm{NTL_{i,t-1}} > -\hat{\alpha}_2/\hat{\alpha}_3$. This is true if $\mathrm{NTL_{i,t-1}}$ is larger than 0.28. As a reference, all settlements studied recorded an NTL value above 0.28 at least one during the 2018-2020 period. [35] In other words, settlements receiving a higher number of displaced populations experience a higher NTL growth if the Nighttime Light level is not too small to start with.

Table 4.5 Regression of NTL on net inflows of persons in settlement (cumulated over 2018-2020)

[35] NTL levels around Badghis settlement evolved between 0.03 and 2.215 during the 2028-2020 period (see Table 3.3).

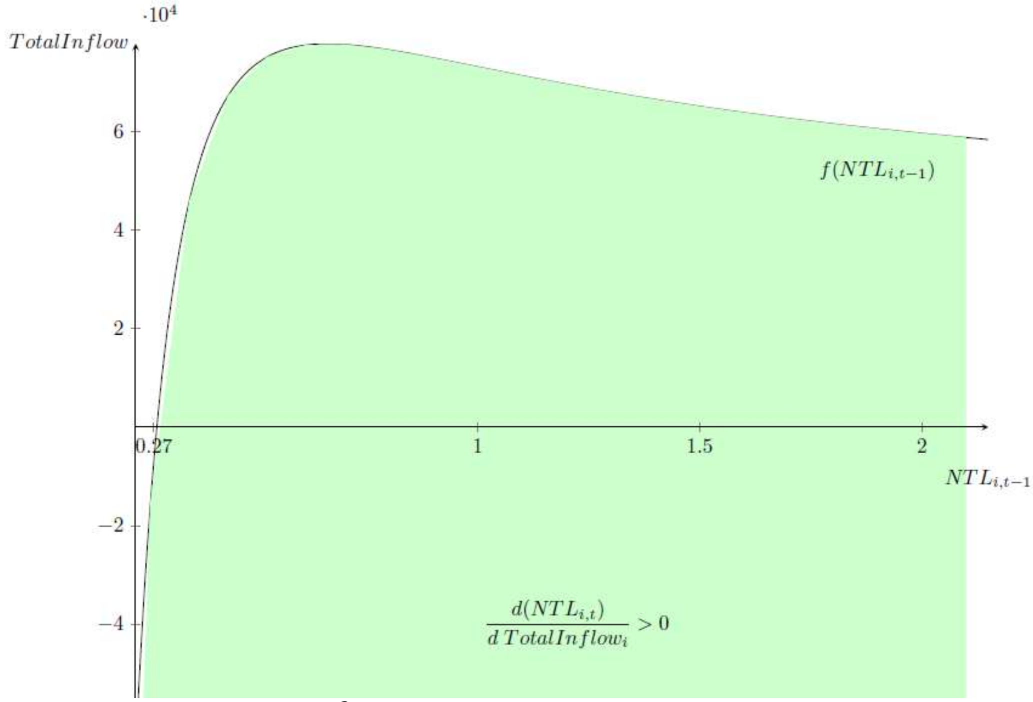
	Y = NTL level	
	(1)	(2)
VARIABLES	Linear	Quadratic
α_1; Lagged NTL	0.744***	1.162***
	(0.00805)	(0.0220)
α_4; (Lagged NTL)^2		-0.337***
		(0.0140)
α_2; Total Inflow	-5.40e-07***	-2.10e-06***
	(9.38e-08)	(4.65e-07)
α_3; (Lagged NTL)*Total Inflow	1.88e-06***	6.36e-06***
	(1.97e-07)	(1.62e-06)
α_5; (Lagged NTL)^2* Total Inflow		4.44e-06***
		(1.05e-06)
α_6; (Total Inflow)^2		0*
		(0)
α_7; (Lagged NTL)*(Total Inflow)^2		0
		(0)
α_8; (Lagged NTL)^2*(Total Inflow)^2		-5.99e-11***
		(0)
α_9; Population in 2017	1.07e-05***	5.86e-06***
	(7.73e-07)	(7.50e-07)
α_{10}; % of agriculture in 2020	8.21e-05**	0.000129***
	(4.18e-05)	(3.97e-05)
α_0; Constant	0.0219***	-0.0964***
	(0.00616)	(0.00855)
Observations	8,497	8,497
Number of settlements	293	293
District Fixed Effect	Yes	Yes
Time Fixed Effects	Yes	Yes

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Second, when introducing a quadratic to allow for non-linearity the results suggest that the effect is concave on displacement inflows, and that the quadratic terms on lagged NTL will depend on the size of Total Inflow. In other words, the settlements that show the largest level of expansion are those with the lowest level of displacement inflows to start with. On the other hand, the quadratic terms on lagged NTL are less straightforward to interpret, but a sufficient condition for the effect to be concave in lagged NTL is that Total Inflow > 74,124 (which comes from 4.44e-06/5.99e-11).

While quadratic regression yields mixed results, there are several cases whereby a marginal increase in the *Total Inflow* of displaced persons in a settlement (over the 2018-2020 period) would increase the NTL levels, in settlements that are both net receivers (inbound) and senders (outbound). Figure 4.1 shows the set of combination (*Total Inflow*$_i$, $NTL_{i,t-1}$) such that *Total Inflow*$_i$ is lower than f($NTL_{i,t-1}$) returns an equation-8 positive, i.e., $\frac{\mathrm{d\,NT}_{i,t}}{\mathrm{d}Total\,Inflow_i} > 0$.

Figure 4.1 Conditions(in green) such that equation-8 positive



Note : f(x) = f(x) $\equiv$ $37310.9 - 17647.1/x^2 + 53445.4/x$. Any combination (*Total Inflow*$_i$, $NTL_{i,t-1}$) such that *Total Inflow*$_i$ is lower than f($NTL_{i,t-1}$) returns an equation-8 positive, i.e., $\frac{\mathrm{d\,NTL_{i,t}}}{\mathrm{d}Tota\ \ Inflow_i} > 0$. Source: Authors' computations.

In inbound areas, a marginal increase in *Total Inflow* would increase NTL level when the inflow of displaced person is low enough and there is some level of human activities (NTL > 0.28)**; however it will always drop the NTL level if the settlement serves as important inbound with a *Total Inflow* larger than 77,255 displaced person.** [36] This is because equation-8 is positive only if $\mathrm{Total\ Inflow_i} < \mathrm{f\,(NTL_{i,t-1})}$ with f(x) $\equiv$ $37310.9 - 17647.1/x^2 + 53445.4/x$. A graphic analysis is displayed in Figure 4.1, whereby a marginal increase in *Total inflow* would raise the NTL as long as Total inflow smaller than f(NTL) – see grey shaded area. We have that $\mathrm{f(NTL_{i,t-1})}$ <0 for all $\mathrm{NTL_{i,t-1}} \in]0;\ 0.28[$. The function f is decreasing from 0 to 0.67, and increasing from 0.67 onward, with f(0.66) = 77,255.[37] In addition, equation-8 is negative when $\mathrm{NTL_{i,t-1}} = 0$, and $\lim_{x \to +\infty} f(x)$ =37,071.

In outbound areas, a marginal increase in the net inflow of displaced person would raise the NTL levels whenever the total outflow is negative enough, i.e., $\mathrm{Total\ Inflow_i} < \mathrm{f}(\mathrm{NTL_{i,t-1}}) < 0$. As can be seen in the lower part of the graph in Figure 4.1, $\lim_{\mathrm{NTL_{i,t-1}} \to 0} f(\mathrm{NTL_{i,t-1}})$ =-∞, whenever Total Inflow is sufficiently negative, we find that equation-8 is positive for all $\mathrm{NTL_{i,t-1}}$>0.

[36]Equation-8 can be written as $\frac{\mathrm{d\,NTL_{i,t}}}{\mathrm{d}X_i} = -2.10\mathrm{e}^{-06} + 6.36\mathrm{e}^{-06} NTL_{i,t-1} + 4.44e^{-06} NTL^2_{i,t-1} + 2.\mathrm{X_i}(-5.99\mathrm{e}^{-11} NTL^2_{i,t-1})$. When $\mathrm{X_i}$ is positive, equation-8 is positive if 0 < $\mathrm{X_i}$ < $\frac{6.36\ \mathrm{NTL_{i,t-1}} + 4.44(\mathrm{NT\ \ _{i,t-1}})^2 - 2.10}{11.98\mathrm{e}^{-05}(\mathrm{NTL_{i,t-1}})^2}$ = f($\mathrm{NTL_{i,t-1}}$). However, the numerator is positive as long as $\mathrm{NTL_{i,t-1}} >$ 0.28; and equation-8 negative if $\mathrm{NTL_{i,t-1}} = 0.28$. In addition, f(x) can be re-written as with f(x) = $37310.9 - \frac{17647.1}{x^2} + \frac{53445.4}{x}$

[37] The derivative is such that f'(x) = (2.11-6.21x)/x[3]

In settlements that serve as outbound areas for a small number of displaced persons, a marginal decrease in outflows (i.e., increase in *Total Inflow*) would drop the NTL level if the past NTL level is small enough. Equation-8 is negative whenever $\mathrm{f}(\mathrm{NTL}_{\mathrm{i,t-1}}) < \text{Total Inflow}_{\mathrm{i}} < 0$. In other words, in outbound settlements with small outflows, there exists a range of NTL (smaller than 0.28) such that a marginal increase in outflow (i.e., reduction in the total outflows) would reduce the NTL levels.

5. Robustness tests

5.1. Variation in settlements' buffer sizes

Robustness checks will be run, by modifying the monthly NTL values associated to each settlement, through changes in the size of the catchment area around settlements GPS points (from 1km to 2 and 5km radius). As discussed earlier, we do not have information on the actual settlements' borders, and must rely on averaging NTL values around the given GPS location. While the study focused on a 1km-by-1km cell, this section extracts the average NTL values observed within 2 and 5km radius around the IOM settlements GPS point. As a reference, a 2km distance would take roughly a 20min walk, while 5km would be completed in 1hour. Summary statistics are available in Table 5.0.1.

Table 5.0.1. Summary statistics of NTL levels, average across 2km and 5km radius around GPS points

Variable	Observations	Mean	Std. Dev.	Min	Max
NTL (average across 2km radius)	9,702	0.31	0.11	0.04	1.63
NTL (average across 5km radius)	9,702	0.31	0.14	0.03	1.85

Source: Authors' computation

Linear results are robust to the construction of NTL average based on a 2km and 5km radius instead of 1km, and the 5km radius results suggest an even stronger positive correlation between *Total Inflow* and NTL growth. The 2km radius analysis suggests that a marginal increase in *Total Inflow* would increase the NTL levels in settlements which had a past NTL levels larger than 0.64. The 5km radius analysis shows that a marginal rise in *Total Inflow* would always be associated with an increase in NTL levels.

Quadratic results become more complex when using the NTL average based on a 2 and 5km radius instead of 1km, as some results differ.

In inbound areas:

- **Using the 2km radius, a marginal increase in *Total Inflow* would increase NTL level when the initial level of human activity is low enough** (for NTL below 0.22), **while the 5km radius analysis suggest that this is true** (for NTL below 0.31), **provided that the settlement is an important receiver of net inflows, i.e.,** the Total Inflow is high enough.
    - Using the 2km radius average for NTL values, equation-8 can be decomposed as $\frac{\mathrm{d\,NTL_{i,t}}}{\mathrm{d}X_i} \equiv$ $\mathrm{h}(NTL_{i,t-1}) + \mathrm{X_i}.\mathrm{f}(NTL_{i,t-1})$, since $\frac{\mathrm{d\,NTL_{i,t}}}{\mathrm{d}X_i} = -3.44\mathrm{e}^{-06} + 1.97\mathrm{e}^{-05}NTL_{i,t-1} - 2.37\mathrm{e}^{-05}NTL_{i,t-1}^2 + 2.\mathrm{X_i}(9.40e^{-12} - 5.95\mathrm{e}^{-11}NTL_{i,t-1} + 7.66\mathrm{e}^{-11}NTL_{i,t-1}^2)$. One finds that f(NTL) is positive whenever NTL < 0.22 or whenever NTL >0.56. The function h(NTL) is positive when NTL in [0.25; 0.79].
    - Using the 5km radius average for NTL values, equation-8 can be decomposed as $\frac{\mathrm{d\,NTL_{i,t}}}{\mathrm{d}X_i} = -5.1\mathrm{e}^{-06} + 1.87\mathrm{e}^{-05}NTL_{i,t-1} - 5.75\mathrm{e}^{-05}NTL_{i,t-1}^2 + 2.\mathrm{X_i}(1.88e^{-11} - 6.05\mathrm{e}^{-11}NTL_{i,t-1}) \equiv$ $\mathrm{h(NTL)} + \mathrm{X_i}.\mathrm{f}(NTL_{i,t-1})$, with $= Total\ Inflow$. One finds that $\mathrm{f}(NTL_{i,t-1})$>0 whenever $NTL_{i,t-1} < 0.31$, and that the function *h* is

always negative. Hence, equation-8 is positive whenever past NTL levels are above 0.31 and the *Total Inflow* is large enough.

- **Using the 2km radius, a marginal increase in *Total Inflow* would increase NTL level in areas with already some level of human activity** (NTL > 0.56) **as soon as the inflow of displaced persons is high enough; while the 5km radius infirm this finding.** Using the 5km NTL average, there exists no level of net inflow Total Inflow high enough so that an increase in net inflow would raise the NTL levels, even in areas with some level of human activities

In outbound areas, a marginal increase in the net inflow of displaced person would raise the NTL levels whenever the total outflow is negative enough in areas with past NTL levels high enough (above 0.56 for the 2km analysis, and above 0.31 for the 5km analysis).

Table 5.0.2 Regression of NTL on net inflows of persons in settlement, 2km and 5km radius

	Y = NTL (x-km radius)			
	x= 2Km		x= 5km	
	(1)	(2)	(3)	(4)
VARIABLES	Linear	Quadratic	Linear	Quadratic
α_1; Lagged NTL	0.878***	0.839***	0.816***	0.916***
	(0.00591)	(0.0212)	(0.00741)	(0.0260)
α_4; (Lagged NTL)^2		0.0320*		-0.155***
		(0.0184)		(0.0242)
α_2; Total Inflow	4.90e-07***	-3.44e-06***	-4.42e-08	-5.10e-06***
	(8.45e-08)	(6.25e-07)	(7.70e-08)	(4.37e-07)
α_3; (Lagged NTL)*Total Inflow	-7.66e-07***	1.97e-05***	6.86e-07***	1.87e-05***
	(1.66e-07)	(2.79e-06)	(1.29e-07)	(1.63e-06)
α_5; (Lagged NTL)^2* Total Inflow		-2.37e-05***		-5.75e-06***
		(2.92e-06)		(1.35e-06)
α_6; (Total Inflow)^2		9.40e-12 ***		1.88e-11***
		(3.46e-12)		(2.49e-12)
α_7; (Lagged NTL)*(Total Inflow)^2		-5.95e-11***		-6.05e-11***
		(0)		(9.03e-12)
α_8; (Lagged NTL)^2*(Total Inflow)^2		7.66e-11***		0
		(0)		(0)
α_9; Population in 2017	3.70e-06***	3.77e-06***	9.56e-06***	7.15e-06***
	(4.18e-07)	(4.29e-07)	(6.58e-07)	(6.42e-07)
α_{10}; % of agriculture in 2020	-3.25e-05	-3.06e-05	6.02e-05*	0.000108***
	(2.38e-05)	(2.38e-05)	(3.47e-05)	(3.35e-05)
α_0; Constant	-0.0240***	-0.0125*	-0.00620	-0.0233***
	(0.00397)	(0.00669)	(0.00525)	(0.00811)
Observations	8,497	8,497	8,497	8,497
Number of settlements	293	293	293	293
District Fixed Effect	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes

Standard errors in parentheses

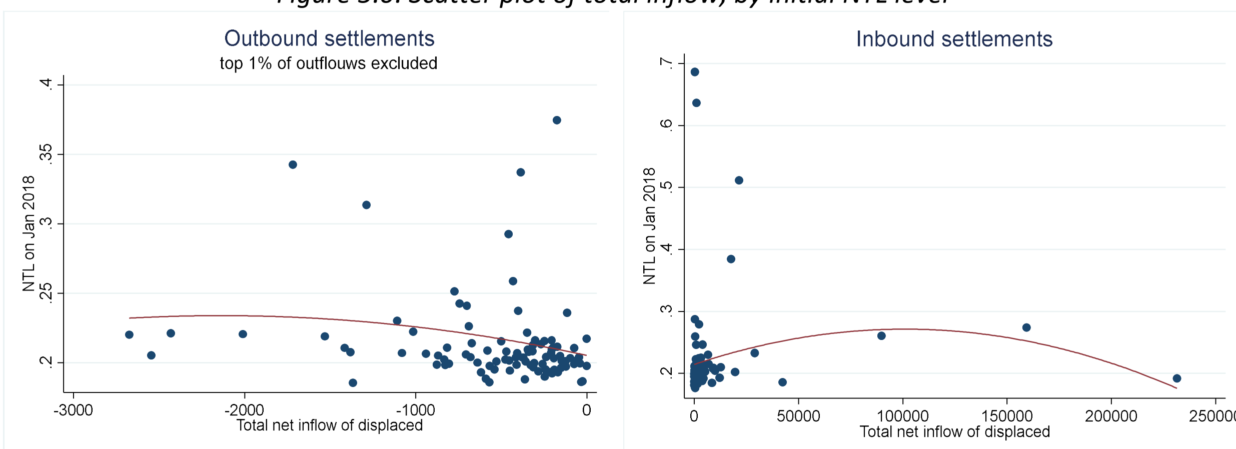
*** p<0.01, ** p<0.05, * p<0.1

5.2. Does displacement affect NTL differently based on initial urbanization levels?

This section attempts to study even further how displacement flows may impact NTL growth differently based on the initial level of human activity and urbanization, e.g., depending on whether they will have to set up a camp or can settle in an already existing urban center. While the initial regressions attempted to control for this by including initial population level (in 2017), this section goes further by allowing the relationship between NTL and displacement to depend on initial NTL levels (as of January 1st, 2018), or by replacing the main independent variable to account for displacement inflows as a share of the existing population.

Note: There is however no clear evidence of a correlation between displacement flows and initial NTL levels. No correlation was observed between total displacement outflows and initial NTL levels, and the dynamic observed in inbound settlements seems to be mostly driven by outliers (Figure 5.0).[38]

Figure 5.0. Scatter plot of total inflow, by initial NTL level



Source: Authors' computations.

The impact of displacement flows on NTL growth might depend on the relative size of the settlement, we hence test for the robustness of our result when changing the main independent variable to account for net inflow of displaced population as a *share* of total population in population. Displaced households settling into a new settlement might have a very different effect on NTL growth (through modification in urbanization and human activity), depending on whether they have to build a new settlement or settle in an already urbanized area. In particular, the speed of infrastructure construction and access to additional services is likely to differ, hence resulting in different NTL growth.

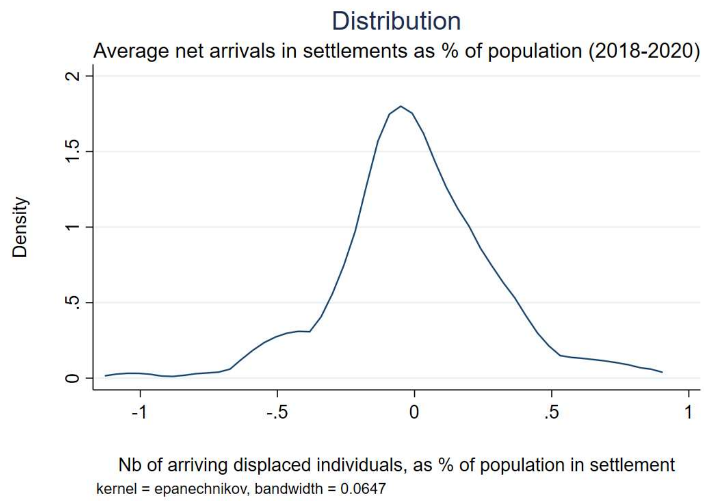
Variables

For the sake of robustness, the analysis will therefore be completed by replacing the independent variable with two alternative measures of net inflow of persons as a proportion of settlement population. First, the variable *Average Inflow Share* accounts for the average share of displaced population hosted in a given period (i.e., the average net inflow of displaced persons hosted in a settlement as a percentage of the total population on a given period). Figure 5.1 shows the density of the *Average Inflow Share* is variable across settlements in Badghis, while summary statistics can be observed

[38] The dynamic remains unchanged when including the top 1% outflows, i.e., settlements with more than -5000 net total inflow (more than 5000 individuals having fled the settlement – in net).

in Table 5.1. Second, the variable *Total Inflow Share* is constructed as the sum of the net inflow of displaced persons divided by the cumulated population in settlement (both summed over each observation available for 2018-2020). Figure 5.2 shows the density of the *Total Inflow Share* is variable across settlements in Badghis, while summary statistics can be observed in Table 5.2

Figure 5.1 Distribution of average net inflow of displaced persons in settlements (as share of population)



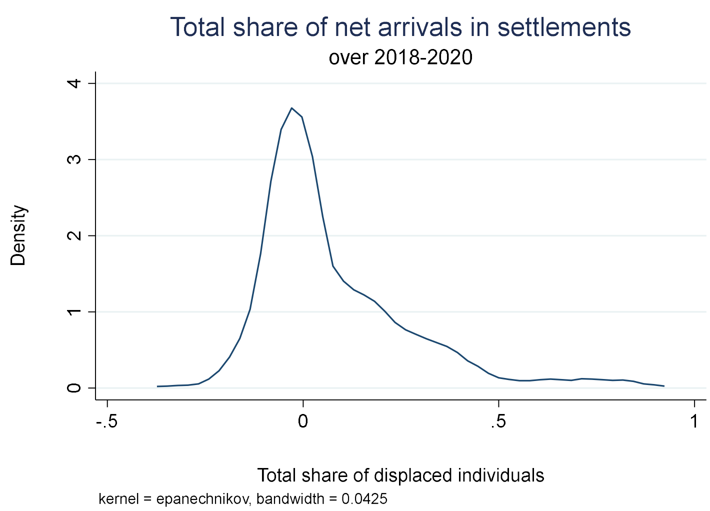
Source: Authors' computation based on IOM DTM 2020.

Table 5.1. *Summary statistics for the cumulative net inflow in Badghis settlements*

Variable	Nb settlements	Mean	Std. Dev.	Min	Max
Average Inflow Share	294	.0018038	.280172	-1.065	.8400924

Note: The net inflow of person can be negative if the population that migrate out of the settlement (either internationally or within Afghanistan) is higher than the number of displaced persons settling in.

Figure 5.2. *Distribution of the total net inflow of persons in settlements (as % of population)*



Source: Authors' computation based on IOM DTM 2020.

Table 5.2. Summary statistics for the cumulative net inflow as a share of cumulated population

Variable	Nb settlements	Mean	Std. Dev.	Min	Max
$Total\ Inflow$ Share	294	0.079	0.194	-0.331	0.881

Source: Authors' computation based on IOM DTM 2020.

Results

There is a risk that NTL growth depends on the size of displaced population relative to the total settlement population, we therefore test the robustness of our results by first accounting for the average net influx of displaced person, as a share of settlement population. That is, we first replace our main independent variable *Total Inflow* by the variable $Average\ Inflow\ Pop$, representing the cumulative net inflow of persons as a share of population in settlement (see Section 4.1 for descriptive statistics). The linear regression following equation-4 is are displayed in column 1 of Table 5.3, the quadratic regression (equation-7) in column 2. Then, results are replicated by using the *Total Inflow Share* (rather than the average), in columns 3 and 4.

The linear regression is robust to the inclusion of the alternative variables; a marginal increase in net relative inflow of displaced population (relative to settlement population) do increase the NTL, as long as past NTL level is not too small. Assuming a linear relationship, equation-6 is positive as long as past NTL levels are larger than 0.24, see column (1) of Table 5.3. An increase in the *average* net inflow of displaced persons (as a proportion of population) raises the NTL level as long as past NTL level is not too small. Similarly, an increase in the *total* net inflow of displaced persons (as a proportion of cumulated population) raises the NTL level (equation-6 is positive) as long as past NTL levels are larger than 0.22. As a reference, all IOM settlements in Badghis recorded an NTL value above both thresholds at least one during 2018-2020. [39]

Table 5.3. *Regression of NTL on average and total share of displaced persons as share of population*

	Y = NTL levels			
	X= Average Inflow Share		X = Total Inflow Share	
	(1)	(2)	(3)	(4)
VARIABLES	Linear	Quadratic	Linear	Quadratic
α_1; Lagged NTL	0.674***	1.214***	0.646***	1.211***
	(0.00890)	(0.0222)	(0.00961)	(0.0226)
α_4; (Lagged NTL)^2		-0.385***		-0.426***
		(0.0143)		(0.0154)
α_2; Total Inflow	-0.0799***	0.0179	-0.0772***	0.0765**
	(0.00613)	(0.0123)	(0.00822)	(0.0309)
α_3; (Lagged NTL)*Total Inflow	0.337***	-0.254***	0.363***	-0.590***
	(0.0173)	(0.0593)	(0.0185)	(0.136)
α_5; (Lagged NTL)^2* Total Inflow		0.742***		1.188***
		(0.0717)		(0.121)
α_6; (Total Inflow)^2		-0.0796***		-0.184***
		(0.0217)		(0.0482)
α_7; (Lagged NTL)*(Total Inflow)^2		0.527***		1.047***
		(0.102)		(0.195)

[39] NTL levels around Badghis settlement evolved between 0.03 and 2.215 during the 2028-2020 period (see Table 3.3).

α_8; (Lagged NTL)²*(Total Inflow)²		-0.736***		-1.285***
		(0.115)		(0.159)
α_9; Population in 2017	8.80e-06***	5.17e-06***	8.68e-06***	4.97e-06***
	(7.66e-07)	(7.47e-07)	(7.66e-07)	(7.43e-07)
α_{10}; % agriculture in 2020	8.28e-05**	0.000108***	7.05e-05*	0.000108***
	(4.07e-05)	(3.91e-05)	(4.08e-05)	(3.92e-05)
α_0; Constant	0.0561***	-0.106***	0.0638***	-0.101***
	(0.00633)	(0.00852)	(0.00647)	(0.00857)
Observations	8,497	8,497	8,497	8,497
Number of settlements	293	293	293	293
District Fixed effect	Yes	Yes	Yes	Yes
Time Fixed effects	Yes	Yes	Yes	Yes

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The robustness analysis using quadratic regressions on both NTL and displacement measures yields mixed results, column 2 and 4 of Table 5.3.

There are several cases whereby a marginal increase in the *Average Inflow Share* (relative to cumulated population) increases the NTL levels, e.g. in inbound settlements with large average net inflow and past NTL levels within a medium-low range, or in important outbound settlements with both low and high past levels of NTL.[40]

- In inbound settlements, a sufficient condition for equation-8 to be positive is that past NTL level lies between 0.35 and 0.50. As a reference, this corresponds to around 20% of observations in Badghis. Another sufficient condition is that past NTL ranged between [0.22;0.50] and that the average inflow share be sufficiently large.
- **An increase in the *Average Inflow Share* (average net inflow of displaced relative to settlement population) will reduce the NTL as soon as the past NTL and *Average Inflow Share* are high enough.** A sufficient condition for equation-8 to be negative is that the *Average Inflow Share* (X_i) is high enough and the past NTL level is higher than 0.5. Around one quarter of Badghis settlements reached that NTL threshold at least once during the 2018-2019 period.
- **In settlements that serve as important outbound areas, a reduction in the average outflow share (increase in the average inflow share) would increase the NTL both in areas that have very low or medium-high level of human activity.** In outbound areas, a sufficient condition for a marginal increase in the *Average Inflow Share* to increase the NTL is that the average outflow is large enough, in areas with low past NTL levels (below 0.22) or high NTL levels (larger than 0.5).

There are several cases whereby a marginal increase in the total net inflow (relative to cumulated population) increases the NTL levels, e.g., in already urbanized outbound areas with high NTL, and in important inbound areas with medium human activity levels.

- In inbound areas, a sufficient condition for equation-8 to be positive is that NTL in [0.26,0.56] in inbound area. It is negative if past NTL levels are larger than 0.55 and inflows are large enough.

[40] When $X_i = \text{Average Inflow Share}$, equation-8 becomes $\frac{\mathrm{d\,NTL_{i,t}}}{\mathrm{d}X_i} = \frac{\mathrm{d\,NTL_{i,t}}}{\mathrm{d}X_i} = -0.254.NTL_{i,t-1} + 0.736.NTL_{i,t-1}^2 + 2.\mathrm{X_i}(-0.0796 + 0.527.NTL_{i,t-1} - 0.736.NTL_{i,t-1}^2)$. The term between brackets is positive when $NTL_{i,t-1}$ ∈[0.22;0.50]. In addition, $-0.254.NTL_{i,t-1} + 0.736.NTL_{i,t-1}^2 > 0$ when $NTL_{i,t-1} > 0.35$.

That is, a marginal increase in the total inflow of displaced persons as share of cumulated population would decrease the NTL in important inbound areas that already have medium-high level of human activity.[41]

- As before, low past NTL value (below 0.26) or medium-high (above 0.56) associated with a large negative *Total Inflow Share* (important outbound) also results in equation-8 being positive. That is, in urbanized settlements from which people flee massively, a marginal decrease in the number of people leaving the settlement (i.e., a net increase in Total Inflow Share) will be associated with an increase in NTL levels.

6. Can we use NTL growth to predict displacement?

Having established a positive correlation between NTL growth and displacement flows we aim to test whether this allows us to predict migration patterns, i.e., whether a settlement mostly served as an inbound or outbound area First, we create the binary variable *Inbound,* which equals 1 if the settlement was a net receiver of displaced persons (*Total Inflow* $\geq$0) over the 2018-2020 period. The *Inbound variable* takes value 0 if it is an outbound area (*Total Inflow* <0). Out of the 294 settlements, 149 were net receivers (50.7%), and 145 served as outbound areas from which people mostly left (49.3%).

We then run a set of regression to see whether cumulated displacement inflows over the 2018-2020 period can be predicted using the average NTL levels and average growth rate in settlement. For a settlement *i* in district *j*, we are studying the following regression, of the inflow variable *X* (namely *Total Inflow, Average Inflow Share* or *Total Inflow Share*), on the average NTL level, and the consecutive increase of NTL, the 2017 initial population level, the share of income derived from agriculture (in 2020), and the district fixed effect δ_j.

Let's define the outcome variable Y= { *Total Inflow, Average Inflow Share* or *Total Inflow Share*} , and the regressors $\Delta NTL_{i,j,t-1} = NTL_{i,j,t} - NTL_{i,j,t-1}$, with *mean(*$NTL_{i,j}$*)* the average $NTL_{i,j}$ across all periods between January 2018-January 2020. We estimate the following regressions

$$Y_{i,j} = \beta_0 + \beta_1 mean(NTL_{i,j}) + \beta_2 \mathrm{mean}(\Delta NTL_{i,j,}) + \beta_3 mean(\Delta NTL_{i,j,})^2 + \beta_4 Pop2017_{i,j} + \beta_5 Agric2020 + \delta_j + u_{i,j}$$

(equation-10)

Results are displayed in Table 6.1, and show that the average NTL is a good predictor of the total inflow of displaced population across 2018-2020. A marginal increase in the average NTL by one unit would raise by 129,797 persons the cumulated inflow over the period (coefficient significant at 99%). This relationship seems to be concave, at 95% confidence interval.

Table *6.1. Regression of Total inflow on average NTL levels and growth*

	(1)
VARIABLES	**Y = Total Inflow**

[41] $\frac{\mathrm{d\,NTL_{i,t}}}{\mathrm{d}X_i} = 0.0765 - 0.59. NTL_{i,t-1} + 1.188. NTL^2_{i,t-1} + 2.\mathrm{X_i}(-0.184 + 1.047. NTL_{i,t-1} - 1.285. NTL^2_{i,t-1})$
= h(NTL) + 2Xf(NTL) with f(NTL)>0 if NTL in [0.26;0.56] and h(NTL) >0 for all NTL.

mean(NTL)	129,797***
	(46,556)
$[\text{mean(NTL)}]^2$	-86,305**
	(35,639)
mean(Δ NTL)	231,355
	(161,720)
Population 2017	-0.940
	(1.112)
% agriculture 2020	-176.5***
	(54.32)
Constant	-18,024
	(11,402)
Observations	293
R-squared	0.238
District Fixed Effect	Yes

Standard errors in parenthesis
*** p<0.01, ** p<0.05, * p<0.1

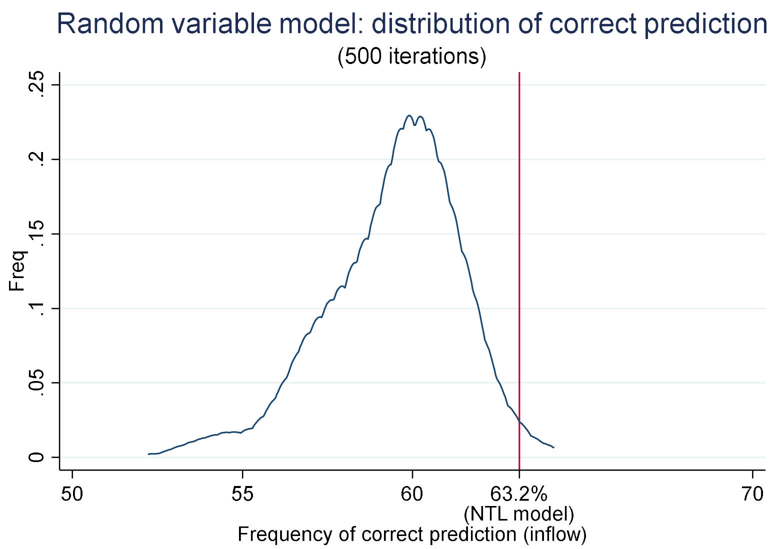
The model yields promising results in terms of using NTL data to predict whether a settlement is a net receiver or sender of displaced persons over the 2018-2020, as it correctly classifies 63.2% of settlements. We then use a reverse methodology to predict whether a settlement is an Inbound area using coefficients obtained in Table 6.1, by constructing the variable *Predicted Inbound,* which equals 1 if the predicted values for *Total Inflow* are positive, and 0 if they are negative. We then compare the actual *Inbound* classification with the predicted outcome. If the null hypothesis is H0: outbound settlement, the model has a false positive rate of 40 percent and a false negative rate of 34 percent. Overall, 63.2 percent of settlements are correctly classified. Given limited accessibility, this level of accuracy can inform preliminary assessments and resource mobilization prior, which can then be confirmed with on-the-ground validation when feasible.

Table 6.2. Inference Predicted Inbound using NTL variations (column 1, Table 6.1)

		Actual *Inbound*		
		0 = Outbound	1 = Inbound	Total
Predicted Inbound	0=Outbound	True Negative 87 (1-α=.60)	False Negative 51 (β=.34)	138
	1=Inbound	False Positive 57 (α=.40)	True Positive 98 (1-β=.66)	155
	Total	144	149	293

Using NTL to predict displacement flows (NTL model) performs better than a similar analysis using random variable as main regression (Random model, over 500 iterations). To test the robustness of our results, we run the same reverse specification and replace the NTL regressors by a random variable that follows a normal law. This process is iterated 500 tomes. As expected, the random variables regressors are rarely significant to determine the *Total Inflow* variable (equation-10): out of the 500 iterations, the estimated β_1 β_2 and β_3 return a p-value value smaller or equal to 0.01 in only 1% of the iteration respectively. Moreover, the 500 iterations of the Random model return an average probability 59.4% of accurately predicting a settlement into a net receiver (inbound) or outbound of displacement flows, with a standard deviation of 1.96. As seen previously, the NTL model returns a probability of accurately categorizing the settlement of 63.2%, which is almost 2 standard-deviation larger than the mean obtained in the random model (Figure 6.1).

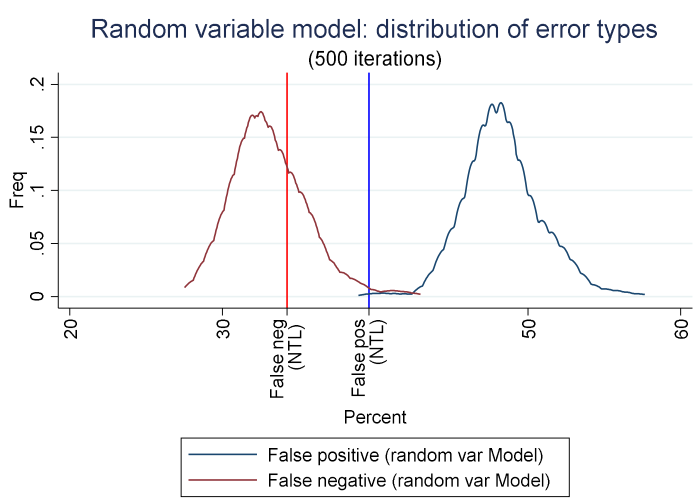
Figure 6.1. Distribution of accurate prediction when using a random variable (500 iteration)



Source: Authors' computation.

In addition, the NTL model mostly always returns a smaller rate of false positive, even though it has a significant probability of returning a slightly higher rate of false negative, (Figure 6.2)**.** The average rate of false positive obtained with the random variables is at 48% (which is 8% point larger than when using NTL), but the rate of false negative is at 33.1%, which is less than 1% lower than the average false negative in the main model.

Figure 6.2. Distribution of false predictions with random variable model (over 500 iterations)



Source: Authors' computation.

Using reverse engineering method hence yields promising results, yet the methodology needs to be improved, which stresses the need to collect high frequency data on displacement and socio-economic outcomes at the disaggregated level.

7. Conclusion

This work investigated whether geospatial data could be used to proxy for displacement flows and settlement expansion, using NTL data. By focusing on Badghis province it provided relevant insight on a province that was particularly affected by recent climatic shocks, including the 2018 drought. We find that large displacement movements can indeed be observed by NTL satellite imagery. Settlements classified as high inflow net inflow of displaced persons over the 2018-2020 timespan (high inflow receivers) have larger NTL growth than others. Results remain robust when looking at the net inflow of persons as a share of the existing population. Accounting for population in 2017 as a proxy for initial levels of urbanization does not change our findings. It is also robust to using the continuous levels of net inflow in given settlement, rather than a binary classification high/low. Finally, the quadratic regressions suggest a lock-in effect, whereby people tend to stay in the place they migrated to.

This work provides a first basis to better understanding the drivers and spatial characteristics of settlement growth. The methodology can be used for investigating the predictability of population displacement based on variations in geospatial data (e.g., shifts in vegetation density due to drought). Several next steps have been identified to push the analysis further, among which include expanding the analysis to other regions. Evidence from the ground suggests that there are regional differences across displacement patterns based on geography, administration, and the like. An important lock-in effect has been witnessed in Herat, for instance, whereby IDPs who settle there do not return to their origin location. Future work would investigate patterns between NTL growth of significant locations country-wide, within the context of recent IOM work on determinant factors to return.

8. Appendix A: variables definitions

8.1. Definitions of variables in IOM dataset.

Returnees are Afghan nationals who have returned to Afghanistan in the assessed location after having spent at least six months abroad. This includes both documented returnees (Afghans who were registered refugees in host countries and requested voluntary return with UNHCR and relevant national authorities) and undocumented returnees (Afghans who returned spontaneously or were deported from host countries, irrespective of whether or not they were registered refugees with UNHCR and relevant national authorities).

Arrival IDPs are Afghans who fled from other settlements in Afghanistan and have arrived and presently reside at the assessed location - host community, as a result of, or in order to avoid, the effects of armed conflict, generalized violence, human rights violations, protection concerns, or natural and human-made disasters.

Total Inflow = *Returnees + IDPs*

Returned IDPs are Afghans who have returned to their home or place of origin in the assessed location or settlement from which they had fled as IDPs in the past, as a result of, or in order to avoid, the effects of armed conflict, generalized violence, human rights violations, protection concerns, or natural and humanmade disasters.

Fled IDPs are Afghans who have fled from an assessed location or settlement within which they previously resided and now currently reside in a different settlement in Afghanistan, as a result of, or in order to avoid, the effects of armed conflict, generalized violence, human rights violations, protection concerns, or natural and human-made disasters.

8.2. Construction of our variables

We construct the variable *Total Inflow* as the cumulative net inflow (labeled *net_inflow*) at final date (December 2020). For each settlement *i,* the variable is hence defined

$$\text{Total Inflow}_i = \sum_{t=2018}^{2020} \text{net\_inflow}_i$$

with *net_inflow*$_i$= *number of returnees from abroad*$_i$ + *number of returning IDPs*$_i$ + *number of new IDPs in settlement*$_i$ - *(Out Migrants*$_i$ + *Fled IDPs*$_i$*)*

9. Appendix B: Geospatial Data and Methodology

The geospatial methodology used for this analysis extracts monthly descriptive statistics of Nighttime Lights data to individual buffers around IOM settlement locations. This process is known in Geographic Information Systems (GIS) as "zonal statistics", and is a commonly used method for collecting raster, or gridded data (e.g., NTL, precipitation) into zones or areas of interest to the user (e.g., buffers, administrative districts). This general approach was used in Google Earth Engine, a free cloud computing and data repository, which enables large-scale, satellite imagery and other analyses, including the extraction of zonal statistics.[48]

For the purposes of exploring drought-driven displacement, we kept the analysis constrained to Badghis Province in Northwestern Afghanistan, but the same approach is easily replicated in other regions or even country wide. The input data used in this analysis were the UN IOM DTM dataset from 2020, and the Nighttime Lights data were monthly composites from January 2018 to December 2020.[42]

To replicate the methodology, one may open the Google Earth Engine script at the universal link[43] or download the script directly from GitHub.[44] In the GitHub repository there is a folder titled "NTL," which contains the monthly NTL script that can be extracted to the buffers, as well as a script for extracting a monthly average of NTL data for the entire area of interest. The user then may update the boundaries of the analysis by simply uploading a new dataset and preserving the object name 'table' so that the script will apply the same processes on the new area of interest. The user would also update the export location to ensure that the file saves to a location on their personal Google Drive account (or alter the

---

[42] Due to reflectance issues from collection, NTL from June are unavailable for most of Afghanistan. As such, there was no NTL data used for the month of June 2018-2020.

[43] Belanger, J. Google Earth Engine Script. Retrieved from: https://code.earthengine.google.com/4b069482a54f1fb6041d2e2cff25834d

[44] Stewart, B., Chamorro, A.E., Martine, L. and Belanger, J. World Bank - GEE_Zonal. GitHub Repository. Retrieved from: https://github.com/worldbank/GEE_Zonal

code to save it as an Earth Engine 'asset', or dataset. The outputs of the methodology are a panel dataset in the form of a CSV (unless otherwise specified in the code) and can be exported from Drive onto one's local machine for further statistical analysis.

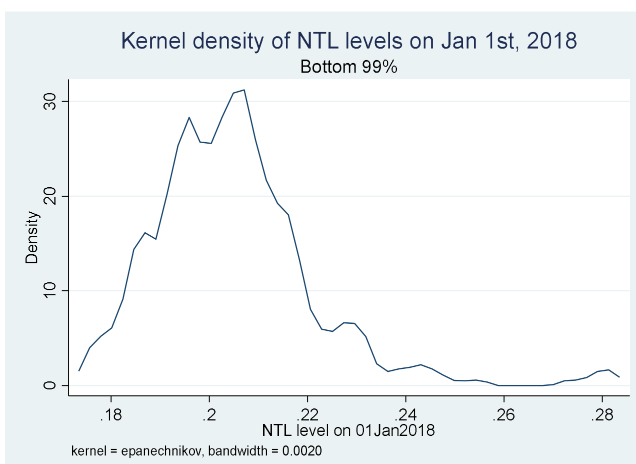
10. Appendix C: Differences based on initial NTL levels

Using a Jenks decomposition to identify natural breaks in the distribution of settlements' initial NTL level as of January 1st, 2018, settlements are classified into two groups (high or low initial NTL) to measure the urbanization starting point. On average, settlements had an initial NTL level of 0.217, with a minimum of 0.175 and a maximum of 1.058. A Jenks decomposition is used to identify two groups of settlement (higher or lower initial NTL), based on the distribution of NTL levels (on Jan 1st, 2018) throughout the Badghis settlement. The Jenks method allows to create two groups such that the variance within a group is minimized, while the variance between groups is maximized. The categorization is such that 289 settlements are associated with the low NTL group, with a minimum of 0.175 and a maximum of 0.414. On the other hand, 5 settlements are classified into the higher NTL initial level, with a minimum NTL of 0.474 and a maximum of 1.058.

Table C.1. Summary statistics of NTL levels as of Jan 1st, 2018

	Number of settlements	Mean	Std. Dev.	Min	Max
Summary statistics for whole sample					
Sample: 294 settlements	294	0.217	0.074	0.175	1.058
Summary statistics by group					
Sample: Group "Lower NTL"	289	0.209	0.030	0.175	0.414
Sample: Group "Higher NTL"	5	0.694	0.232	0.474	1.058

Figure C.1. Kernel density of initial NTL level (Jan 1st, 2018) – top 1% excluded.



Source: Authors' computations

The impact that displacement flows may have on the NTL growth does depend on initial level of NTL, as urbanized areas do not seem to be impacted by inflows of displaced. We run a linear and quadratic regression following equation-4 and equation-7, on two different sample: the 145 observations with high initial NTL levels (5 settlements across 2018-2020), and the 8381 observations with lower initial NTL levels (289 settlements).

The linear and quadratic regression show that larger displacement inflows are always correlated with larger NTL growth in settlement with higher initial NTL levels, at, 95% confidence interval level, see columns 3 and 4 of Table C.2.

The linear and quadratic regressions analysis on the settlements starting with lower initial NTL levels return mixed results. Table C.2. column (1) show that equation-5 is positive, i.e., the NTL growth increases with the Total Inflow, as the interaction term is positive. The linear regression suggests that a marginal increase in net inflows of persons would increase the NTL in places with medium-high levels of human activities, i.e., with NTL above 0.30. The quadratic regression displays interesting patterns:

- **In outbound settlements, a reduction in the net outflows of displaced persons raises the NTL level in outbound areas with important human activity.** A sufficient condition for a marginal decrease in outflows (i.e., an increase in Total Inflow) to be associated with an increase in NTL levels is that past NTL be larger than 0.9.[45]
- **In inbound areas, a marginal increase in net inflows of persons would increase the NTL in places with small levels of human activities,** i.e., with NTL between 0 and 0.29.

Table C.2. Regression of NTL on net inflows of persons in settlement on two different samples (higher vs lower initial NTL levels)

	Y = NTL levels			
	Sample: Small initial NTL		Sample: Large initial NTL	
	(1)	(2)	(3)	(4)
VARIABLES	Linear	Quadratic	Linear	Quadratic
α_1; Lagged NTL	0.563***	0.917***	0.570***	1.201***
	(0.00947)	(0.0210)	(0.0911)	(0.297)
α_4; (Lagged NTL)^2		-0.393***		-0.392**
		(0.0147)		(0.164)
α_2; Total Inflow	-9.09e-07***	-4.09e-06***	1.55e-05**	-1.55e-05
	(7.78e-08)	(3.87e-07)	(6.09e-06)	(0.000128)
α_3; (Lagged NTL)*Total Inflow	3.01e-06***	1.34e-05***	4.04e-06	0.000347
	(1.67e-07)	(1.40e-06)	(5.61e-06)	(0.000285)
α_5; (Lagged NTL)^2* Total Inflow		3.44e-06***		-9.81e-05
		(9.64e-07)		(0.000161)
α_6; (Total Inflow)^2		1.39e^{-11}***		-7.17e-10
		(0)		(6.01e-09)
α_7; (Lagged NTL)*(Total Inflow)^2		-3.20e-11***		-1.32e-08

[45] $\frac{\mathrm{d\,NTL_{i,t}}}{\mathrm{d}X_i} = -4.09e^{-06} + 1.34e^{-0}\ NTL_{i,t-1} + 3.44e^{-06}NTL_{i,t-1}^2 + 2.\mathrm{X_i}(1.39\mathrm{e}^{-11} - 3.20\mathrm{e}^{-11}NTL_{i,t-1} - 5.30e^{-11}NTL_{i,t-1}^2) \equiv$ h(NTL) + 2e^{-11}Xf(NTL) with f(NTL) positive if NTL in [0,0.29] and h(NTL) positive if NTL>0.9.

		(0)		(1.29e-08)
α_8; (Lagged NTL)^2*(Total Inflow)^2		-5.30e-11***		3.55e-09
		(0)		(7.24e-09)
α_9; Population in 2017	1.06e-05***	7.21e-06***	7.83e-05**	-1.92e-05
	(6.93e-07)	(6.44e-07)	(3.28e-05)	(2.04e-05)
α_{10}; % agriculture in 2020	-6.40e-05*	-1.85e-05	0.00769***	0.00641***
	(3.40e-05)	(3.11e-05)	(0.00270)	(0.00207)
α_0; Constant	0.109***	-5.30e-11***	-0.812**	0
	(0.00596)	(0)	(0.403)	(0)
Observations	8,352	8,381	145	145
Number of settlements	288	289	5	5
District Fixed effect	Yes	Yes	Yes	Yes
Time Fixed effects	Yes	Yes	Yes	Yes

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

11. Appendix D. Alternative data source

IOM's CBNA collects information on arrival IDPs twice a year based on key informant interviews at the level of settlements, which are aggregated into districts and regions. In 2020, new displacements were reported across 1,069 settlements in 7 districts of the Badghis province. In CBNA, respondents are asked to indicate the percentages of IDPs that arrived at a settlement due to conflict and natural disasters; in the dataset, displacements are grouped into four categories, such as induced by conflict, conflict-natural disasters, natural disasters, and natural disasters-conflict.

OCHA's dataset includes information on conflict-induced displacements at the level of districts, collected through the alert system; additionally, OCHA reports the dates of displacements. In 2020, there were 66 incidents (dates) of displacement due to conflict, and those IDPs arrived in 2 districts of the Badghis province.

Comparisons of IOM and OCHA data on new displacements were performed for the districts reported by both organizations. IOM reported more conflict-induced displacements than OCHA in the Ab Kamari district but no settlements with only conflict-induced arrival IDPs in the Qala-e-Naw district. At the same time, the IOM's numbers of IDPs due to conflict and conflict-natural disasters exceed the OCHA's numbers of conflict-induced IDPs in both districts. Effective comparisons of IOM and OCHA data on displacements may not be conducted due to differences in their data collection methodologies.

Period of comparison

The IOM's CBNA contains data on new displacements in Afghanistan during 2018-2020; the dataset that we currently have includes 4 periods: January-December 2018, January-June 2019, January-June 2020, and January-December 2020. The OCHA data on conflict-induced displacements are available for 2016-2020.

Table A1. New displacements in the Badghis region reported by IOM and OCHA.

Organization	2016	2017	2018	2019	2020

				Jan-June 2019	July-Dec 2019	Jan-June 2020	July-Dec 2020
IOM	-	-	168197[46]	32983	-	19747	24819
OCHA	15157	27361	9918	24113	2053	1810	6072

A comparative analysis of IOM and OCHA data on IDPs in the Badghis region was performed for January-December 2020, except for correlation analysis, which was done for 2019-2020.

Description of IOM data on displacement

IOM's CBNA provides information on internally displaced people (IDPs) twice a year based on key informant interviews at the levels of settlements, districts, and provinces in Afghanistan. In particular, for the period of January-December 2020, the data are available for 1069 settlements and 7 districts of the Badghis region. Summary statistics for new displacements in 2020 are presented in Table 2. The largest numbers of IDPs arrived in Bala Murghab (22,491 new IDPs; on average, 91 people in each of 247 settlements) and Qala-e-Naw (12,289 new IDPs; on average, 90 persons in 136 settlements) districts of Badghis province.

Table A2. The summary statistics for Arrival IDPs (new displacements) in Badghis province in 2020.[47]

Districts	**Mean**	**Std. Dev.**	**Freq. (# of settlements)**
Ab Kamari	7.5846774	25.656578	248
Bala Murghab	91.05668	283.35422	247
Ghormach	13.2	39.750008	20
Jawand	16.623077	65.459647	130
Muqur	21.822485	46.781577	169
Qadis	15.058824	31.127211	119
Qala-e-Naw	90.360294	204.55805	136
Total	41.689429	162.17494	**1,069**

In 2020, 44,566 IDPs arrived in the Badghis region due to various reasons. Table A3 shows the breakdown of numbers of IDPs that arrived due to conflict, natural disasters, and their combinations[48]. It additionally specifies the numbers of IDPs due to conflict (9,504) and due to conflict and conflict-natural disasters (30,571).

Table A3. IDPs arrived in the Badghis region in 2020.

	Arrival IDPs due to:		
	conflict, conflict-natural disasters, natural disasters, and	conflict and conflicts-natural disasters	conflict

[46] The number of ArrivalIDPs2018 in IOM's 2018 CBNA is too large and most likely to include displacements in previous periods (not only IDPs that arrived in 2018). This number cannot be compared with OCHA's number.

[47] Stata command: .tab ADM2NameEnglish, sum (ArrivalIDPs2020).

[48]The IOM questionnaire asks to indicate the percentages of IDPs that arrived at a settlement due to conflict and natural disasters, which then in dataset are grouped into 4 categories: conflict, conflict-natural disasters, natural disasters, and natural disasters-conflict.

	natural disasters-conflict		
Districts			
Ab Kamari	1,881	1,503	154
Bala Murghab	22,491	16,715	9,005
Ghormach	264	257	84
Jawand	2,161	189	0
Muqur	3,688	2,207	240
Qadis	1,792	1,190	21
Qala-e-Naw	12,289	8,510	0
Total	**44,566**	**30,571**	**9,504**

Description of OCHA data on displacement

OCHA reports conflict-induced displacements for districts and provinces of Afghanistan. OCHA dataset additionally includes dates of displacement. According to OCHA, in 2020, there were 66 incidents (dates) of displacement in the Badghis region (Table A4). Table A4 presents summary statistics for conflict-induced IDPs based on displacement incidents (not comparable to IOM data because IOM's summary statistics is for settlements).

Table A4. The summary statistics for conflict-induced IDPs in 2020.

Districts	**Mean**	**Std. Dev.**	**Freq. (# of displacement incidents)**
Ab Kamari	20	0	1
Qala-e-Naw	120.95385	93.139417	65
Total	119.42424	93.251861	**66**

As reported by OCHA, in 2020, conflict-induced IDPs settled in two districts of the Badghis province Qala-e-Naw (7,882 people) and Ab Kamari (20 people) (Table A5).

Table A5. Conflict-induced IDPs settled in the Badghis region in 2020.

Districts	**Displaced individuals (or arrival IDPs)**
Ab Kamari	20
Qala-e-Naw	7,862
Total	**7,882**

Comparison of IOM and OCHA data

IOM's CBNA reports arrival IDPs for settlements (in 2020, 1,069 settlements in the Badghis region), which are then aggregated into districts (7 districts in the region). OCHA's data only contains information for districts; in 2020, conflict-induced IDPs arrived in two districts of the Badghis province.

Table A6 compares displacements reported by OCHA to those recorded by IOM for two districts that are included in the OCHA's dataset. For example, OCHA reported the arrival of 7,862 IDPs due to conflict in Qala-e-Naw district. According to IOM, settlements in this district reported 9,350 IDPs that arrived due to conflict and conflict-natural disasters (but 0 IDPs if to look at settlements that reported the arrival of IDPs due to conflict only). This hints that OCHA data may also include displacement due to natural disasters along with those due to conflict, even though the former is not specified.

Table A6. Comparison of OCHA and IOM’s data on displacements in the Badghis region in 2020.

	OCHA	IOM		
Reason for displacement **Districts**	conflict	conflict; conflict-natural disasters; natural disasters; natural disasters-conflicts	conflict; conflict-natural disasters	conflict
Ab Kamari	20	1,881	1,503	154
Qala-e-Naw	7,862	12,289	8,510	0
Other districts	0	30,396	20,558	9,350
Total	**7,882**	**44,566**	**30,571**	**9,504**

Correlation analysis of OCHA and IOM’s data on internal displacement was conducted based on a limited number of comparable observations (districts) (Table A7). Pearson’s correlation coefficient is 0.8162 ($p<0.01$), which, however, may not be interpreted because of too few observations.

Table A7. Arrival IDPs in 2019-2020 (used for correlation analysis)

Periods	**Districts**	**IOM**	**OCHA**
Jan-June 2019	Ab Kamari	1071	0
	Bala Murghab	20445	17335
	Ghormach	273	0
	Jawand	1127	0
	Muqur	1288	0
	Qadis	1503	0
	Qala-e-Naw	11432	6778
Jan-June 2020	Ab Kamari	972	20
	Bala Murghab	11011	0
	Ghormach	0	0
	Jawand	700	0
	Muqur	1686	0
	Qadis	763	0
	Qala-e-Naw	4615	1790
July-Dec 2020	Ab Kamari	909	0
	Bala Murghab	11480	0
	Ghormach	264	0
	Jawand	1461	0
	Muqur	2002	0
	Qadis	1029	0
	Qala-e-Naw	7674	6072

Changes in the numbers of Arrival IDPs (stock of IDPs) in 2018-2020

In 2020, the number of IDPs (protracted IDPs) in the Badghis region decreased by 24,327 compared to 2018 (Table A8). In particular, this effect was achieved due to IDPs leaving Qala-e-Naw (-67,888) and Ab Kamari (-693) districts. In other districts, the numbers of IDPs increased as compared with those in 2018. These numbers should be interpreted with caution as the government of Afghanistan disputes IOM's numbers of protracted IDPs.

Table A8. Changes in the numbers of IDPs in 2020 vs. 2018 according to IOM data

Districts	2018	2020	Absolute change	Percentage change
Ab Kamari	7,806	7,113	-693	-8.9
Bala Murghab	9,985	44,118	34,133	341.8
Ghormach		1,146	1,146	-
Jawand	3,319	6,138	2,819	84.9
Muqur	7,203	12,319	5,116	71.0
Qadis	8,069	9,109	1,040	12.9
Qala-e-Naw	233,900	166,012	-67,888	-29.0
Total	**270,282**	**245,955**	**-24,327**	**-9.0**