Weather and Conflicts in Afghanistan — Appendix

A Additional Data Details **2**

A.1 Data Sources 2

A.2 District Assignment of UCDP GED Events 3

A.3 Area-Weighting 4

A.4 Data Processing 5

B Mathematical Appendix **8**

C Additional Descriptive Statistics **12**

D Additional Robustness Checks **14**

E Estimation Method **29**

F References of Appendix **30**

A Additional Data Details

A.1 Data Sources

A shapefile from the Empirical Studies of Conflict Project (ESOC) define the division of Afghanistan into administrative units and their respective boundaries. The shapefile is accessible at https://esoc.princeton.edu/files/administrative-boundaries-398-districts.

The dataset on conflict events is the UCDP GED Global version 17.2. It is the most recent version of the UCDP GED. The dataset is accessible as a shapefile at http://ucdp.uu.se/downloads/.

Raster data on temperature comes from a product developed by the LIS team at NASA GSFC. The raster data is the NASA LIS version 7 Noah-36. The data is not publicly available, but NASA GSFC provides subsets of the data upon permission. Contact information is available at https://www.nasa.gov/content/contact-goddard.

Raster data on daily precipitation comes from the CHIRPS v2.0 Global Daily netCDF products. The raster data is accessible at the FTP server ftp://ftp.chg.ucsb.edu/pub/org/chg/products/CHIRPS-2.0/global_daily/netcdf/p05/.

To construct my opium suitability index, I used several sources. First, global land cover maps from GlobCover 2009 at http://due.esrin.esa.int/page_globcover.php (Arino et al., 2012). Second, DIVA-GIS-specific world climate data between 1950 and 2000 at 2.5 arcminutes resolution from WorldClim version 1.3 at http://www.diva-gis.org/climate. Third, two river networks over parts of Asia and Southwest Asia from the United States Geological Survey (USGS) mapping product Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales (HydroSHEDS) at https:

//hydrosheds.cr.usgs.gov/datadownload.php?reqdata=15rivs.

The UNODC provide indicative district-level data on opium cultivation. Contact information is available at https://www.unodc.org/unodc/en/about-unodc/contact-us.html. The data is also accessible from the UNODC Afghanistan Opium Surveys 2005-2016 available at https://www.unodc.org/unodc/en/crop-monitoring/index.html?tag=Afghanistan. Information on the typical period for opium planting at the provincial level comes from the UNODC Afghanistan opium surveys (UNODC, 2008; UNODC, 2013) available at https://www.unodc.org/unodc/en/crop-monitoring/index.html?tag=Afghanistan.

A.2 District Assignment of UCDP GED Events

Here I highlight a caveat related to the UCDP GED Global version 17.2. It concerns the fact that I cannot confirm that the district polygons in the ESOC shapefile contain the correct coordinates used in the UCDP assignment of UCDP GED events to districts. Because though the UCDP register the longitude and latitude of a UCDP GED event to the centroid of a district if the event is known to occur at the district-level, but not at a more fine-grained level, they use no, and cannot provide any, shapefile with information on the particular district centroid used. Instead, the UCDP use gazetteers. The gazetteers used for all UCDP GED events is unknown. However, most of them are in the GEOnet Names Server at https://www.nga.mil/ProductsServices/GeographicNames/Pages/default.aspx (Petterson, Therese, and Stina Högblad at Uppsala University, personal communication, March 13 and 20, 2018).

To investigate this further, I noted that the UCDP internally track administrative changes in a data structure called a geotree. Upon contact, the UCDP

provided the geotree for Afghanistan in the form of a table. The table confirms that compared to the ESOC shapefile, from July 2005 and onwards, the UCDP has used a more fine-grained subdivision of Afghanistan for all UCDP GED events and that the geotree register all districts in the ESOC shapefile. Thus, there is suggestive, but not conclusive, evidence that the number of false assignments is low when I assign UCDP GED events to the districts of Afghanistan as defined by the ESOC shapefile.

A.3 Area-Weighting

The area-weighted averages of temperature and precipitation were computed as unweighted arithmetic averages across degree grid cells in the World Geodetic System 1984 (WGS84) Geographic Coordinate System. Formally, let w_{gdk} denote temperature or precipitation in grid cell g in district d day k. Then the area-weighted average is

$$w_{dk} \equiv \frac{\sum_{g \in \mathcal{G}_d} w_{gdk}}{|\mathcal{G}_d|}, \tag{A.1}$$

where $\mathcal{G}_d$ is the set of grid cells whose centroids fall inside that of district d's boundary as defined by a polygon in the ESOC shapefile.

There are two objections to this area-weighting procedure. The first is the following. Since an oblate spheroid can approximate the figure of the Earth, the area of each grid cell varies in longitude and latitude. Therefore it would be sensible to first project the georeferenced weather data to a suitable coordinate system with units in meters.[45] Then one average across meter grid cells of

[45]For Afghanistan—as an Atlas of Earth shows—it would be sensible to project the raster data on the Universal Transverse Mercator (UTM) 41N and 42N zones. This projection is believed to introduce little distortion along both dimensions as it projects small chunks of the surface of the Earth onto a flat surface (Dell, 2009).

equal area. My response is that the method proposed by the objection may introduce more error into the measured weather variables as it involves the use of a resampling algorithm when projecting the data. Furthermore, since districts are tiny, grid cell areas are approximately constant within districts of Afghanistan.

The other objection against my weighting scheme emphasizes that I do not weight by, e.g., population or the area of agricultural land. My response is that such a weighting scheme is unnecessary in this case as I focus on the reduced-form total effect of weather variations on conflict incidence. Other weighting schemes would emphasize some particular aspect. For example, weighting by population size is sensible if we, e.g., have the hypothesis that rising temperature levels can affect people's propensity to do violence against others.[46] Another example is to weight by the agricultural land area, which is more sensible if we focus on some agricultural mechanism that is supposed to explain the weather-conflict relationship. However, since I focus on the overall, rather than any particular, average treatment effect of weather variations on conflict incidence, neither of these weighting schemes is appropriate for my research question.

A.4 Data Processing

Data processing was first carried out using ArcPy. ArcPy is a module in Python 2.7 included in ArcGIS 10. I wrote Python scripts that imported the ArcPy geoprocessing tools to process the georeferenced datasets. The scripts accomplished the following tasks.

- To match UCDP GED events to districts of Afghanistan as defined by the

[46] See Baysan et al. (2015) for a recent discussion on the role of noneconomic psychological and physiological factors in driving the temperature-conflict relationship.

ESOC shapefile of Afghanistan. Specifically, to intersect the UCDP GED and ESOC shapefile.

- To compute the area-weighted average of precipitation for each day and district of Afghanistan. Specifically, to for loop over a folder of the CHIRPS v2.0 daily precipitation datasets and compute the area-weighted average for all districts within each loop.

- To compute the area-weighted average of temperature for each day and district of Afghanistan. Specifically, to iteratively spawn subprocesses, each of which processes four days of NASA LIS version 7 Noah-36 temperature data at a time.[47]

Together with the ArcPy module and DIVA-GIS (Hijmans et al., 2001) I also accomplished the following task:

- To compute the opium suitability index. First, normalize the GlobCover 2009 global land cover map with the normalized values in Table 2 in Kienberger et al. (2017). Second, with DIVA-GIS, compute a normalized climatic suitability index of opium poppy (*Papaver somniferum*) based on tolerable climatic conditions specified in Table 3 in Kienberger et al. (2017). Third, construct a proxy for water availability by computing and normalizing the river density across Afghanistan based on two river networks over parts of Asia and Southwest Asia from the USGS HydroSHEDS. Fourth, compute a weighted arithmetic average based on (Analytical Hierarchy Process) weights in Kienberger et al. (2017) that are suggested by

[47] Due to the high resolution of the temperature datasets, the size of a folder containing all daily temperature datasets is about 3.5 terabytes as for each day there is a corresponding dataset of size about 500 megabytes. Since the ArcPy module (ultimately ArcGIS 10) reserves memory space for each computation within a for loop, I had to circumvent the finite barrier of the operating system by processing the data using a set of subprocesses.

a sample of six expert consultants.[48] Finally, aggregate to the district-level and normalize to a value between 0 and 1.[49]

The output data that resulted from the completion of the above tasks were imported to Stata 14 for cleaning and analysis. Replication files are available upon request (including results not shown).

[48] In contrast to Kienberger et al. (2017) I do not account for varying soil suitability across Afghanistan. Two reasons explain this. One, the six experts consulted in Kienberger et al. (2017) assigned a small weight (11 percent) to its importance as a measure of environmental poppy suitability. Two, the exact weighting scheme of the FAO-74 soil classification system was not disclosed by Kienberger et al. (2017).

[49] Normalizing is given by the transform $x_i \mapsto \frac{x_i - \min_i(x_i)}{\max_i(x_i) - \min_i(x_i)}$ for all districts i.

B Mathematical Appendix

Daily weather data can be used to approximate nonlinear weather-conflict relationships.[50] To see this, suppose that the weather-conflict relationship over some fixed sample period $\tau = [\underline{t}, \bar{t}]$ is

$$C_{dt} = \sum_{k \in \mathcal{K}_t} f(w_{dtk}) + \eta_{dt}, \quad t \in \tau, \tag{B.1}$$

where f is a continuous function; C_{dt} is some measure of conflicts (e.g., conflict incidence) in district d time period t (e.g., year-month); w_{dtk} is a weather event of district d day k in time period t that is a realization from a climatic element whose empirical distribution function across the days has a compact support $\mathcal{S}_{dt}$ that is specific for each district d and time period t;[51] $\mathcal{K}_t$ is the set of days in time period t; and η_{dt} is a nuisance parameter containing controls, determinants, fixed effects and error terms in a regression model with C_{dt} as the dependent variable (e.g., my baseline specification (1)). In other words, for each district, the assumption is that effect of weather on conflicts during a given year-month is additively separable into a sum of continuous diurnal weather effects.

The importance of this assumption is its implication. For if given, my baseline specification retrieves inherently nonlinear effects at the daily level. Indeed, as Proposition 1 below shows, Assumption B.1 implies that a linear parametric function with parameters describing averages of nonlinear weather effects approximates the effect of the particular weather phenomena on conflict.

Proposition 1 (Binned Weather-Conflict Dose-Response Function)**.** *Assumption*

[50] See Subsection 4.1 in Hsiang (2016) for an overview on how to identify nonlinear effects of weather variations.

[51] The support define the physical limits of weather variations. For a formalization, see Hsiang (2016).

B.1 justifies the approximation

$$C_{dt} \approx \sum_b \beta_b w_{dt}^b + \eta_{dt}, \quad t \in \tau, \tag{B.2}$$

where for a bin b of $\mathcal{S}_{dt}$, $w_{dt}^b = \sum_{k \in \mathcal{K}_t} \mathbf{1}_b(w_{dtk})$ with $\mathbf{1}_b(w_{dtk}) = 1$ if $w_{dtk} \in b$ and 0 otherwise.

Proof. Let $t \in \tau$ and partition $\mathcal{S}_{dt}$ into nontrivial and nonoverlapping bins.[52] Then, since f is continuous, by the Mean Value Theorem for Integrals, for each bin $b = [\underline{w}_b, \overline{w}_b]$ there is a real $\xi_b \in b$ such that

$$f(\xi_b) = \frac{1}{\overline{w}_b - \underline{w}_b} \int_{\underline{w}_b}^{\overline{w}_b} f(w)\, dw. \tag{B.3}$$

The above equation justifies the approximation

$$f(w_{dtk}) \approx \sum_b f(\xi_b) \mathbf{1}_b(w_{dtk}). \tag{B.4}$$

Equation B.3 and Assumption B.1 then implies

$$C_{dt} = \sum_{k \in \mathcal{K}_t} f(w_{dtk}) + \eta_{dt} \tag{B.5}$$

$$\approx \sum_{k \in \mathcal{K}_t} \sum_b f(\xi_b) \mathbf{1}_b(w_{dtk}) + \eta_{dt} \tag{B.6}$$

$$= \sum_b \underbrace{f(\xi_b)}_{\beta_b} \underbrace{\sum_{k \in \mathcal{K}_t} \mathbf{1}_b(w_{dtk})}_{w_{dt}^b} + \eta_{dt}. \tag{B.7}$$

■

From Proposition 1, two important corollaries follows. First, (B.2) shows that the parameters $\{\beta_b\}_b$ can be retrieved by regressing C_{dt} on $\{w_{dt}^b\}_b$, after accounting for terms in η_{dt}.

[52] The partitioning is possible as the empirical support is countable.

Second, each β_b identifies an average effect of potentially nonlinear effects of weather on conflict, as seen in (B.3). It is therefore beneficial concerning functional form specification to tighten the length of each bin as it makes the functional form restriction described above more plausible. Though decreasing the length of each bin reduces the approximation error theoretically, in practice, it increases the measurement error (cf. footnote 33) and also the number of parameters to be estimated so that the degrees-of-freedom decreases. In the limit as the length of each bin approaches zero, the number of in-sample perfectly multicollinear w_{dt}^b increases and estimation cannot be carried out using OLS. This trade-off cannot be parametrically resolved using known statistical inferential tools. Instead, the trade-off was experimentally addressed by cutting the support of the empirical weather distribution at end points and then choosing bins of equal lengths within the remaining support such that the nonlinear weather-conflict relationship was retrieved, as discussed in Section 4 and 5.

Augmenting Assumption B.1 with the following assumption yields my baseline specification (1):

$$C_{dt} = \sum_{k \in \mathcal{K}_t} \boldsymbol{f}(\boldsymbol{w}_{dtk}) + \eta_{dt}, \tag{B.8}$$

where $\boldsymbol{f}$ is a multivariate function and $\boldsymbol{w}_{dtk}$ a vector of weather variables. Under the assumption that temperature and precipitation are sufficient statistics for capturing the weather-conflict relationship I let $\boldsymbol{f}$ be bivariate and $\boldsymbol{w}_{dtk} = (T_{dtk}, P_{dtk})$. To acquire (1) I assume that the effect of temperature and precipitation is additively separable; i.e., that

$$\boldsymbol{f}(\boldsymbol{w}_{dtk}) = f_1(T_{dtk}) + f_2(P_{dtk}), \tag{B.9}$$

where f_1 and f_2 are two univariate continuous functions. Now, by analogy, following the steps of Proposition 1 we acquire (1).

To conclude: I have argued that my baseline specification (1) approximate any continuous weather-conflict relationship (linear or nonlinear). However, I do not derive the statistical approximation error.[53] Lastly, I note that one can relax the assumption that the functions are invariant across space and time by interacting weather variables with, e.g., regional or seasonal dummies. It is also simple to include temporally and spatially lagged functions (see, e.g., Appendix D).

[53] It would be of interest to statistically assess the error made. One idea is the following. First, view the empirical support of temperature and precipitation as a set of 2-cells. Second, with, e.g., 14 temperature-day, and 13 precipitation-day, bins as in my baseline specification, define $14 \cdot 13 = 182$ 2-cells. Then, adjust (1) by excluding the temperature- and precipitation-day bins, and including all these 2-cells. Lastly, test if the temperature response functions vary by precipitation 1-cells. This test was carried out on my dataset. However, because of the fixed effects, many 2-cells were omitted due to multicollinearity, making the empirical content of the test uncertain.

C Additional Descriptive Statistics

This section provides additional descriptive statistics. Figure C.1 illustrates the alternative temperature- and precipitation-day bins used in Section 5.2, and Table C.1 provide descriptive statistics relating to the analysis in Section 5.3.

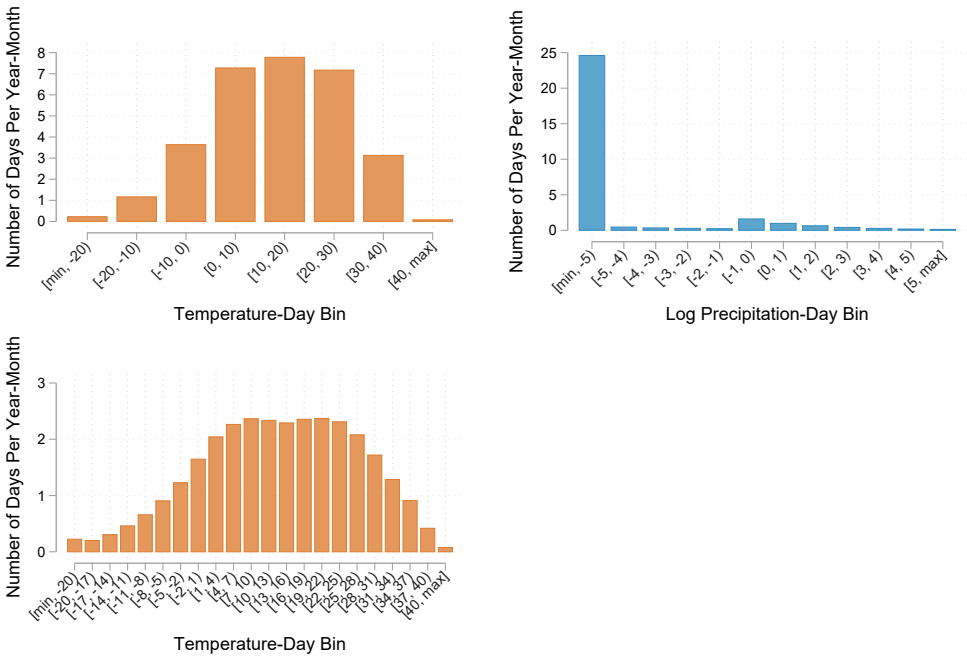


FIGURE C.1. ALTERNATIVE TEMPERATURE- AND PRECIPITATION-DAY BINS

Notes: The figure shows the average distribution of daily average temperature and precipitation across 8 and 22 temperature-day bins (upper and lower left panel) and 13 log precipitation-day bins (upper right panel). Each bar represents the average number of days per year-month in each temperature or precipitation category across all 398 districts of Afghanistan over the sample period July 2005 to December 2016. Minimum daily temperature is about -41 °C, and maximum daily temperature is about 45 °C. Maximum daily precipitation is about 117 mm. See the text for more details.

Source: Author's calculations based on data from the CHIRPS and NASA GSFC.

TABLE C.1—DESCRIPTIVE STATISTICS BY OPIUM GROWING STATUS: WEATHER AND CONFLICTS

			SD	
	Observations	Mean	Overall	Within
Panel A: Opium Growing Districts				
Battle-related deaths†				
All	36,432	1.59	8.98	8.17
If > 0	5,636	10.28	20.78	17.53
$\mathbf{1}$(Battle-related deaths > 0)†	36,432	0.15	0.36	0.30
Conflict Events†				
All	10,453			
Government of Afghanistan vs. Taleban	9,537			
Taleban vs. Civilians	450			
Government of Afghanistan vs. IS	253			
Daily Temperature (°C)‡	1,205,688	14.50	13.26	3.66
Daily Precipitation (mm)‡	1,205,688	0.92	3.10	2.94
Opium Suitability Index$^{\mathbb{P}}$	264	0.38	0.16	
Panel B: Non-Opium Growing Districts				
Battle-related deaths†				
All	18,492	1.06	6.73	6.73
If > 0	2,460	7.96	16.89	13.30
$\mathbf{1}$(Battle-related deaths > 0)†	18,492	0.13	0.34	0.29
Conflict Events†				
All	4,015			
Government of Afghanistan vs. Taleban	13,246			
Taleban vs. Civilians	196			
Government of Afghanistan vs. IS	10			
Daily Temperature (°C)‡	611,978	10.88	12.88	3.61
Daily Precipitation (mm)‡	611,978	1.02	3.39	3.23
Opium Suitability Index$^{\mathbb{P}}$	134	0.38	0.19	

Notes: This table reproduces the information in Table 1 and information on opium suitability by opium growing status. A district is opium growing if it has in any year between 2006 to 2016 cultivated opium poppies. The summary statistic Overall SD stands for the overall standard deviation of the corresponding variable. The summary statistic Within SD stands for the overall standard deviation of the corresponding variable after removing district-month fixed effects. The variable $\mathbf{1}$(Battle-related deaths > 0) is 1 if there is at least one battle-related death, and 0 otherwise. The variable Opium Suitability Index is the constructed district-level index for environmental suitability for opium poppy cultivation. The acronym IS stands for Islamic State (of Iraq and Syria). The sample period is July 2005 to December 2016. All 398 districts are included in the sample. Numbers are correct to two decimal places. †Measured at district-year-month-level. ‡Measured at district-year-month-day-level. $^{\mathbb{P}}$Measured at district-level.

Source: Author's calculations based on data from the CHIRPS, NASA GSFC and UCDP.

D Additional Robustness Checks

I now show that my baseline results in Section 5.1 are robust to a series of tests.

Falsification Test.—I plot estimates of $\{(\beta_i^{-1}, \gamma_j^{-1})\}_{(i,j)}$ from

$$C_{dt} = \sum_{l=-1}^{1} \Big(\sum_i \beta_i^l T_{d,t-l}^i + \sum_j \gamma_j^l P_{d,t-l}^j \Big) + \delta_{dm} + \pi_{pt} + q_{dm}(y) + \epsilon_{dt}. \tag{D.1}$$

That is, I include one year-month leads of the temperature- and precipitation-day bins to my baseline specification. Since I do not expect future diurnal variation in temperature and precipitation to affect current conflict incidence, I hypothesize that the coefficients on the leads are all equal to zero. However, using an F-test, I reject this hypothesis at the 5, but not 1, percent α-level. Still, as made clear by Figure D.1, this seems to be due to a seemingly spurious coefficient on the $[0.2, 0.4)$ mm bin, as all others are insignificant. Indeed, if I ignore the $[0.2, 0.4)$ mm bin, I cannot reject the hypothesis at the 10 percent α-level. This suggests that, overall, forward temperature and precipitation play no significant role in driving conflict incidence.

Different Standard Error Corrections.—In my baseline specification, I two-way cluster at the district- and year-month-level. This clustering design allows for serial correlation within districts and spatial correlation within year-months. This section tests for robustness to three alternative designs.

The baseline design does not account for temporally lagged spatial dependence (e.g., the dependence of observations in two districts separated in time). I, therefore, check for robustness to two other two-way clustering designs. First I cluster by district and region-year. This design allows for arbitrary correlation within region-years (minimum cluster size is 72). That is, all district-months

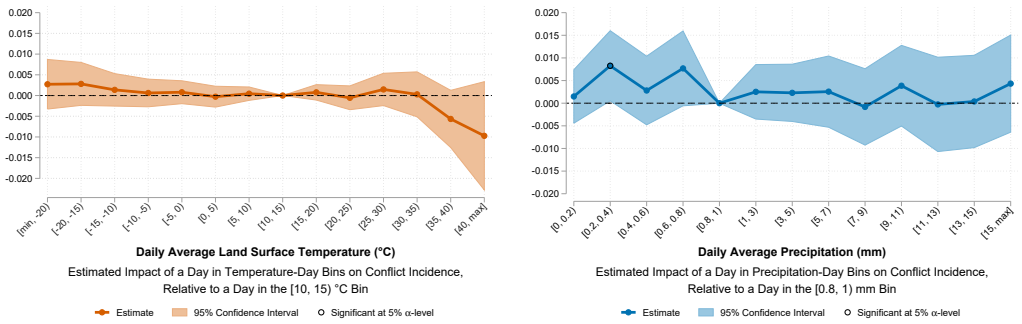


FIGURE D.1. CONFLICT INCIDENCE RESPONSE FUNCTIONS NEXT YEAR-MONTH

Notes: These figures plots estimates and 95 percent confidence bands of the one year-month lead conflict incidence temperature and precipitation response functions obtained by estimating (D.1). The omitted temperature- and precipitation-day bins are $[10, 15)$ °C and $[0.8, 1)$ mm. The number of observations is 54,128 from a balanced panel of 398 districts and 136 year-months. Mean conflict incidence is about 15 percent. Standard errors two-way clustered at the district- and year-month-level.

that lie in the same region-year (e.g., north of Afghanistan in 2006) are allowed to depend on each other both spatially and temporally. Second, since the former assumes unobservables in district-months in different regions are unrelated, I also cluster by district and season-year, which allows for correlation between all district-months within a season-year (e.g., spring 2006) (minimum cluster size is 47).

I also consider robustness to Conley (1999) spatial-HAC standard errors. The spatial-HAC correction account for heteroskedasticity and within-time spatial and within-location serial correlation of unobservables.[54] For the spatial correction, I use a uniform kernel that is assumed to discontinuously fall from 1 to 0 at some spatial cutoff distance (Conley, 2010). I let the cutoff distance be so large that arbitrary spatial correlation is allowed for (i.e., vanishes at 1,000 kilometers). For the HAC correction, I use the Newey-West (Bartlett) kernel

[54]Implementation is based on Fetzer (2014) and Hsiang, Meng and Cane (2011).

that weight pairs of observations within districts such that the weights decay linearly across time periods. I assume that there is no bound to the serial correlation (i.e., vanishes at 100,000 year-months). The Conley (1999) spatial-HAC standard error correction is conceptually similar to my baseline two-way clustering design but is not computationally equivalent.

Figure D.2 show the result from re-estimating the baseline specification and employing the three above adjustments. Note that I also include the baseline standard error correction and one-way clustered standard errors for completeness. Since I am unable, for technical reasons, to include the trend component $q_{dm}(y)$ when using the Conley (1999) spatial-HAC standard error correction, I further exclude $q_{dm}(y)$ for comparability. The figure shows that the two-way clustering designs yield nearly identical confidence bands, but the baseline two-way clustering design produces a wider confidence band at the highest temperature-day bin. Confidence bands based on the Conley (1999) spatial-HAC correction are less conservative. The one-way clustered confidence bands are only marginally smaller than the two-way equivalents. In conclusion, in comparison to these adjustments, my baseline standard error correction method does not overestimate standard errors and is in a worst case scenario too conservative.

Temporal Lag Length.—I consider the significance of delayed effects of higher order by adding two to five year-month lags to my baseline specification (1). I find in Figure D.3 that none of these delayed effects are significant for either temperature or precipitation.

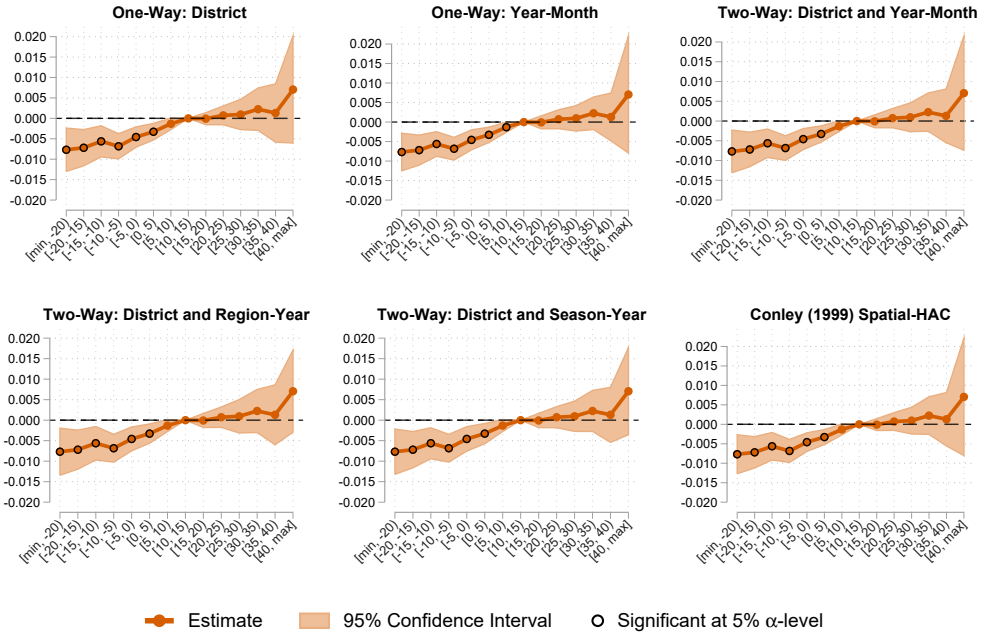
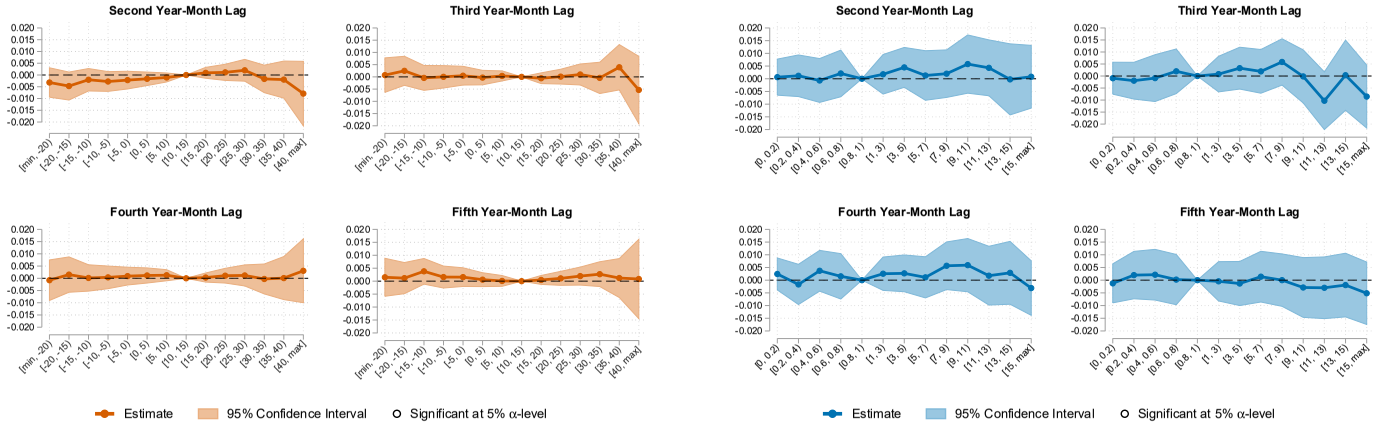


FIGURE D.2. CONFLICT INCIDENCE TEMPERATURE RESPONSE FUNCTIONS—CONFIDENCE BANDS BY STANDARD ERROR CORRECTION METHOD

Notes: These figures plot estimates and 95 percent confidence intervals of $\{\beta_i^0\}_i$ obtained by estimating (1) after excluding the trend component $q_{dm}(y)$. Standard errors are corrected using six different standard error correction methods. Upper left panel: One-way clustering at the district-level (cluster size is 398). Upper middle panel: Clustering at the year-month level (cluster size is 137). Upper right panel: Two-way clustering at district- and year-month-level (minimum cluster size is 137). Lower left panel: Two-way clustering at the district and region-year level (minimum cluster size is 72). Lower middle panel: Two-way clustering at the district and season-year level (minimum cluster size is 47). Lower right panel: Conley (1999) spatial-HAC correction with 1,000 km as the spatial distance cutoff in a uniform kernel weight function, and Newey-West (Bartlett) kernel with 100,000 year-months as the temporal distance cutoff. The number of observations is 54,526 from a balanced panel of 398 districts and 137 year-months. Mean conflict incidence is about 15 percent.



A. LAGGED TEMPERATURE RESPONSE FUNCTIONS

B. LAGGED PRECIPITATION RESPONSE FUNCTIONS

FIGURE D.3. LAGGED CONFLICT INCIDENCE RESPONSE FUNCTIONS

Notes: These figures plots estimates and 95 percent confidence intervals of two to five year-month lagged conflict incidence temperature and precipitation response functions obtained by estimating (1) with the second to fifth temporal lag order of all temperature- and precipitation-day bins added. The number of observations is 52,934 from a balanced panel of 398 districts and 133 year-months. Mean conflict incidence is about 15 percent. Standard errors two-way clustered at the district- and year-month-level.

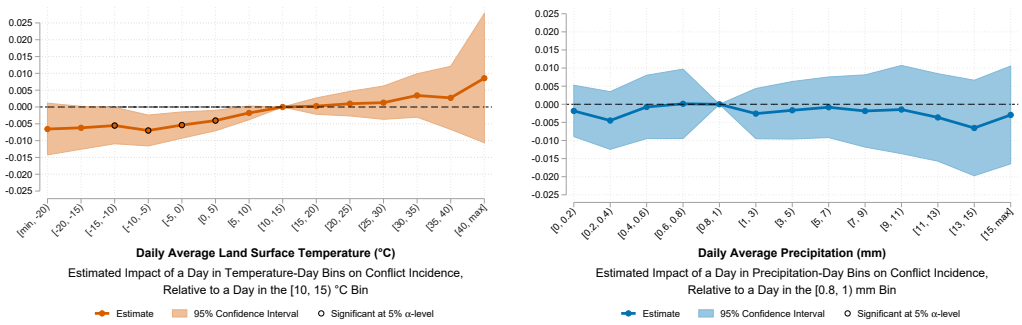
Spatial Spillovers.—Climatic events can cause outcomes to be displaced across space and spill over to neighboring districts. Spillovers at the provincial-level have been accounted for by the province-year-month fixed effects of my baseline specification. However, I do not consider spillovers across provincial boundaries. Spillovers across provincial boundaries due to weather variations can be partly accounted for by the following model:

$$C_{dt} = \sum_{l=0}^{1} \sum_{\pi=0}^{3} \Big(\sum_{i} \beta_i^{\pi l} T^i_{\pi d,t-l} + \sum_{j} \gamma_j^{\pi l} P^j_{\pi d,t-l} \Big) + \delta_{dm} + \pi_{pt} + q_{dm}(y) + \epsilon_{dt}, \quad \text{(D.2)}$$

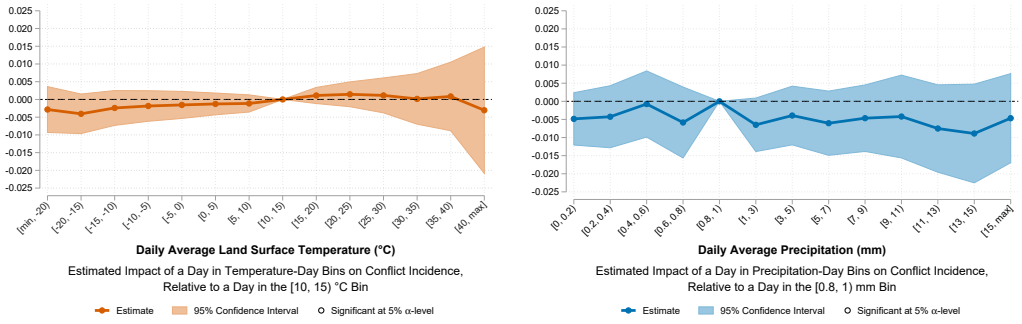
where $T^i_{\pi d,t-l}$ ($P^j_{\pi d,t-l}$) for $\pi > 0$ denote the average number of days with temperature (precipitation) levels in bin i (j) of districts whose centroids lie within a $(\pi - 1, \pi]$ one hundred kilometer band from the centroid of district d and $T^i_{0d,t-l} \equiv T^i_{d,t-l}$ ($P^j_{\pi d,t-l} \equiv P^i_{d,t-l}$). The coefficients $\{(\beta_i^{\pi l}, \gamma_j^{\pi l})\}_{(i,j)}$ for $\pi > 0$ capture spatial spillovers such that, e.g., $\beta_i^{\pi l}$ is the estimated impact on conflict incidence of—in all districts within a $(\pi - 1, \pi]$ one-hundred kilometers radius—exchanging one day with temperature levels in the omitted bin to a day in bin i in l year-months past to the current year-month.

Figure D.4a and D.4b show the baseline estimates when including average climate exposure within 0 to 300 kilometers in temperature and precipitation. The baseline results are by and large robust to this alternative specification. Figure D.5 and D.6 plot the spatially lagged temperature and precipitation response functions in the current and prior year-month. I find that spatial spillovers are insignificant.[55]

[55] I note that results are similar if one drops the province-year-month fixed effects except that contemporaneous average temperature exposure within a $(0, 100]$ km radius from a district d significantly affect current conflict incidence in district d (not shown). This finding provides an additional argument for the inclusion of province-year-month fixed effects in the baseline specification (1).



A. CURRENT YEAR-MONTH

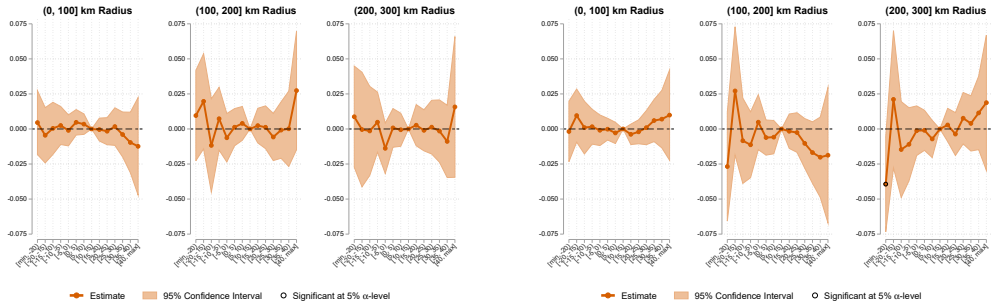


B. PRIOR YEAR-MONTH

FIGURE D.4. CONFLICT INCIDENCE TEMPERATURE AND PRECIPITATION RESPONSE FUNCTIONS—CONTROLLING FOR REMOTE TEMPERATURE AND PRECIPITATION VARIATION

Notes: The left and right figure of Panel A (B) plots estimates and 95 percent confidence bands of the contemporaneous (one year-month lagged) conflict incidence temperature and precipitation response functions obtained by estimating (D.2). The omitted temperature- and precipitation-day bins are $[10, 15]$ °C and $[0.8, 1)$ mm. The number of observations is 54,526 from a balanced panel of 398 districts and 137 year-months. Standard errors two-way clustered at the district- and year-month-level.

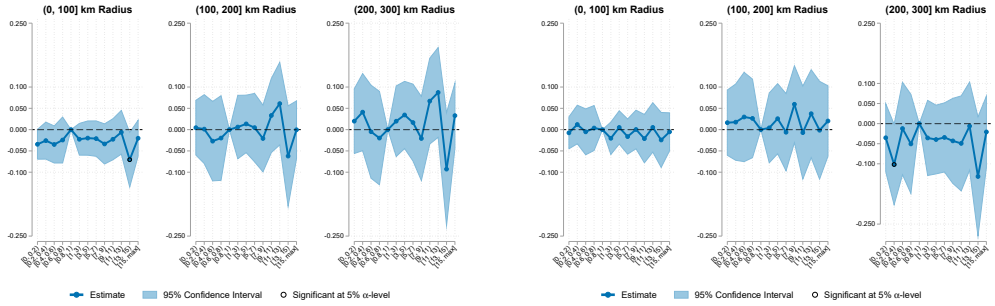
Grid Cell Fixed Effects.—Provinces and their boundaries are politically defined. These may therefore not suitably capture differences in agro-climatic zones that better describe the regional effects of climate change. To try to account for this, I distribute the 398 districts of Afghanistan into 30 2.5 by 2 lon-



A. CURRENT YEAR-MONTH B. PRIOR YEAR-MONTH

FIGURE D.5. CONFLICT INCIDENCE SPATIAL TEMPERATURE LAGS

Notes: These figures plots estimates and 95 percent confidence intervals of $\{\beta_i^{\pi l}\}_i$ for all $\pi \in \{1,2,3\}$ and $l \in \{0,1\}$ obtained by estimating (D.2). The number of observations is 54,526 from a balanced panel of 398 districts and 137 year-months. Mean conflict incidence is about 15 percent. Standard errors two-way clustered at the district- and year-month-level.



A. CURRENT YEAR-MONTH B. PRIOR YEAR-MONTH

FIGURE D.6. CONFLICT INCIDENCE SPATIAL PRECIPITATION LAGS

Notes: These figures plots estimates and 95 percent confidence intervals of $\{\gamma_i^{\pi l}\}_i$ for all $\pi \in \{1,2,3\}$ and $l \in \{0,1\}$ obtained by estimating (D.2). The number of observations is 54,526 from a balanced panel of 398 districts and 137 year-months. Mean conflict incidence is about 15 percent. Standard errors two-way clustered at the district- and year-month-level.

gitude by latitude grid cells. Then I replace π_{pm} with π_{gm}, where g denotes a 2.5 by 2 longitude by latitude grid cell in (1). These cells are disjoint and cover all districts of Afghanistan. I find that precipitation still plays no role and that fewer coefficients are significant at the 5 percent α-level for the temperature conflict incidence relationship, although some are still significant at the 10 percent α-level. However, the caveat with the grid cell fixed effects is that since I assign district centroids to grid cells, there are grid cells that cover districts but not their centroids. This problem does not occur for the province-year-month fixed effects as provinces are by construction polygons that cover each of its districts perfectly. I, therefore, prefer results based on the baseline set of fixed effects.

Controlling for District-Year Fixed Effects.—Climate change adaptation may be the result of observed or expected extreme weather events. If there is inter-annual district-level climate change adaptation (e.g., improved water resource management) that affect opportunity costs to conflict and peace, my baseline estimators may be subject to omitted variable bias. Though this seems unlikely to be the case as Afghanistan only recently began to prepare rural communities for climate change (NEPA, UNEP and WFP, 2016), I here control for this potential omission by adding district-year fixed effects to my baseline specification. Figure D.8 reproduce the baseline results under this alternative specification. I find that the baseline results are by and large robust to the inclusion of district-year fixed effects.[56]

Conflict Intensity.—My focus on conflict incidence reflects my interest in ex-

[56]Note that the fact that the estimates in Figure D.8 are closer to zero compared with the baseline estimates may be a result of exacerbated attenuation bias if there were no district-year-level omitted variables in the baseline specification.

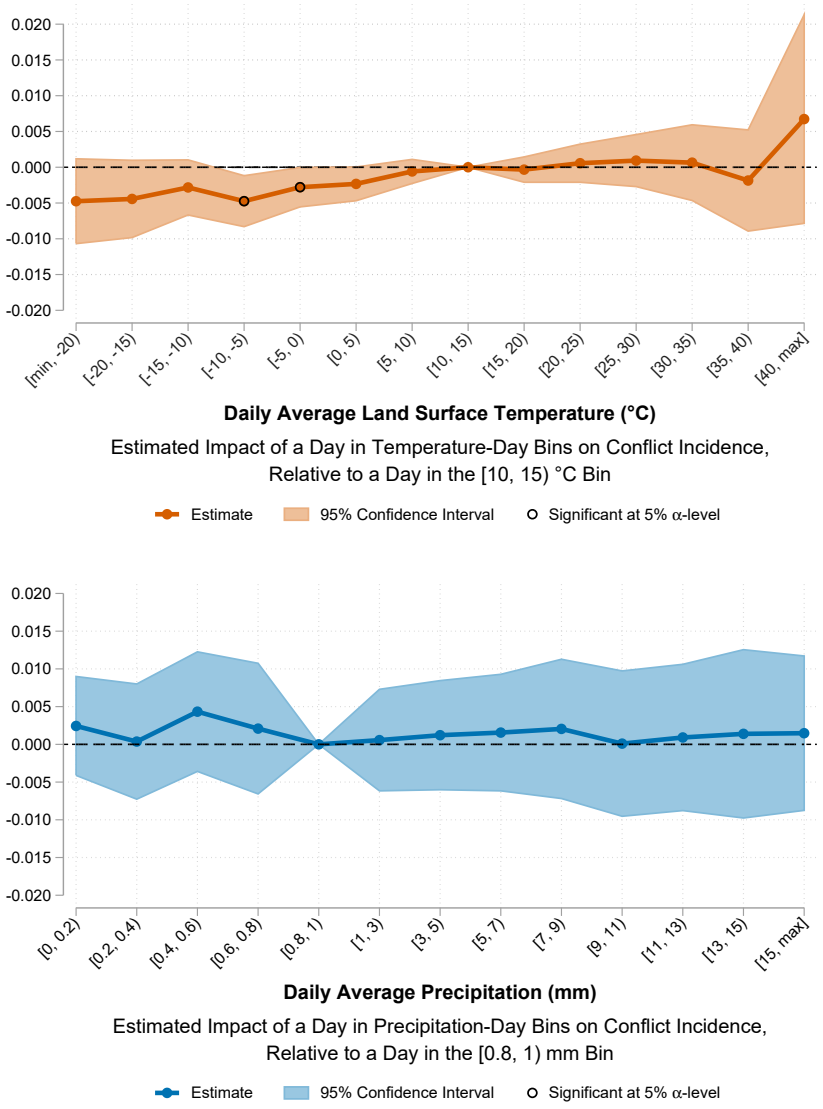
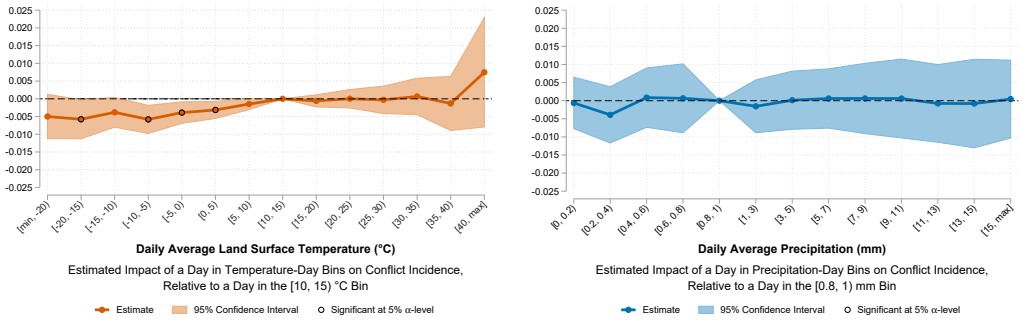
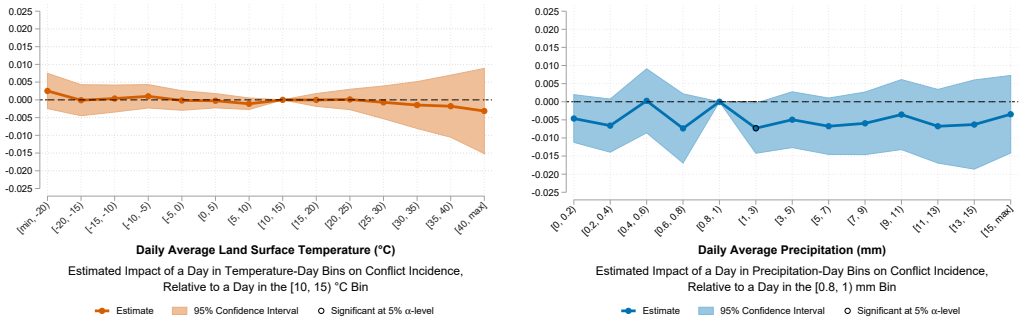


FIGURE D.7. CONFLICT INCIDENCE RESPONSE FUNCTIONS—ALTERNATIVE REGIONAL FIXED EFFECTS

Notes: These figures show temperature and precipitation response functions in the current year-month obtained by estimating (1) after replacing province-year-month fixed effects with year-month-varying 2.5 by 2 longitude-latitude grid cell fixed effects. The number of observations is 54,526 from a balanced panel of 398 districts and 137 year-months. Mean conflict incidence is about 15 percent. Standard errors two-way clustered at the district- and year-month-level.



A. CURRENT YEAR-MONTH



B. PRIOR YEAR-MONTH

FIGURE D.8. CONFLICT INCIDENCE RESPONSE FUNCTIONS—CONTROLLING FOR DISTRICT-YEAR FIXED EFFECTS

Notes: The left and right figure of Panel A (B) plots estimates and 95 percent confidence bands of the contemporaneous (one year-month lagged) conflict incidence temperature and precipitation response functions obtained by estimating (1) after adding district-year fixed effects. The omitted temperature- and precipitation-day bins are $[10, 15]$ °C and $[0.8, 1)$ mm. The number of observations is 54,526 from a balanced panel of 398 districts and 137 year-months. Mean conflict incidence is about 15 percent. Standard errors two-way clustered at the district- and year-month-level.

plaining the general presence of conflicts. An alternative definition of violence is conflict intensity measured as the number of battle-related deaths. If a significant relationship between weather and conflict incidence but not between weather and conflict intensity, or vice verse, is found, this does not imply that one or the other model does not identify the parameters of interest. However, if results stand in stark contrast to each other, this may cast doubt on my baseline result. Furthermore, the relationship between weather and conflict intensity is in itself interesting.

To estimate the relationship between weather and conflict intensity I fit the number of battle-related deaths to a Poisson regression model. Specifically, for the number of battle-related deaths BRD_{dt} in district d year-month t I impose the following probability density function:

$$\mathbb{P}(BRD_{dt} = n|\boldsymbol{x}_{dt}) = \exp\left(\mu(-\boldsymbol{x}_{dt})\right)\frac{\mu(\boldsymbol{x}_{dt})^n}{n!}, \quad n = 0, 1, \ldots, \tag{D.3}$$

where the link function $\mu(\boldsymbol{x}_{dt}) \equiv \mathbb{E}(BRD_{dt}|\boldsymbol{x}_{dt})$ provides a parametric form for the conditional mean of conflict intensity given all covariates $\boldsymbol{x}_{dt}$:

$$\mu(\boldsymbol{x}_{dt}) = \exp\left(\sum_{l=0}^{1}\left(\sum_i \beta_i^l T_{d,t-l}^i + \sum_j \gamma_j^l P_{d,t-l}^j\right) + \delta_{dm} + \pi_{pt}\right). \tag{D.4}$$

For this model, $100\beta_i^l$ ($100\gamma_j^l$) approximate the percentage change in the conditional mean number of battle-related deaths from exchanging a day in the omitted temperature-day (precipitation-day) bin to a day with temperature (precipitation) levels in bin i (j).

Estimating a Poisson regression model is more suitable than a log-linear model of conflict intensity for three reasons. First, it accounts for the fact that the number of battle-related deaths is a count variable and easily handles ob-

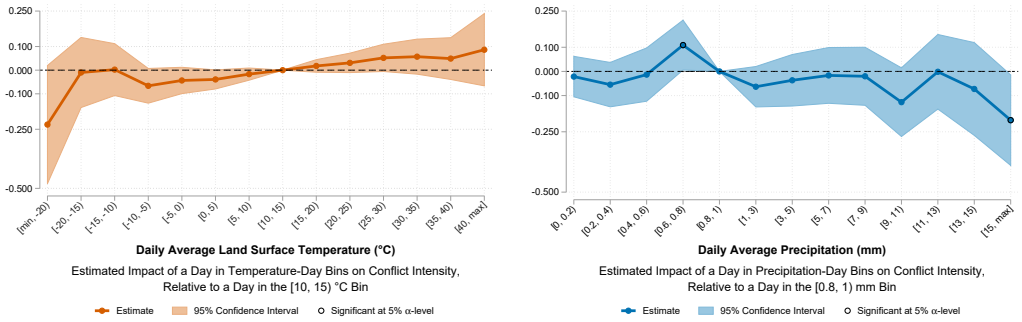
served values of conflict intensity equal to zero.[57] Second, even if the Poisson distributional assumption does not correctly describe the empirical distribution of the number of battle-related deaths, maximum likelihood estimation produces unbiased estimates of the coefficients if (D.4) correctly describes the conditional mean of conflict intensity (Wooldridge, 1997, 1999). Third, the Poisson regression model does not suffer from the incidental parameters problem (Cameron and Trivedi, 2005).

There are however three primary drawbacks to my Poisson regression model. First, I have to omit district-month-specific yearly trends $q_{dm}(y)$. Second, I can only account for serial correlation in the error terms by clustering at the district-level, thereby ignoring spatial correlation. Third, estimation of Poisson fixed effects models tend to lead to some loss of data as observations who do not vary within a group specified by a fixed effect (e.g., a district-month) are dropped as these do not contribute to maximizing the likelihood (Cameron and Trivedi, 2005). In any case, I believe that my Poisson model is the best among sub-optimal solutions to study conflict intensity and treat results from my Poisson model as highly indicative only.

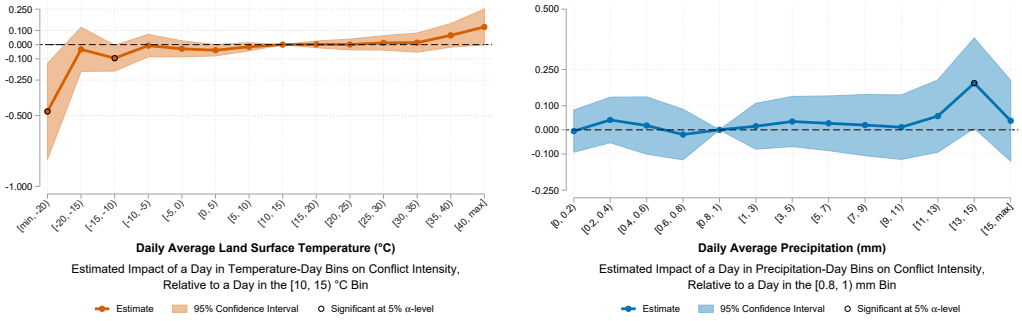
Figure D.9 show response functions for conflict intensity obtained by estimating (D.3). The estimation method drops observations for which the number of battle-related deaths is constantly zero within a fixed effects group. Conse-

[57] This, combined with the fact that the number of observations with zero values on the number of battle-related deaths is high, explains why I do not estimate a log-linear model as a robustness check. Some authors (e.g., Dube and Vargas (2013)) handle such a problem by replacing the variable with the log of the variable plus a small number ε (e.g., 0.001) to account for zero values. However, results from replacing BRD_{dt} with $\log(BRD_{dt} + \varepsilon)$ are highly sensitive to the choice of ε in the case of a high number of observations for which BRD_{dt} is zero. Indeed, I find that estimates can mechanically change and switch sign when experimentally varying ε over $(0.001, 0.0001)$.

quently, a loss of about 55 percent observations occurs. This substantial loss of data makes these results highly suggestive. Keeping this in mind and evaluating the results as they are I find that the role of precipitation and temperature in driving the number of battle-related deaths is barely significant. Also, three coefficients on contemporaneous and lagged precipitation seem spuriously significant as the pattern seem inexplicable. Nonetheless, the model prediction for contemporaneous temperature is consistent with the baseline results as higher contemporaneous temperature levels tend to increase the number of battle-related deaths. The predictions are also quantitatively meaningful, with, e.g., a decrease by about 3.9 percent (s.e. about 2.1 percent) in the mean number of battle-related deaths from contemporaneously exchanging a day with temperature levels between 10 and 15 °C to a day with temperature levels between 0 and 5 °C.



A. CURRENT YEAR-MONTH



B. PRIOR YEAR-MONTH

FIGURE D.9. CONFLICT INTENSITY RESPONSE FUNCTIONS

Notes: The left and right figure of Panel A (B) plots estimates and 95 percent confidence bands of the contemporaneous (one year-month lagged) conflict intensity temperature and precipitation response function obtained by estimating (D.3) with link function (D.4). The omitted temperature- and precipitation-day bins are $[10, 15]$ °C and $[0.8, 1)$ mm. The number of observations is 24,530 from an unbalanced panel of 350 districts and 137 year-months. Mean number of battle related deaths is about 3.14. Standard errors are clustered at the district-level.

E Estimation Method

I estimate my Poisson regression model with the Stata package `poi2hdfe` by Paulo Guimarães based on a procedure outlined in Figueiredo, Guimarães and Woodward (2015). The remaining multidimensional fixed effects regression models are estimated with a computationally efficient iterative and graph-theoretic estimation method developed by Correia (2016). The Stata package `reghdfe` that implement the estimation method by Correia (2016) allow for, e.g., two-way clustering and interaction of fixed effects with continuous variables. More notable is that the `reghdfe` package automatically drops singletons, i.e., fixed effects groups with only one observation. In this thesis, I only drop singletons for the conflict onset and ending model in Section 5.2. The reason is that maintaining singletons can overstate statistical significance (Correia, 2015). Though this may seem to be a non-optimal solution as we lose some information, there is currently no general solution on how to treat singletons.

F References of Appendix

Arino, Olivier, Jose Julio R. P., Vasileios Kalogirou, Sophie Bontemps, Pierre Defourny, and Eric van Bogaert. 2012. "Global Land Cover Map for 2009 (GlobCover 2009)." *European Space Agency and Université catholique de Louvain*.

Baysan, Ceren, Marshall Burke, Felipe Gonzalez, Solomon M. Hsiang, and Edward Miguel. 2015. "Economic and Non-Economic Factors in Violence: Evidence from Organized Crime, Suicides and Climate in Mexico." *Unpublished manuscript, University of California, Berkeley, California, USA*.

Conley, Timothy G. 1999. "GMM Estimation with Cross Sectional Dependence." *Journal of Econometrics*, 92(1): 1–45.

Conley, Timothy G. 2010. "Spatial Econometrics." In *Microeconometrics*. 303–313. London: Palgrave Macmillan.

Correia, Sergio. 2015. "Singletons, Cluster-Robust Standard Errors and Fixed Effects: A Bad Mix." *Unpublished manuscript, Duke University, North Carolina, USA*.

Correia, Sergio. 2016. "A Feasible Estimator for Linear Models with Multi-Way Fixed Effects." *Unpublished manuscript, Duke University, North Carolina, USA*.

Dell, Melissa. 2009. "GIS Analysis for Applied Economists." *Unpublished manuscript, Massachusetts Institute of Technology, Massachusetts, USA*.

Dube, Oeindrila, and Juan F. Vargas. 2013. "Commodity Price Shocks and Civil Conflict: Evidence from Colombia." *The Review of Economic Studies*, 80(4): 1384–1421.

Fetzer, Thiemo. 2014. "Social Insurance and Conflict: Evidence from India." *Unpublished manuscript*.

Figueiredo, Octávio, Paulo Guimarães, and Douglas Woodward. 2015. "Industry Localization, Distance Decay, and Knowledge Spillovers: Following the Patent Paper Trail." *Journal of Urban Economics*, 89: 21–31.

Hijmans, Robert J., Luigi Guarino, Mariana Cruz, and Edwin Rojas. 2001. "Computer Tools for Spatial Analysis of Plant Genetic Resources Data: 1. DIVA-GIS." *Plant Genetic Resources Newsletter*, 127: 15–19.

Hsiang, Solomon M. 2016. "Climate Econometrics." *Annual Review of Resource Economics*, 8: 43–75.

Hsiang, Solomon M., Kyle C. Meng, and Mark A. Cane. 2011. "Civil Conflicts Are Associated with the Global Climate." *Nature*, 476(7361): 438.

Kienberger, Stefan, Raphael Spiekermann, Dirk Tiede, Irmgard Zeiler, and Coen Bussink. 2017. "Spatial Risk Assessment of Opium Poppy Cultivation in Afghanistan: Integrating Environmental and Socio-Economic Drivers." *International Journal of Digital Earth*, 10(7): 719–736.

United Nations Office on Drugs and Crime (UNODC). 2008. "Afghanistan Opium Winter Rapid Assessment Survey 2008." Vienna: United Nations, Office on Drugs and Crime.

United Nations Office on Drugs and Crime (UNODC). 2013. "Afghanistan Opium Risk Assessment 2013." Vienna: United Nations, Office on Drugs and Crime.

Wooldridge, Jeffrey M. 1997. "Quasi-Likelihood Methods for Count Data." *Handbook of Applied Econometrics Volume 2: Microeconomics*, 321–368.

Wooldridge, Jeffrey M. 1999. "Distribution-Free Estimation of Some Nonlinear Panel Data Models." *Journal of Econometrics*, 90(1): 77–97.