Quantifying the Influence of Climate on Human Conflict

Instructions for replication

Solomon M. Hsiang, Marshall Burke, and Edward Miguel

This document describes how to replicate the results in Hsiang, Burke, and Miguel "*Quantifying the influence of climate on human conflict*". Full replication requires Stata, Matlab, and R. The document proceeds in 8 steps:

1. Description of the file structure
2. Instructions for replicating the individual point estimates that make up Figures 4 and 5 and the meta-analysis. This step is implemented in Stata.
3. Instructions for replicating the meta-analysis results. This step is implemented in R.
4. Instructions for replicating Figure 2. This step is implemented in Matlab.
5. Instructions for replicating Figure 3.
6. Instructions for replicating Figures 4 and 5. This step is implemented in R.
7. Instructions for replicating Figure 6. This step is implemented in Matlab
8. Instructions for replicating the figures and tables in the SOM.

All the datasets we use are either publicly available, or authors have graciously allowed them to be included in our replication materials. We thank the many authors we interacted with for their generosity in sharing data and in answering our many questions. In particular, we thank Christian Almer, David Blakeslee, David Card, Paul Burke, Halvard Buhaug, Peter deMenocal, Nils Peter Gleditsch, Mariaflavia Harari, Gerald Haug, Richard Larrick, Harry Lee, Lars Lefgren, Marc Levy, Suresh Naidu, John O'Loughlin, Matthew Ranson, Ole Magnus Theisen, and Nina von Uexkull for providing results and/or data. Any errors in the subsequent manipulation of these data are our own.

1 Description of the file structure

All replication files and code are contained in the zipped file `HsiangBurkeMiguel_ReplicationData.zip`. Unzipping this file yields three folders named `code`, `data`, and `output`. The `code` folder contains the following:

- `Replicate_Standardized_Effects.do`. This is a Stata .do file to replicate the estimates used in the meta-analysis and shown in Figures 4 and 5. This script automatically calls `ols_spatial_HAC.ado`, `cgmreg.ado`, `ivreg2.ado`, and `xtivreg2.ado`, which are located in the `state_subfunctions` sub-folder.
- `Replicate_metaAnalysis.R`. This R script produces most of the meta-analysis results in the paper, as well as Figures 4 and 5 in the main text, and Figure S3 and Tables S1, S2, and S3 in the SOM. The script generates the log file `metaAnalysis_log.txt`.
- `Make_Fig_1.do`. This Stata .do file produces Figure 1.
- `Make_Fig_2.m`. This Matlab script produces Figure 2. This script automatically calls subfunctions in the `matlab_subfunctions` sub-folder.

- `Make_Fig_6.m`. This Matlab script produces Figures 6. This script automatically calls subfunctions in the `matlab_subfunctions` sub-folder.
- `Make_Fig_S1_S2.r`. This R script produces Figures S1 and S2.
- `Make_Fig_S4.r`. This R script produces Figure S4.
- `Make_Table_S4.do`. This Stata .do file produces Table S4.

The `data` folder contains a number of sub-folders with individual datasets in them that are called by `Replicate_Standardized_Effects.do`. It also contains:

- `Figure1_data.csv`
- `standardized_effects.csv`. This file collects the individual point estimates from the papers and from `Replicate_Standardized_Effects.do`, and is called by `Make_Fig_4_5.r` to make Figures 4 and 5.
- `Figure6_data.csv`. This file contains the data underlying Figure 6, and is called by `Make_Fig_6.m` to make the figure.
- `Supplement_Data` folder containing `FigS1_data.csv` and `FigS2_data.csv`. These files contain the data underlying Figures S1 and S2 and are called by `Make_Fig_S1_S2.R` to make these figures.

The `output` folder contains the output created by the above files. Some of this output is in turn used to generate additional figures. If the above code is re-run, everything in this folder will be re-created.

2 Replicating individual point estimates

We begin by briefly describing each estimate we utilize, and whether the corresponding dataset is included in the replication materials or whether no re-analysis of the data was required and the estimate was just pulled directly from the relevant paper. The reader is referred to the SOM for further detail on each paper and the particular estimate we extract from it.

- Auliciems and DiBartolo 1995. We draw estimates directly from the published paper.
- Bergholt and Lujala 2012. We re-analyze their publicly-available replication data, http://file.prio.no/Journals/JPR/2012/49/1/Bergholt%20and%20Lujala%2049(1)%20Replication.zip. The data are included in our replication materials.
- Bohlken and Sergenti 2011. We re-analyze their publicly-available replication data, http://file.prio.no/Journals/JPR/2010/47/5/Replication%20data%20Bohlken%20and%20Sergenti%2047(5).zip. The data are included in our replication materials.
- Brückner and Ciccone 2011. We re-analyze their publicly available replication data, available from the Econometrics Society website: http://www.econometricsociety.org/. The data are included in our replication materials.
- Buhaug 2010. We draw estimates directly from the published paper and follow-up letter, as detailed in the SOM. The "within" standard deviation for the temperature variable is calculated using the Burke et al 2009 data.
- Burke et al 2009. We re-analyze their publicly available replication data, http://elsa.berkeley.edu/~emiguel/pdfs/climate_conflict_replication.zip. The data are included in our replication materials.

- Burke 2012. We re-analyze his publicly available replication data, https://crawford.anu.edu.au/pdf/staff/acde/paul_burke/Burke_BEMacro.zip. The data are included in our replication materials.
- Burke and Leigh 2010. We re-analyze their publicly available replication data. http://www.aeaweb.org/aej/mac/data/2008-0115_data.zip. The data are included in our replication materials.
- Card and Dahl 2011. We re-analyze their publicly available replication data, http://dss.ucsd.edu/~gdahl/football-violence-code.html. The data are included in our replication materials.
- Couttenier and Soubeyran 2012. We draw estimates directly from the published paper.
- Dell, Jones, and Olken 2012. We re-analyze their publicly available replication data, http://www.aeaweb.org/aej/mac/data/2010-0092_data.zip. The data are included in our replication materials.
- Fjelde and von Uexkull 2012. Data and replication code were kindly provided to us by the authors, and the authors have given us permission to include them in our replication materials.
- Harari and La Ferrara 2011. We draw estimates directly from the paper.
- Hendrix and Salehyan 2012. http://file.prio.no/Journals/JPR/2012/49/1/Replication%20Data%20Hendrix%20and%20Salehyan.zip. The data are included in our replication materials.
- Hidalgo et al 2010. We re-analyze their publicly available replication data, http://dvn.iq.harvard.edu/dvn/dv/restat/faces/study/StudyPage.xhtml?globalId=hdl:1902.1/15305&studyListingInde 3_8163c462bc4224d979014e6c501e. The data we provide in the replication materials is a subset of what is available in the full dataset.
- Hsiang, Meng, and Cane 2011. We re-analyze their publicly available replication data, http://dl.dropbox.com/u/3011470/CONFLICT/replication_HSIANG_MENG_CANE_2011.zip.The data are included in our replication materials.
- Jacob, Lefgren, and Moretti 2007. Data and replication code were kindly provided to us by the authors, and the authors have given us permission to include them in our replication materials.
- Larrick et al 2011. Data and replication code were kindly provided to us by the authors, and the authors have given us permission to include them in our replication materials. The underlying data from which their dataset is constructed are available here: http://www.retrosheet.org/.
- Levy et al 2005. Data and replication code were kindly provided to us by the authors, and they have given us permission to include them in our replication materials.
- Maystadt, Ecker, and Mabiso 2013 . We draw estimates directly from the published paper.
- Miguel, Satyanath, and Sergenti 2004. We re-analyze their publicly available replication data, http://elsa.berkeley.edu/~emiguel/pdfs/mss_repdata.dta and http://elsa.berkeley.edu/~emiguel/pdfs/mss_rep_results.do. The data are included in our replication materials.
- Miguel 2005. These data were not previously publicly available, but we include them in our replication materials.
- O'Laughlin et al 2012. We re-analyze their publicly available replication data, http://www.colorado.edu/ibs/climateconflict/manuscripts.html. The data are included in our replication materials.

- Ranson 2012. Data and replication code were kindly provided to us by the author, and he has given us permission to include them in our replication materials.
- Sekhri and Storeygard 2012. We draw estimates directly from the published paper.
- Theisen 2012. We re-analyze his publicly available replication data, http://file.prio.no/Journals/JPR/2012/49/1/Theisen%2049%20Replication.zip. The data are included in our replication materials.
- Theisen, Holterman and Buhaug 2012. Data and replication code were kindly provided to us by the authors, and they have given us permission to include them in our replication materials. The underlying PRIO-GRID data are available here: http://www.prio.no/Data/PRIO-GRID/.

The *Stata* code to replicate the estimate from each study is given in the file `Replicate_Standardized_Effects.do`. We provide replication code to compute the estimates as well as the log file of *Stata* output for all the estimates (`Replicate_Standardized_Effects.log`). The estimates are then manually collected in the `standardized_effects.csv` file. `Replicate_Standardized_Effects.do` also outputs a `.csv` file for a subset of the studies, which are then used to create Figure 2.

Replicating all the results requires installation of `ols_spatial_HAC.ado`, `cgmreg.ado`, `ivreg2.ado`, and `xtivreg2.ado`, which are provided in the `state_subfunctions` sub-folder.[1] The code is set to run out of the `code` folder, and should not require any extra installation of these .ado files.

3 Replicating the meta-analysis

The R script `Replicate_metaAnalysis.r` takes as an input the estimates in `standardized_effects.csv`, and outputs the meta-analysis results discussed in Section 4.2 of the main text, and outputs Figure S3 and Tables S1, S2, and S3 from the SOM.

4 Replicating Figure 1

This Stata .do file `Make_Fig_1.do` produces the two panels of Figure 1, which we then combine in Illustrator for publication.

5 Replicating Figure 2

The 12 panels in Figure 2 are created by the Matlab script `Make_Fig_2.m`. This script takes as input the `.csv` files created by `Replicate_Standardized_Effects.do`, and creates a non-parametric "watercolor" regression plot for each study in the figure. We then combine these panels in Illustrator to make the full figure. This script calls `watercolor_reg.m`, which is provided in the `matlab_subfunction` folder. More info on watercolor regression can be found here: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2265501.

6 Replicating Figure 3

Figure 3 was created by the authors using a combination of R and Adobe Illustrator. Each subfigure was either pulled directly from the relevant published article, or reconstructed from the underlying

[1] More information on `ols_spatial_HAC.ado` is available here: http://www.fight-entropy.com/2010/06/standard-error-adjustment-ols-for.html. More info on the multi-way clustering routine `cgmreg.ado` can be found in Colin Cameron, Jonah Gelbach, and Douglas L Miller, "Robust Inference with Multi-way Clustering" (NBER Working Paper, 2006). Code for `cgmreg.ado` can currently be found on Jonah Gelbach's website: http://gelbach.law.yale.edu/~gelbach/ado/cgmreg.ado.

data where possible and plotted in R. The sub-figures were then combined in Adobe Illustrator. The underlying data we obtained were as follows:

- Buntgen et al 2011. Data are available here: http://www.ncdc.noaa.gov/paleo/pubs/buentgen2011/buentgen2011.html
- Yancheva et al 2007. Data are on the NOAA server, available by searching for "Yancheva" at the following URL: http://gcmd.nasa.gov/KeywordSearch/Home.do?Portal=GCMD&MetadataType=0
- Cullen et al 2001. Data are available here: ftp://ftp.ncdc.noaa.gov/pub/data/paleo/contributions_by_author/cullen2000/cullen2000.txt
- Buckley et al 2010. Data were obtained via personal communication with the lead author, Brendan Buckley bmb@ldeo.columbia.edu.

7 Replicating Figures 4 and 5

The R script `Replicate_metaAnalysis.r` creates Figures 4 and 5, taking as input the estimates in `standardized_effects.csv`. These figures are then cleaned up slightly in Illustrator for publication.

8 Replicating Figure 6

Figure 6 is based on data from 21 global climate models in the CMIP3 archive that reported the A1B emissions scenario (see the journal article for more details). The data file we provide in the replication materials, `Figure6_data.csv`, is the mean pixel-level change in temperature across all models, divided by the historical inter-annual standard deviation of temperature at that pixel. The Matlab script `Make_Fig_6.m` produces the map in Figure 6, which we then clean up in Illustrator for publication.

9 Replicating figures and tables in the SOM

Figures S1 and S2 are created by the R script `Make_Fig_S1_S2.r`, using data in the `output/Supplement_Data/` folder. Figure S3 and Tables S1, S2, and S3 in the SOM are created by the R script `Replicate_metaAnalysis.r`. Figure S4 is created by the R script `Make_Fig_S4.r`.