

Reconciling disagreement over climate–conflict results in Africa

Solomon M. Hsiang^{a,b,1} and Kyle C. Meng^{c,d,1,2}

^aGoldman School of Public Policy, University of California, Berkeley, CA 94720; ^bNational Bureau of Economic Research, Cambridge, MA 02138; ^cWoodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ 08544; and ^dOffice of Economic Policy and Analysis, Environmental Defense Fund, New York, NY 10010

Edited by B. L. Turner, Arizona State University, Tempe, AZ, and approved November 27, 2013 (received for review August 26, 2013)

A recent study by Burke et al. [Burke M, Miguel E, Satyanath S, Dykema J, Lobell D (2009) *Proc Natl Acad Sci USA* 106(49):20670–20674] reports statistical evidence that the likelihood of civil wars in African countries was elevated in hotter years. A following study by Buhaug [Buhaug H (2010) *Proc Natl Acad Sci USA* 107(38):16477–16482] reports that a reexamination of the evidence overturns Burke et al.’s findings when alternative statistical models and alternative measures of conflict are used. We show that the conclusion by Buhaug is based on absent or incorrect statistical tests, both in model selection and in the comparison of results with Burke et al. When we implement the correct tests, we find there is no evidence presented in Buhaug that rejects the original results of Burke et al.

climate change | temperature | security

Understanding whether climate change may elevate levels of human conflict is a major scientific question with critical implications for society. In a pivotal and controversial article, Burke et al. (1) exploit year-to-year variation in countries’ temperature and precipitation to identify the causal effect of these variables on the incidence of civil war in sub-Saharan countries during 1981–2002. In their preferred specification, Burke et al. (1) report that a 1 °C increase in average temperature elevates the probability of civil war by 0.043, a 39% increase relative to the average rate of war during the period. Subsequent studies have obtained similar findings in modern Africa for intergroup conflict at local scales (2–4) and civil conflict at the continental scale (5), as well as at various scales elsewhere around the world (6, 7). However, work by Buhaug (8, p. 16480) reports that the original findings by Burke et al. “do not hold up to closer inspection.” This disagreement is widely cited by the media, policymakers, and other researchers (9–11) as statistical evidence that climatic conditions might not influence modern human conflict in sub-Saharan Africa. The extremity and persistence of this disagreement has recently prompted Solow (12, p. 180) to “call for peace on climate and conflict [research],” suggesting that “such disagreements indicate that a deeper look behind the statistics is warranted.”

Here we take a deeper look behind the statistics and reconcile the contradictory findings of Burke et al. (1) and Buhaug (8) by correcting the model selection and model comparison procedures used in Buhaug—both of which require formal statistical tests that are absent from the original analysis but are needed to draw the conclusions stated in Buhaug. We address four key errors made in the original analysis of Buhaug: *i*) controls for unobserved time-invariant confounders (country-fixed effects) and time-variant confounders (country-specific trends) are discarded without testing for their joint significance in the model; *ii*) qualitatively and quantitatively different conflict variables are compared with each other without first standardizing their units of measure; *iii*) the original results of Burke et al. are rejected based on comparisons of a new model with a null hypothesis of zero effect, rather than comparisons of the new model with the original results; and *iv*) coefficients and SEs in a logit regression are not converted to units of conflict risk before they are

evaluated. When we correct for these errors, we find that the results of Buhaug are not distinguishable from findings in Burke et al. Although we do not attempt to verify whether the finding of Burke et al. is correct here (we refer readers to refs. 6 and 7 for evaluations of Burke et al.), our findings invalidate the claim that the analysis of Buhaug overturns Burke et al. as well as Buhaug’s stronger statement that “[C]limate characteristics and variability are unrelated to short-term variations in civil war risk in sub-Saharan Africa” (8, p. 16481).

Results

In a multipronged critique of Burke et al. (1), Buhaug (8) displays 12 alternative estimates (models 2–13 in Buhaug) presented across three tables that are incorrectly compared with Burke et al.’s benchmark model (model 1 in Buhaug). Of the four critical errors listed above, *i* invalidates Buhaug’s inferences based on Buhaug’s table 1 (denoted BT1), *ii* and *iii* invalidate Buhaug’s inferences based on Buhaug’s table 2 (denoted BT2), and *i*, *ii*, *iii*, and *iv* invalidate Buhaug’s inferences based on Buhaug’s table 3 (denoted BT3). We replicate and examine BT1–BT3 separately and respectively in Tables 1–3 of this article.

In BT1, Buhaug (8) removes country-fixed effects and country-specific trends from the statistical model used in Burke et al. (1), terms that were included in the original analysis to account for all time-invariant and linearly trending confounding variables. Buhaug infers from BT1 that the finding of Burke et al. is incorrect because the parameter estimates of the model change in magnitude and in significance; however, this interpretation does not logically follow from the results in BT1. Drawing inferences

Significance

Whether climatic changes affect civil conflicts has been the subject of intense academic debate. Much of this controversy originates from a highly cited dispute between a previous PNAS paper—which finds that civil war incidence in sub-Saharan Africa is associated with increasing local temperature—and a subsequent rebuke of this result, also published in PNAS. We reexamine this apparent disagreement by comparing the statistical models from the two papers using formal tests. When we implement the correct statistical procedure, we find that the evidence presented in the second paper is actually consistent with that of the first. We conclude that the original grounds for the dispute over whether the climate–conflict relationship exists were erroneous.

Author contributions: S.M.H. and K.C.M. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

See Commentary on page 2054.

¹S.M.H. and K.C.M. contributed equally to this work.

²To whom correspondence should be addressed. E-mail: kmeng@princeton.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1316006111/-DCSupplemental.

from a model without fixed effects and trends requires stronger assumptions about African countries than were originally assumed by Burke et al. Specifically, it assumes that the average rate of conflict is the same for all countries in sub-Saharan Africa and that the conflict rate do not exhibit a trend (common or country-specific) over time. Thus, Buhaug's approach in BT1 implicitly requires the assumption that all countries in sub-Saharan Africa are comparable over space and time in the factors that influence conflict risk, which may include cultural patterns, geopolitics, natural resources, colonial history, international trade patterns, government policies, and geographic constraints. This assumption is not tested by Buhaug. A formal test of Buhaug's assumption that countries have the same average risk of conflict and/or trends in conflict (i.e., country-fixed effects and/or country-specific trends all equal zero) is the calculation of a F-statistic that jointly tests whether average conflict risk and/or trends in conflict are statistically different from zero (13, pp. 150–157). In Table 1 we replicate the results in BT1 and conduct this test on the relevant terms in the benchmark model of Burke et al. We reject the null hypotheses that country-fixed effects are jointly zero ($P < 0.0001$), that country-specific trends are jointly zero ($P < 0.0001$), and that country-fixed effects and country-specific trends are jointly zero ($P < 0.0001$). (Because of limited degrees of freedom, we are mechanically unable to conduct joint F-tests with SEs clustered at the country level. However, estimates using country-clustered SEs are nearly identical in magnitude to heteroscedastic-robust SEs. By necessity, we use these latter values to conduct a F-test. The replication file is provided in [Supporting Information](#).) These results strongly reject the assumption that models 2–4 in BT1 are correctly specified. The misspecification of these models is likely to alter Buhaug's parameter estimates relative to the Burke et al. finding regardless of whether or not the Burke et al. result is correct, so the observation that these parameter estimates differ from Burke et al. does not provide grounds to reject the Burke et al. result.

In BT2, the definition of conflict used in Burke et al. (1), the incidence of war years, is compared with models that examine alternative definitions of conflict incidence and various measures of conflict outbreak. No justification is provided for why these alternative variables should respond to temperature similarly to the incidence of war years. However, if we assume that such a justification exists, then this comparison should be made properly. The coefficients presented in BT2 represent changes in the probability of observing a conflict event; however, the different

types of conflict events occur with very different likelihoods (Table 2) and thus must be standardized before changes in these likelihoods can be compared. The most likely form of conflict is termed “incidence 25+” in Buhaug (8) and occurs with an unconditional probability of 0.254 in the sample (roughly once every 4 y), whereas the least likely is termed “outbreak +1000” and occurs with probability 0.012 (roughly once every 100 y). Comparing changes in probability for events that differ this much in their underlying likelihood is an apples-to-oranges comparison, because a probability change of 0.01 for incidence 25+ is a 4% change in the average risk of this event, whereas a 0.01 probability change for outbreak +1000 is an 83% change. We correct for the large differences in the underlying risk of these different forms of conflict by converting Buhaug's results into units of percentage change in the likelihood of conflict per 1 °C change in temperature. We do this by dividing each dependent variable by its average likelihood of occurring before running the regressions in BT2, essentially converting outcomes into units of relative risk that permit a valid apples-to-apples comparison across different outcome measures (6). We replicate the results of BT2 in Table 2 following this standardization. Standardization makes it clear that the results in Buhaug exhibit very large uncertainties, with two 95% confidence intervals that span effects ranging from –100% to +100% per 1 °C. This high level of sampling variability might explain why none of these coefficients were statistically significant in Buhaug; however, this uncertainty must also be accounted for when comparing these results to Burke et al., as we do below.

In Table 2, we test whether the results in each of these models is different from the main result presented in Burke et al. (1) using seemingly unrelated regression (SUR), an approach that allows us to formally test whether or not two different regression models return statistically different results (14) (also 15, p. 153). Intuitively, this approach asks whether the “regression lines” describing the relationship between temperature and conflict in Burke et al.'s and Buhaug's (8) analyses are statistically different from one another, while taking into account the fact that the studies are using related conflict outcomes where disturbances may be correlated. This test is necessary because it is extremely unlikely that any two studies using different data sets will recover identical results, due to sampling variability, even if the true underlying relationship is the same for both studies—so observing that two regression results are not identical is not sufficient evidence to conclude that the underlying relationship is different. Instead, we must ask whether the difference in the

Table 1. Testing model validity

	Burke et al. (1) war years 1000+	Buhaug (8) model 2 war years 1000+	Buhaug model 3 war years 1000+	Buhaug model 4 war years 1000+
Temperature _t	0.043 (0.022)	–0.001 (0.019)	0.013 (0.030)	0.011 (0.016)
Observations	889	889	889	889
R-squared	0.657	0.341	0.460	0.012
Country-fixed effects	Yes	No	Yes	No
Country trends	Yes	Yes	No	No
<i>Testing if Additional Model Assumptions Are Rejected Using F-Test</i>				
Assumption[s]	All countries have same intercept		Countries have no trend	All countries have same intercept and no trend
F-test (P value)	0.0000		0.0000	0.0000

This table replicates Buhaug table 1. The first column replicates Burke et al.'s benchmark model, which is also model 1 in Buhaug. Models 2–4 in Buhaug remove elements of this model (country-specific fixed effects and/or trends), implicitly making new assumptions about the data's structure. Bottom panel presents *P* values from a F-test under a null hypothesis that the omitted fixed effects and/or trends are all equal to zero in the benchmark model. Standard error in parentheses.

Table 2. Testing for disagreement between results when alternative conflict variables are used

	Burke et al. (1) war years 1000+ (standardized)	Buhaug (8) model 5 incidence 1000+ (standardized)	Buhaug model 6 outbreak 1000+ (standardized)	Buhaug model 7 incidence 25+ (standardized)	Buhaug model 8 outbreak 25+ (standardized)	Buhaug model 9 outbreak 100+ (standardized)
Probability of occurrence	0.110	0.190	0.012	0.254	0.052	0.030
Temperature _t	0.390 (0.197)	-0.030 (0.110)	-0.408 (1.046)	0.060 (0.156)	-0.165 (0.504)	0.532 (0.790)
Temperature _{t-1}	0.120 (0.211)	-0.130 (0.147)	-0.755 (1.233)	-0.121 (0.128)	-0.083 (0.505)	-0.598 (0.581)
Precipitation _t	-0.209 (0.471)	0.326 (0.318)	-1.001 (4.212)	0.508 (0.281)	1.065 (1.316)	-0.455 (2.465)
Precipitation _{t-1}	0.227 (0.443)	0.296 (0.324)	0.205 (2.847)	0.093 (0.271)	0.352 (1.370)	-0.321 (2.017)
Observations	889	889	889	889	889	769
R-squared	0.657	0.765	0.090	0.652	0.130	0.099
<i>Testing Whether Coefficients Differ from Burke et al. using SUR (P value)</i>						
Temperature _t		0.0558	0.4388	0.1299	0.2638	0.8558
Temp _t , temp _{t-1}		0.1392	0.4563	0.1598	0.4276	0.3700
All four variables		0.1290	0.2843	0.1453	0.4429	0.4333

This table replicates Buhaug table 2. All regressions contain country fixed effects and country-specific trends with standard errors clustered by country, shown in parentheses. The unconditional probability of occurrence is shown and is used to standardize each conflict outcome. For regression coefficients shown, a 0.1 effect implies a 10% change relative to average risk levels. We estimate Buhaug models 5–9 simultaneously with the Burke et al. model using seemingly unrelated regression (SUR) to test a null hypothesis that coefficients from the two models are the same in bottom panel.

regression results is large enough that it is unlikely to be caused by sampling variability alone. This approach differs markedly from the analysis in Buhaug, where it was claimed that new results overturned the findings of Burke et al. without conducting any formal statistical comparisons between the two sets of models. We correct this error by testing whether the effect of current temperature, the effects of current and lagged temperature (jointly), and the effects of all weather variables (jointly) on alternative conflict measures from Buhaug differ significantly from the result reported in Burke et al. As shown in Table 2, of these 15 comparisons, only one (the effect of current temperature on incidence 1000+) is marginally significant at the 10% level. Thus, we fail to reject the hypothesis that Buhaug’s results are different from Burke et al.’s result because the magnitude of observed differences would be expected based on sampling variability alone. We conclude that once the variables in BT2 are standardized and intermodel comparisons are correctly implemented, the results in BT2 provide no support for the claim that the results of Buhaug are different from those of Burke et al.

In BT3, a logistic regression lacking country-fixed effects and country-specific trends is presented using an alternative measure of conflict. The effect of temperature is again not directly compared with the results of Burke et al. (1). Furthermore, the effect of temperature is displayed in raw coefficients from a nonlinear logit regression, which are difficult to interpret in terms of conflict risk and across conflict measures with different underlying likelihoods. To make the effect on this alternative conflict measure comparable and interpretable in terms of conflict risk, we replicate the main results from BT3 but report the estimated effects of

temperature in terms of relative risk ratios in Table 3. This conversion makes it immediately clear that the results reported in BT3 are not statistically different from the benchmark result in Burke et al., because the Burke et al. result is contained within all four confidence intervals reported in BT3. Moreover, under all four models the upper bound of the 95% confidence interval for the effect of +1 °C dramatically exceeds the estimate reported in Burke et al., indicating that the relative risk for conflict may rise as much as 547 (model 12, the lowest upper bound) and 14.5 duodecillion (model 13, the highest upper bound). We think it is implausible that the true effect of temperature on conflict can be as large as these upper bounds suggest. Rather it seems more likely that the models in BT3 are not properly specified. Regardless, these estimates do not indicate that the relative risk ratio of 1.39 implied by Burke et al. is too large.

Discussion

We find that the disagreement in findings between Buhaug (8) and Burke et al. (1) vanishes after applying the appropriate set of model selection and model comparison tests. The large statistical uncertainty reported in Buhaug causes the results to not be statistically different from the findings reported in Burke et al. Furthermore, the high statistical uncertainty reported in Buhaug indicates that Buhaug’s statistically precise conclusion “Climate not to blame for African civil wars” (8, p. 16477) is inconsistent with the evidence presented.

It is important to note that our findings neither confirm nor reject the results of Burke et al. (1). Our results simply reconcile the apparent contradiction between Burke et al. and Buhaug (8)

Table 3. Relative risk ratio from +1 °C

	Burke et al. (1) implied war years 1000+	Buhaug (8) model 10 outbreak 25+	Buhaug model 11 outbreak 25+	Buhaug model 12 outbreak 25+	Buhaug model 13 outbreak 25+
Upper bound effect (95% CI)		8.62 × 10 ⁶	7.10 × 10 ⁴	546.9	1.45 × 10 ⁴⁰
Average effect of temperature	1.39	0.0199	3.27 × 10 ⁻⁶	5.73 × 10 ⁻⁹	2.46 × 10 ⁻⁵⁷
Lower bound effect (95% CI)		4.60 × 10 ⁻¹¹	1.51 × 10 ⁻¹⁶	6.01 × 10 ⁻²⁰	4.20 × 10 ⁻¹⁵⁴

This table replicates Buhaug table 3. Estimates are relative risk ratios from +1 °C. Models described in Buhaug. CI, confidence interval.

by demonstrating that Buhaug does not provide evidence that contradicts the results reported in Burke et al. Notably, however, other recent analyses obtain results that largely agree with Burke et al. (2–5), so we think it is likely that analyses following our approach will reconcile any apparent disagreement between these other studies and Buhaug.

Finally, we argue that the statistical procedures and reasoning used to obtain our conclusions are broadly applicable and should form the basis for future comparisons between statistical findings in applied research. As such, formal statistical tests must be used in order for a new study to overturn a previous result.

Materials and Methods

Data and regression models are detailed in Burke et al. (1) and Buhaug (8). Replication code and data are available in the [Supporting Information](#).

The benchmark model in Burke et al. (1) (model 1 in Table 1) is

$$y_{it} = c + \beta T_{it} + \gamma T_{i,t-1} + \delta P_{it} + \eta P_{i,t-1} + \mu_i + \theta_i year_t + \epsilon_{it} \quad [1]$$

where y_{it} is 1 if a conflict occurs in country i in year t and zero otherwise, T_{it} is current local temperature, $T_{i,t-1}$ is lagged local temperature, P_{it} is current local precipitation, $P_{i,t-1}$ is lagged local precipitation, μ_i is a vector of country-specific constants (fixed effects), θ_i is a vector of country-specific time trends, and ϵ_{it} are disturbances. In Table 1, we test the assumption of Buhaug (8) models 2–4 that components of Eq. 1 are not statistically significant. For model 2 we test the assumption that $\mu_i = 0$ for all i , for model 3 we test the assumption that $\theta_i = 0$ for all i , and for model 4 we test the assumptions that both $\mu_i = 0$ and $\theta_i = 0$ for all i . Each test is implemented using a F-test that allows us to simultaneously test these multiple-condition hypotheses.

Buhaug (8) models 5–9 in Table 2 are directly compared against the benchmark model in Burke et al. (1) using SUR. This procedure allows cross-model restriction tests when there may be correlation in the error structure across models and/or samples (14, 15). This procedure is necessary because it is likely that the various conflict outcomes considered in BT2 by Buhaug are closely related events and thus may induce correlation across samples and fitted models. For example, a country-year event passing the 1,000 battle-deaths threshold used in Burke et al. will necessarily pass the 25 battle-deaths threshold used in Buhaug model 7. By using SUR, we characterize the extent of correlation between models and account for this structure when determining whether two models provide statistically different results. In Table 2 we compare parameter estimates using the Burke et al. benchmark definition of conflict y_{it}^1 and alternative definitions y_{it}^2 provided by Buhaug. Both outcomes are assumed to have similar underlying structure, although the parameters describing this structure may differ between outcomes

$$y_{it}^1 = c^1 + \beta^1 T_{it} + \gamma^1 T_{i,t-1} + \delta^1 P_{it} + \eta^1 P_{i,t-1} + \mu_i^1 + \theta_i^1 year_t + \epsilon_{it}^1 \quad [2]$$

$$y_{it}^2 = c^2 + \beta^2 T_{it} + \gamma^2 T_{i,t-1} + \delta^2 P_{it} + \eta^2 P_{i,t-1} + \mu_i^2 + \theta_i^2 year_t + \epsilon_{it}^2 \quad [3]$$

and we are interested in testing the following three hypotheses:

$$H1: \beta^1 = \beta^2 \quad [4]$$

$$H2: \beta^1 = \beta^2, \gamma^1 = \gamma^2 \quad [5]$$

$$H3: \beta^1 = \beta^2, \gamma^1 = \gamma^2, \delta^1 = \delta^2, \eta^1 = \eta^2 \quad [6]$$

while also accounting for the fact that ϵ_{it}^1 and ϵ_{it}^2 are correlated. H1 tests whether the current temperature term is different across the two models. H2 jointly tests for differences across models for the current and lagged temperature terms while H3 jointly tests for differences across models for all four weather terms. To test H1–H3 while accounting for this cross-outcome correlation using SUR, we stack two samples (one from each study) and estimate a single equation

$$y_{it}^{1 \cup 2} = c^1 + c^D D + \beta^1 T_{it} + \beta^D T_{it} D + \gamma^1 T_{i,t-1} + \gamma^D T_{i,t-1} D + \delta^1 P_{it} + \delta^D P_{it} D + \eta^1 P_{i,t-1} + \eta^D P_{i,t-1} D + \mu_i^1 + \mu_i^D D + \theta_i^1 year_t + \theta_i^D year_t D + \epsilon_{it}^1 \quad [7]$$

where D is a dummy variable equal to 1 if an observation comes from the second sample (Eq. 3), and the vector of outcomes is

$$y_{it}^{1 \cup 2} = \begin{bmatrix} y_{it}^1 \\ y_{it}^2 \end{bmatrix} \quad [8]$$

which enables us to test H1 by examining the statistical significance of β^D and likewise for H2 and H3. Our estimated uncertainty in the estimate of β^D accounts for correlation between ϵ_{it}^1 and ϵ_{it}^2 .

Buhaug (8) models 10–13 in Table 3 are directly compared against the benchmark model in Burke et al. (1) by converting the marginal effect estimated in a logistic regression into relative risk ratios

$$\frac{\Delta y + \bar{y}}{\bar{y}} \quad [9]$$

where Δy is the change in conflict risk for a +1 °C warming and \bar{y} is the average conflict risk in the sample.

ACKNOWLEDGMENTS. We thank Halvard Buhaug and Marshall Burke for comments and supplying replication files. We thank Edward Miguel, Amir Jina, and James Rising for comments.

- Burke MB, Miguel E, Satyanath S, Dykema JA, Lobell DB (2009) Warming increases the risk of civil war in Africa. *Proc Natl Acad Sci USA* 106(49):20670–20674.
- O'Loughlin J, et al. (2012) Climate variability and conflict risk in East Africa, 1990–2009. *Proc Natl Acad Sci USA* 109(45):18344–18349.
- Harari M, La Ferrara E (2012) Conflict, climate and cells: A disaggregated analysis. *IGIER 461* (Univ Bocconi, Milan, Italy).
- Hsiang SM, Burke M, Miguel E (2013) Reconciling Temperature-conflict Results in Kenya. CEGA Working Paper Series (Center for Effective Global Action, University of California, Berkeley), No 032. Available at <http://www.escholarship.org/uc/item/9ct8g2zr>. Accessed December 18, 2013.
- Hsiang SM, Meng KC, Cane MA (2011) Civil conflicts are associated with the global climate. *Nature* 476(7361):438–441.
- Hsiang SM, Burke M, Miguel E (2013) Quantifying the influence of climate on human conflict. *Science* 341(6151):1235367, 10.1126/science.1235367.
- Hsiang SM, Burke M (2013) Climate, conflict and social stability: What does the evidence say? *Clim Change*, 10.1007/s10584-013-0868-3.
- Buhaug H (2010) Climate not to blame for African civil wars. *Proc Natl Acad Sci USA* 107(38):16477–16482.
- Bernauer T, Siegfried T (2012) Climate change and international water conflict in Central Asia. *J Peace Res* 49:227–239.
- Scheffran J, Brzoska M, Kominek J, Link PM, Schilling J (2012) Climate change and violent conflict. *Science* 336(6083):869–871.
- Gleditsch N (2012) Whither the weather? Climate change and conflict. *J Peace Res* 49(1):3–9.
- Solow AR (2013) Global warming: A call for peace on climate and conflict. *Nature* 497(7448):179–180.
- Wooldridge JM (2006) *Introductory Econometrics: A Modern Approach* (South-Western College Publishing, Cincinnati, OH), 3rd Ed.
- Clogg CC, Petkova E, Haritou A (1995) Statistical methods for comparing regression coefficients between models. *Am J Sociol* 100(5):1261–1312.
- Wooldridge JM (2010) *Econometric Analysis of Cross Section and Panel Data* (MIT Press, Cambridge, MA), 2nd Ed.