Is Temperature Exogenous?

The Impact of Civil Conflict on the Instrumental Climate Record in Sub-Saharan Africa

Kenneth A. Schultz
Dept. of Political Science
Stanford University

Justin Mankin
Department of Geography, Dartmouth College
Lamont-Doherty Earth Observatory of Columbia University

Draft Date: November 2, 2018

ABSTRACT

Research into the effects of climate on political and economic outcomes assumes that short-term variation in weather is exogenous to the phenomena being studied. However, weather data are derived from stations operated by national governments, whose political capacity and stability affect the quality and continuity of coverage. We show that civil conflict risk in Sub-Saharan Africa is negatively correlated with the number and density of weather stations contributing to a country's temperature record. This effect is both cross-sectional—countries with higher average conflict risk tend to have poorer coverage—and cross-temporal—civil conflict leads to loss of weather stations. Poor coverage induces a small downward bias in one widely used temperature data set, due to its interpolation method, and increases measurement error, potentially attenuating estimates of the temperature-conflict relationship. Combining multiple observational data sets to reduce measurement error almost doubles the estimated effect of temperature anomalies on civil conflict risk.

We would like to acknowledge helpful comments on earlier drafts from Frances Moore, Zeke Hausfather, and the anonymous reviewers. We also benefited from the advice and assistance of Tufa Dinku of Columbia University, Karen McCourt of the United Kingdom Meteorological Office, Bernard Gomez of the World Meteorological Organization, and Andrew Lederman of the United States Department of State.

A growing body of research has examined the effect of climate variation on political and economic outcomes, including the incidence of violent conflict (for reviews, see, Carleton and Hsiang 2016; Hsiang, Burke, and Miguel 2013). In particular, a number of studies have found that high temperature anomalies are associated with an elevated risk of civil conflict, particularly in the developing world (Burke, Miguel, Satyanathm, Dykema, and Lobell 2009; O'Loughlin, Witmer, Linke, Laing, Gettelman, and Dudhia 2012; O'Loughlin, Linke, and Witmer 2014; Bollfrass and Shaver 2015; Linke, O'Loughlin, Gettelman, and Laing 2017; Maystadt, Calderone, and You 2015). A central assumption underlying these studies is that the climate variables of interest—e.g., temperature, precipitation, drought—are exogenous to the outcome being explained. For this reason, weather shocks are increasingly used as instrumental variables in models of civil conflict and political instability (see, e.g., Miguel, Satyanath, and Sergenti 2004; Dube and Vargas 2013; Ritter and Conrad 2016).

However, while actual temperature and precipitation are not affected by economic and political outcomes—at least in the short run—their measurement can be. The most extensive modern records are derived from readings taken at weather stations distributed unevenly around the globe, and the tasks of establishing, staffing, and maintaining these stations fall under the jurisdiction of national governments.[1] As a result, political conditions could influence the instrumental record in at least two ways. First, the ability to establish and maintain weather stations may be related to the state's governing and bureaucratic capacity, factors that have been shown to influence a variety of outcomes, including civil conflict and economic growth (Fearon and Laitin 2003; Besley and Persson 2010; Hendrix 2010). Second, violence and instability may

[1] We discuss the availability of satellite-based observations in the conclusion.

lead to the destruction of facilities or divert government resources away from the collection of weather data, creating gaps in the record that are directly caused by the outcome of interest. For example, when the Central African Republic fell into civil war in December 2012, nine of its twelve weather stations participating in the Global Climate Observing System (GCOS) Surface Network (GSN) stopped reporting within months.[2] Reports from the country indicate that the stations were destroyed by fighting and the staff forced to flee.[3]

Given the importance of research into climate impacts, we need to know whether and how the climate record might be influenced by political and economic outcomes. This is particularly pressing given the fact that the number of weather stations has declined globally in recent decades, particularly in sub-Saharan Africa. Although estimates of local meteorological conditions do not require the presence of weather stations in the immediate vicinity, or even within the same country, fewer stations lead to greater reliance on interpolation and thus greater potential for bias or measurement error (Dell, Jones, and Olken 2014, 747–50). Previous work has shown that station loss has no appreciable impact on estimates of global average temperature (Lawrimore, Menne, Gleason, Williams, Wuertz, Vose, and Rennie 2011, 16–17), though there is concern that coverage gaps in polar regions have downwardly biased some estimates of recent warming (e.g., Cowtan and Way 2014). But there has been no examination of the causes of

[2] Station performance data from http://www1.ncdc.noaa.gov/pub/data/gcos/WW_REG1_POR_summary (accessed Sept. 28, 2017).

[3] Athanase Yambele, Director of Meteorology and Hydrology, Central African Republic, email message to author, September 27, 2017.

station loss or whether coverage gaps affect the high resolution climate data used in research on political and economic impacts.[4]

In this paper, we examine the effect of civil conflict risk on the instrumental temperature record in sub-Saharan Africa (SSA), the region that has been the subject of most of the scholarly attention (Burke et al. 2009, 2010; Couttenier and Soubeyran 2014; O'Loughlin et al. 2012; O'Loughlin et al. 2014; Miguel et al. 2004; Linke et al. 2017; Maystadt, Calderone, and You 2015).[5] We establish four main results. First, civil conflict risk is negatively associated with the number and density of weather stations contributing to a country's temperature record. There is evidence of both a cross-sectional effect—i.e., countries with higher conflict risk tend to have poorer coverage—as well as a cross-temporal effect—i.e., the incidence of civil conflict leads to loss of weather stations. Second, the most severe coverage gaps are associated with a small downward bias in estimated temperature anomalies in the high resolution temperature series generated by the University of East Anglia's Climatic Research Unit (CRU), the data set used in six of the eight studies cited above. This bias is due to manner in which those data deal with areas of sparse coverage to create a gridded product. Third, station coverage gaps are also

[4] An exception is Dell, Jones, and Olken (2012, appendix table 15) who show that coverage is not influenced in the short-run by economic growth or leadership turnover, their two main dependent variables.

[5] Other studies have examined the effects of climate shocks on other forms of violence, such as interpersonal violence, crime, and interstate conflict (see Hsiang et al. 2013). We focus on civil conflict since it is the most common outcome of interest in the political science literature.

associated with greater measurement error. Comparing four observational data sets, we show that variation in temperature estimates is greater in more conflict-prone states.

Collectively, these results imply that estimates of the effect of temperature on conflict risk could *understate* the magnitude of that relationship. The final result establishes the plausibility of this conjecture by re-estimating the relationship between temperature and civil conflict in SSA using a regression calibration approach designed to reduce measurement error when multiple mis-measured proxies are available (Carroll, Ruppert, Stefanski, and Crainiceanu 2006). Doing so yields an estimated effect that is almost twice what we find using any of the individual temperature data series without correction.

These results have direct implications for work on climate and conflict, as well as for efforts to accurately estimate the impacts of climate change, including the social costs of carbon emissions. In addition, this paper contributes to a growing body of research into how political conditions affect the data generating process behind commonly used statistics (Jerven 2013; Hollyer, Rosendorff, and Vreeland 2014; Kelley and Simmons 2015; Merry 2016; Lee and Zhang 2016). In some applications, missingness or poor data quality can be usefully leveraged to identify gaps in state capacity or incentives to conceal information. But the larger concern is that resulting data are missing, noisy, or biased in ways that are correlated with a host of political outcomes. In the case of research that uses weather as an instrumental variable, the results of this paper raise a caution flag. The message is *not* that weather cannot be used as an instrument; rather, it is that researchers should not assume that the exogeneity assumption holds. As always, a careful understanding of the data and how they are generated is needed to ensure that assumption is valid.

1. The Effect of Conflict on Coverage

Governments collect weather data to support a variety of functions, including agriculture, aviation, construction, science, and defense. The primary tasks of meteorological data collection—including building, maintaining, staffing, and inspecting weather stations—are overseen by national meteorological and hydrological centers. Stations themselves vary substantially in terms of the technology of the instruments, the type of observations they make, and how the data are recorded. Weather monitoring at an airport, for example, is more sophisticated than a thermometer and rain gauge at a school, where readings might be written on paper by a volunteer. At the international level, the World Meteorological Organization (WMO) supports national governments and creates standards, particularly for stations used for long-term climatological monitoring (World Meteorological Organization 2011). Thus, global data sets used in most research incorporate observations from the subset of stations that meet the necessary standards of quality, precision, and longevity. These can be expensive to create and maintain and require either reliable automation or operation by trained personnel (World Meteorological Organization 2008).

Patterns of Weather Station Coverage

Figure 1 shows the number of stations per year in SSA in the period 1900-2016. Station data are derived from two sources: the CRU high resolution times-series data (Harris, Jones, Osborn, and Lister 2014) and the monthly gridded data from the Berkeley Earth surface temperature project (BEST) (Rohde, Muller, Jacobsen, Perlmutter, and Mosher 2013).[6] We focus on these two data sets because of their widespread use in the literature on climate impacts and because both projects make the underlying station data easily available. The two sources exhibit significant overlap, but they have different standards for station inclusion. The CRU data are most restrictive because they require a station to have sufficient coverage in the 1961-90 period to establish the local climatology against which anomalies are estimated. The BEST data include the CRU stations but also draw on additional sources and use techniques that allow for the inclusion of stations with relatively short reporting periods. For each data source, two series are drawn in Figure 1: solid lines represent the number of stations that reported a valid temperature in all 12 months of the year, and dashed lines count the number of stations that reported at least one monthly temperature in that year.

--- FIGURE 1 ABOUT HERE---

Both series tell a similar story. Weather station coverage in SSA grew sharply in the decade after World War II, reaching its maximum during the 1960s, when most countries in this region became independent states. From that point on, there is a decline in the number of stations, particularly those that report for the full year. The decline is less pronounced in the BEST data, which experience a recovery in numbers, though much of that growth is due to a single country, South Africa.

This pattern broadly mirrors global trends, but SSA has experienced both the poorest coverage overall and the deepest losses (United Nations Economic Commission for Africa

[6] The CRU data are version 3.25, which covers 1901-2016. The BEST data are continually updated; the data were accessed May 4, 2017.

2011). The decline has multiple causes. Concerted efforts to retrospectively fill in historical data, combined with the fact that many stations do not report in real time, can inflate past observations relative to current ones. The drop after 1990 was due in part to the collapse of the Soviet Union, which supported weather data collection in several regions (Rohde et al. 2013, 7–8; Dell et al. 2014, 747–48). The pattern also reflects actual station loss as well as declining performance, evident in the growing gap between the number of stations that report at least one monthly average and those that report the full year. This gap suggests that many stations that physically exist either fail to record a temperature in some months, fail to report the data, or record a temperature that is discarded due to quality concerns.[7]

Crucially, the decline in weather stations in SSA is not uniform. Figure 2 depicts the location of all weather stations reporting temperature in the CRU (panel a) and BEST (panel b) data since 1946. Those indicated with a solid star reported for at least one month in the period 2010-16, while hollow stars identify stations that did not report in those years. The map shows significant regional disparity in both existing and defunct stations. Coverage was historically densest in French West Africa and in South Africa, although both regions have seen recent declines. Coverage is sparser in central Africa, particularly the zone running from Angola northeast through the Congo to Somalia and north to Chad. Notably, some countries lack a single station in the CRU data, including Nigeria, Uganda, and Botswana. The additional coverage available in the BEST data is also noticeable, though they exhibit similar regional variation in station density.

[7] Both CRU and BEST apply a variety of quality control criteria to screen out temperature reports that are highly anomalous relative to past readings and/or readings in nearby stations.

--- FIGURE 2 ABOUT HERE---

A country does not need to have a weather station within its borders for researchers to estimate its temperature, as monthly temperature anomalies are significantly correlated at distances of 1200km or more. Figure 3 shows, for each 0.5° grid cell, the number of CRU stations within 1200km of the cell's centroid that reported temperature in January of the years 1985, 1995, 2005, and 2015. Since 1990, areas of sparse coverage opened up from the Horn of Africa to the southwestern coast. Notably, most of the affected countries—Angola, Rwanda, Burundi, Somalia, Uganda, Ethiopia, and the Democratic Republic of the Congo—experienced civil conflict in at least half of years since 1990.

--- FIGURE 3 ABOUT HERE---

Conflict risk and country coverage

To what extent might variation in station coverage be explained by the underlying risk or actual incidence of civil conflict? We define civil conflict as organized political violence between the government and one or more rebel groups that claims at least 25 battle-related deaths in a year. The data for identifying such conflicts comes from the Uppsala Conflict Data Program (UCPD) Armed Conflict Database, version 17.1 (Gleditsch, Wallensteen, Eriksson, Sollenberg, and Strand 2002; Allansson, Melander, and Themnér 2017). Figure 4 presents evidence of a cross-sectional association between conflict risk and weather station coverage based on two kinds of indicators. [8] First, two in-country indicators count the number of weather

[8] Whereas weather station and temperature data exist for all countries and years regardless of when they became independent, civil conflict data pertain to independent states. The sample for

stations located within the country reporting in a given year, normalized per 100,000 sq. km of country area. Second, two distance-based indicators indicate the average number of stations within 1200km of each grid cell in the country, based on the grid cell resolution of the respective data sets.[9] For each indicator, the figure plots each country's average level of coverage over its history against the proportion of years that the country experienced a civil conflict in the period 1946-2016. In each case, the relationship is negative.

--- FIGURE 4 ABOUT HERE---

Two broad mechanisms could contribute to this negative correlation. First, there could be country-level factors—such as poverty, poor state capacity, inhospitable terrain, or low bureaucratic quality—that both make a state vulnerable to conflict and compromise its ability to establish and maintain stations. Second, conflict itself could cause station loss or performance problems either due to direct damage caused by violence or the diversion of government resources away from station staffing and maintenance. Since some station reports are physically collected by international scholars, both of these factors could also create variation in data accessibility, a phenomenon that Hendrix (2017) has documented in a related context.

---

these tests is constructed from country-years in the period 1946-2016 during which a country was independent. Ethiopia is treated as a different state before and after the secession of Eritrea in 1993 and is indicated in the figures by the label "ETH93." Observations on Sudan after the secession of South Sudan (2011) are dropped.

[9] The CRU data use 0.5° grid cells, while the BEST data use 1° cells. Yearly counts are weighted by the proportion of months in the year that the station made valid report.

Before turning to an exploration of these possibilities, it is important to note some complications with analysis of the coverage data. First, the majority of stations in these data sets do not report temperatures in real (or near real) time. Reports are collected retroactively, often with some delay. For example, many stations are updated via the World Weather Records, a compilation that is published once a decade. A station that disappears sometime in a given decade might appear to be lost from the beginning of the decade, a conjecture attested to by drops in station counts at the turn of each decade. Second, there have been continual efforts to add stations to the existing data sets, which means that station counts often increase retroactively from one data release to the next. Thus, at any given time, the existing station count imperfectly captures the true total. Third, to be included in the CRU data, a station had to report consistently through the period 1961-90. As a result, instability-related station loss during this period extends both forward and backward in time, and recently opened stations cannot be included in these data. There would also be survivorship bias if stations that managed to persist through those decades are, for whatever reason, relatively robust. In principle, the BEST station data do not face this latter problem, but to the extent that the project draws on CRU and other sources with similar criteria, it is no wholly immune.

As a result, these data cannot be used to develop a fully-specified model of country coverage, nor do they accurately capture temporal dynamics. We thus undertake the more modest exercise of exploring the correlations that exist between coverage and indicators of or risk factors for civil conflict. Table 1 presents the coefficients from bivariate regressions of each of the four country-year coverage measures on each of a number of indicators of conflict incidence and state capacity. Three measures of civil conflict are considered: a contemporaneous indicator for whether the country experienced a civil conflict in a given year; a cross-sectional measure of the number of years in the period 1961-90 the country experienced civil conflict, to account for the effect of instability in the baseline period; [10] and a regional indicator for whether any country within 1200km was experiencing a civil conflict in that year, which captures both regional instability and coverage effects of weather stations in neighboring states. The table also explores correlations with several characteristics related to state capacity (Hendrix 2010):

- Socioeconomic development: Real GDP per capita (logged) and the rate of infant mortality per 1000 live births.[11]
- Extractive capacity: Total government revenue as a percentage of GDP.[12]
- Governance: A measure of bureaucratic quality from the International Country Risk Guide (ICRG), a measure of government effectiveness developed by the World Bank's World Governance Indicators project (Kraay, Kaufmann, and Mastruzzi 2010), and an index developed by Hollyer, Rosendorff, and Vreeland (2014)

[10] Some countries experienced civil conflict on their territory prior to becoming independent (e.g., Eritrea). These conflicts are captured in the UCDP data, which codes the location of separatist conflicts. These cases are included in the count of 1961-90 conflicts. Neighbors are included if any part was located within 1200km of the country's centroid.

[11] Both are from the World Bank's World Development Indicators.

[12] From the International Centre for Tax and Development, "Government Revenue Dataset," 2018 release (Prichard, Cobham, and Goodhall 2014).

measuring the government’s performance in reporting economic data, which can reflect both transparency and capacity.[13]

- Country features: Indicators for population density, mean elevation, and area (logged).

To aid in comparison of coefficients, all independent variables were standardized. The final column shows the correlation between each independent variable and the incidence of conflict in the country.

--- TABLE 1 ABOUT HERE---

Several patterns stand out. First, there is strong negative correlation between all three indicators of conflict—contemporaneous, historical, and regional—and all station coverage indicators. The correlations with the governance indicators are in the expected (positive) direction, and some are statistically significant. There is also a relationship between mountainous terrain (proxied by mean elevation) and poorer coverage. Associations with other indicators, including economic development, infant mortality, and extractive capacity, are inconsistently and/or unexpectedly signed. These results suggest that variation in coverage is associated with civil conflict incidence as well as with several factors that are known to contribute to conflict risk, especially conflicts in neighboring states (Buhaug and Gleditsch 2008), mountainous terrain (Fearon and Laitin 2003), and ineffective governance.

Conflict and station loss

[13] ICRG data are available for 1984-2016 and missing for 13 countries in the data set. The transparency index is available for 1980-2010. The WGI data are available for 1996-2016.

For reasons discussed earlier, the coverage data are not ideal for understanding cross-temporal dynamics and do not tell us whether conflict directly leads to station loss. To examine this, we turn to station performance data collected by the GCOS. These data report, for each station in the network, the number of hourly observations reported by that station in each month and whether the station reported a monthly summary using CLIMAT, an electronic reporting system. Although some data go back to 1948, there is a very significant missingness from 1961-72, including some years with no reports from any station. Thus, a station enters the sample the first time it makes a report after December 1972 or when the country in which it is located became independent, whichever is later. This yields a sample of 1169 stations.

The data are organized into spells of activity at the station-month level, allowing us to estimate the effect of civil conflict on the probability that such a spell will end. Specifically, we code each station as active in any month in which it made at least one hourly report or issued a CLIMAT summary. Since many stations have intermittent, short-term gaps in reporting, a station is coded as having died if it enters a spell of inactivity that lasts some minimum number of months. For the results reported in the text, a station is considered dead if it goes inactive for at least 24 months. If a station comes back to life after being dead for the minimum period, a new spell of activity begins. The dependent variable equals zero during active months and short-lived periods of inactivity and one in a month in which the station dies. The average monthly death rate is only 0.36 percent, but 58 percent of stations experienced at least one failure.

The main independent variables are indicators for whether or not the country in which the station was located experienced civil conflict in a given month. We operationalize this two ways. First, the UCDP data are used to generate an indicator for whether or not the country was in an episode of civil conflict. An episode of conflict begins when a conflict first hits the 25 battle death threshold and continues until it experiences a year of inactivity. Since onset and termination codings can be imprecise, we also use the UCDP Georeferenced Event Data (GED), version 17.1, which captures actual conflict events, as well as their location, in the period 1989-2016 (Sundberg and Melander 2013). This data set is used to create a series of station-month indicators for whether a violent event took place in the country and whether the closest such event took place 0-10km, 10-50km, 50-100km, or more than 100km from the station. These indicators enter the regressions with a one-month lag. We focus on events that are classified as "state-based violence," which implies fighting between the government and a rebel group.

Table 2 shows estimates from logit models applied to these data. In addition to the conflict variables, all models include a control for the age of the station (logged), a measure of how long the current spell of activity has been going on (introduced as a cubic polynomial), and country and year fixed effects.[14] To ease interpretation, the estimates are expressed as odds ratios, so numbers greater (less) than one imply an increased (decreased) risk of failure.

--- TABLE 2 ABOUT HERE---

Several results stand out. First, a civil conflict episode in the country substantially increases the risk that a station will die, with the estimates in column (1) and (3) implying a 70 or 100 percent increase in that risk in any month of conflict, respectively. Second, the estimates in column (2) show that violent events in the previous month are also associated with a heightened

[14] Station age measures how long a station with a given WMO identification has been reporting. It does not necessarily capture the age of the equipment, which is unknowable.

risk of station failure, with the effect increasing in magnitude and statistical significance the closer the event is. However, column (3) shows that, once we reintroduce the control for an ongoing conflict episode, the effect of violent events becomes smaller, and the coefficients are both individually and collectively insignificant. Although the estimates suggest that violence within 10km of a station has a larger effect than violence that is more than 100km away, we cannot reject the null hypothesis that all of the effects are zero (or 1 in terms of relative risk). The risk of death is also decreasing in station age, suggesting that stations that have been in operation longer are more robust, possibly due to their importance or location.

These findings imply that the effect of civil conflict is not simply due to direct damage from the conflict, since station loss can happen even in months without violent events. Clashes between the government and rebels also may also contribute to station failure, particularly when a station is close to active fighting, but that effect is less robust. These results suggests that stations fail primarily due to diversion of government resources and to a reduction in security in the vicinity, which can lead to their being abandoned or unmaintained. We also note that other kinds of violent events in the GED—rebel groups against another or one-sided violence against civilians—do not have a consistent effects on station loss. Additional tests varying the minimum length of inactivity for a station to be considered dead show that while conflict episodes are robustly associated with long failures lasting a year or more, proximate violence is associated with deaths of shorter duration (Supplementary Information, pp. 1-2).

2. The Effects of Coverage Gaps on Temperature Estimates

We have seen evidence of a negative correlation between weather station coverage and the risk of civil conflict. To understand the implications of this pattern for studies of the climate-conflict link, it is important to consider how variation in coverage could affect observational temperature data. There are two potential concerns: bias and measurement error.

Bias

In general, interpolation should not induce predictably signed biases unless station loss occurs in regions that are systematically warmer or cooler than those in which stations remain. There is no reason to believe that this is the case in SSA. The CRU data, however, rely on a practice that could induce bias. The vast majority of grid cell observations have to be interpolated, since few cells actually contain a weather station. In version 3.xx of the CRU data, this was done by interpolating from the three closest stations within 1200km. Where fewer than three stations were within that range, one or more "dummy" stations were created with a temperature anomaly (relative to 1961-90 baseline) of zero (Harris et al. 2014). Starting with the release of version 4.01 in 2017, CRU is shifting to a new method that allows observations to be influenced by as few as one station; cells with no stations within 1200km are assigned a zero anomaly.[15] In the context of a warming climate, this practice means that cells in poorly covered areas receive artificially low temperatures.

To analyze this bias, we focus on the version 3.xx data, since it was used in prior studies of the climate-conflict relationship, and the newer method is as yet less well documented. In the appendix, we confirm using cell-month level data that observations influenced by at least one dummy station are lower on average than observations that are not (SI, pp. 3-4). The bias is as

[15] For more information, see the release notes for version 4.01, on-line at http://data.ceda.ac.uk//badc/cru/data/cru_ts/cru_ts_4.01/Release_Notes_CRU_TS4.01.txt.

much as 0.3°C, which corresponds to about one-third of a standard deviation. For comparison, the BEST data display no such bias in poorly covered areas, suggesting that the effect is an artifact of the CRU methodology.

Here, we explore this bias at the country-year level. The dependent variable is the annual temperature anomaly in the country, and the key independent variable is a measure of the coverage gap, specifically the proportion of cell-month observations within the country that were influenced by at least one dummy station in that year. Since the bias should larger in warmer years, we interact the coverage gap measure with the average temperature anomaly in countries within 1200km, thereby capturing regional conditions. The model also controls for the regional temperature anomaly and country fixed effects. Standard errors are corrected for cross-sectional spatial dependence within 1200km and panel-specific serial correlation over 5 years using the method proposed by Conley (2008) and implemented by Hsiang (2010).

Figure 5 reports the effect of station coverage gaps on a country's annual temperature anomaly as a function of the regional temperature. Although only CRU uses dummy stations to fill in areas of poor coverage, we present the effect of CRU coverage gaps on both the CRU and BEST temperature anomalies for purposes of comparison. In the CRU data (solid line), coverage gaps are associated with higher temperature estimates when the regional temperature anomaly is 0.1°C or lower; however, once the regional temperature anomaly is above 0.7°C, coverage gaps induce a statistically significant downward bias on the country's estimated temperature anomaly. By contrast, the BEST data are not influenced by this measure, as the effect of a CRU coverage gap is small and indistinguishable from zero under all conditions (dashed line).

--- FIGURE 5 ABOUT HERE---

The predicted bias in the observed CRU data is relatively small but growing due to increasing temperatures and widening coverage gaps. Overall, at the country-year level, the proportion of cell-months affected by a dummy station ranges from 0 to 0.67, but the mean is only 0.014. At the average regional temperature anomaly of 0.30°C, a shift from no gap to one standard deviation above the mean (0.068) is associated with an insignificant upward bias of 0.010°C. However, by 2016, the mean coverage gap was 0.042, and the average regional temperature anomaly had grown to 0.78°C. Under these conditions, moving from no gap to one standard deviation above the mean (0.14) is associated with a bias of -0.066°C, or about 8 percent of the average temperature anomaly in this period. Thus, the bias caused by CRU's use of dummy stations is small but has gotten worse with time.

Measurement Error

Measurement error is likely to arise where station density is so low that estimates are not well constrained by direct observations. Greater reliance on interpolation means that temperature estimates are more dependent on the interpolation methods and assumptions. Moreover, estimates from areas that are densely covered by weather stations are less sensitive to differences in station inclusion criteria. To assess this possibility, we collect data from two additional temperature data sets based on the instrumental record: the National Climatic Data Center's Global Historical Climatology Network-Climate Anomaly Monitoring System (GHCN-CAMS) surface air temperature data (Fan and van den Dool 2008) and the Terrestrial Air Temperature Gridded Monthly Time Series by Willmott and Matsuura (2015). These data, like those from CRU, are reported at the 0.5° grid cell resolution. To make them compatible, the BEST data were downscaled to that resolution via bilinear interpolation. Due to limits on temporal coverage, the combined data set covers 1948-2014. The four temperature series rely on overlapping, though not identical, station data and use different interpolation methods. As a result, they are highly correlated, though not perfectly so. Pairwise correlations in the temperature anomalies range from 0.68-0.81 in the grid cell-month data and 0.75-0.87 when the data are aggregated to country years.

We can use these different data series to explore variation across space and time in the reliability of temperature estimates. Let $Temp_{it}^{*}$ denote the true temperature anomaly in location *i* at time *t*. This quantity is unobserved, but we have four estimates that are measured with error:

$$Temp_{it}^{j} = Temp_{it}^{*} + e_{it}^{j} \text{ with } j = 1,2,3,4. \quad (1)$$

For the sake of tractability, assume that the errors in any observation *it* are independently and identically distributed with mean zero and variance σ_{it}^2, which may vary by time and location depending on the coverage. Under this assumption, we can estimate the variance of the error as $\hat{\sigma}_{it}^2 = \frac{1}{3}\sum_{j=1}^{4}\left(Temp_{it}^{j} - \overline{Temp}_{it}\right)^2$, where $\overline{Temp}_{it}$ is the mean of the reported temperatures. If we let $\hat{v}_i^2 = \frac{1}{n-1}\sum_{t}\left(\overline{Temp}_{it}\right)^2$ estimate the natural variability of the temperature anomaly in each location *i*, the reliability of the anomaly measurement at location *i* and time *t* is $r_{it} = \frac{\hat{v}_i^2}{\hat{\sigma}_{it}^2 + \hat{v}_i^2}$.

This ratio runs from zero to one, with lower values indicating that the temperature is measured with more noise.

When the data are rendered as grid-cell months, the reliability ratio ranges from 0.11 to 0.99, with a mean of 0.80; across country-years, reliability ranges from 0.11 to 0.99, with a mean of 0.85. Figure 6 maps the average reliability in each grid cell for all months 2000-14, along with the locations of CRU weather stations that were active for at least one month in this period.

The low reliability scores in sparsely covered areas is quite apparent. The accompanying box plot shows how the distribution of reliability scores has evolved over time in the country-year data. The overall trend in reliability ratios is similar to the trend over time in weather stations that we saw in Figure 1. In the appendix, we report regressions confirming that reliability varies systematically with weather station coverage, both at the grid-cell month and country-year levels (SI, pp. 5-6).

--- FIGURE 6 ABOUT HERE---

Table 3 explores whether civil conflict risk is associated with greater measurement error in the county-year data using two sets of conflict indicators: contemporaneous indicators for whether a conflict was ongoing in the country or in a country within 1200km and long-run measures of the proportion of years a country experienced civil conflict and the proportion of years that a state within 1200km experienced civil conflict. For purposes of estimation, the dependent variable is $\ln \hat{v}_i^2 / \hat{\sigma}_{it}^2$, or the logged signal to noise ratio, which maps one-to-one onto reliability ratios. The models include a control for the average temperature across the four data sets, measures of the country's mean elevation, latitude (linear and squared), and area (logged), as well as year fixed effects to capture the overall downward trend in reliability. Standard errors are corrected for cross-sectional spatial dependence within 1200km and panel-specific serial correlation over five years using the method proposed by Conley (2008) and implemented by (Hsiang 2010).

--- TABLE 3 ABOUT HERE---

The good news is that reliability does not appear to be systematically correlated with contemporaneous conflict, either in the country or the neighborhood. It is also not correlated with the temperature anomaly, as long as year fixed effects are included. However, there is a pronounced cross-sectional effect, as reliability is systematically lower in conflict-prone states. Based on these estimates, moving average conflict in country from one standard deviation below to one standard deviation above the mean (0 to 0.5) reduces average predicted reliability in 2014 from 0.72 to 0.66.

3. Re-estimating the Effect of Temperature on Civil Conflict

All of the foregoing suggests that conflict risk increases noise in temperature estimates and, in the case of the CRU data, induces a small downward bias during periods of warming. To see how these problems can affect the estimated relationship between temperature shocks and conflict, consider the following benchmark model (Hsiang et al. 2013):

$$Conflict_{it}^{*} = \beta Temp_{it}^{*} + \upsilon_i + \eta_t + \varepsilon_{it} \quad , \tag{2}$$

where $Conflict_{it}^{*}$ denotes the (latent) risk of civil conflict in country *i* and year *t*, $Temp_{it}^{*}$ is the true temperature anomaly in that country-year, υ_i and η_t are country and year fixed effects, respectively, and ε_{it} is the error term. The essential problem is that we observe not the true temperature anomaly, $Temp_{it}^{*}$, but proxies, $Temp_{it}^{j}$, that are measured with error. Standard practice involves using one such data set to estimate the following model, which comes from plugging (1) into (2):

$$Conflict_{it}^{*} = \beta Temp_{it}^{j} + \upsilon_i + \eta_t + \mu_{it} \,. \tag{3}$$

where $\mu_{it} = \varepsilon_{it} - \beta e_{it}^{j}$. Let $\hat{\beta}^{j}$ denote the estimate of β from estimating (3) using temperature series *j*.

Standard results on measurement error tell us that that $\hat{\beta}^{j}$ will be biased towards zero, as noise attenuates the estimate due to the correlation between $Temp_{it}^{j}$ and μ_{it}. In a linear setting with classical measurement error, attenuation is a direct function of the reliability ratio, so at the average reliability score reported above, $\hat{\beta}^j$ would be expected to attenuated by about 15 percent. The dichotomous dependent variable in the present context, however, complicates that calculation, as do the non-classical features of the measurement error.

We have also seen that the variance of the measurement error, σ_{it}, is positively correlated with the mean conflict risk in the country, υ_i, creating cross-sectional heteroscedasticity in the disturbance terms, μ_{it}. While it is standard practice to calculate robust standard errors with clustering on the country, heteroscedasticity can lead to inconsistent estimates in models with a dichotomous dependent variable. Finally, the bias identified in the CRU data implies that $E(\mu_{it})$ is a function of both the conflict risk and the true temperature anomaly. In particular, the cold bias is larger the more conflict-prone the country is and the hotter the true temperature (Figure 5). This effect should induce a negative bias on $\hat{\beta}$ when estimated using these data. Collectively, then, our findings suggest that conflict-related gaps in weather station coverage tend to understate the positive association, if any, between temperature shocks and conflict.

If so, a key implication of this analysis is that the estimated effect of temperature will be higher if steps are taken to reduce the measurement error. One such option is to estimate the relationship using regression calibration (RC), a method for dealing with measurement error when multiple proxies exist (Carroll et al. 2006, chap. 4).[16] The basic idea is to estimate β by

[16] An alternative method would be multiple imputation; see Blackwell, Honaker, and King (2017).

replacing $Temp_{it}^{*}$ in equation (2) with its expectation conditional on the proxies, $Temp_{it}^{j}$, and the other covariates, in this case the country and year fixed effects. This implies a two-step procedure that first generates a linear approximation to $Temp_{it}^{*}$ from the proxy data and then estimates the effect on $\widehat{Temp}_{it}^{*}$ on conflict, with the standard errors in the second stage corrected for the additional parameters from this substitution. This solution does not explicitly deal with heteroscedastic measurement error, but Spiegelman, Logan, and Grove (2011) show that RC performs well even in the presence of moderate heteroscedasticity.

The data are organized into country-year observations, and the dependent variable records whether the country experienced a civil conflict that led to at least 25 battle deaths in that year. Although researchers have used a variety of different models and operationalizations of temperature shocks, a common practice is to estimate a linear probability model with country and year fixed effects (Burke et al. 2009; Hsiang, Burke, and Miguel 2013). While we replicate this basic specification, we note that prior work has generally neglected to take into account the very pronounced first-order autocorrelation in the dependent variable, reflecting a strong tendency for conflicts, once started, to continue. There is good reason to think that the onset of a new conflict episode (or the re-ignition of a conflict after some spell of peace) is driven by different factors than the continuation of a conflict that was already ongoing. One solution to this problem is to restrict the sample to observations in which there was no conflict in the previous year, thereby estimating the effect of a temperature shock on the transition probability from a year without conflict to a year with conflict (Jackman 2000).

Figure 7 summarizes the results of this exercise, showing the estimated coefficient on temperature anomaly using the four temperature data sets and the RC estimate. [17] The tests were run on both the full sample, covering the years 1960-2014, and on the sub-sample that conditions on the absence of conflict in the previous year.[18] Although all estimates are positive, the RC estimate is the largest in both samples and is statistically different from zero when the sample is restricted to cases that did not experience conflict the prior year. The RC estimate in panel (b) is about twice as large as those obtained using the proxy data and imply that a one degree increase in temperature is associated with a 0.073 increase in the probability of civil conflict. The RC estimates also have larger standard errors, in part because they capture the uncertainty in $\widehat{Temp}_{it}^{*}$, rather than treating temperature as noiseless. In the appendix, we compare the RC estimate to the others, both individually and collectively, and show that it is statistically distinguishable from the Willmott estimate and from the average of the estimates obtained across the four data sets (SI, pp. 7-8). Note also that the CRU estimate is not systematically lower than the others, a result that is consistent with the finding that the bias in those data is relatively small at the country-year level.

---

[17] Standard errors are clustered by country. We implemented the RC estimator using the `rcal` command (Hardin, Schmiediche, and Carroll 2003) with standard errors calculated by clustered bootstrap.

[18] South Sudan and Sudan post-2011 are dropped from the sample, since they enter so close to the end of the sample. Starting in 1960 reduces the imbalance in the panel due to countries that became independent earlier. Including earlier years does not change the results, but doing so complicates the estimation of year fixed effects when using the clustered bootstrap.

--- FIGURE 7 ABOUT HERE---

4. Conclusions

Several conclusions flow from this study. First, researchers interested in the link between climate and political, economic, or social outcomes need to think about the processes that generate the climate data and choose sources that are robust to variation in the outcomes being explored. The overall message is *not* that weather cannot be used as an instrument in studies of these outcomes; rather, it is that familiarity with the underlying data is necessary to ensure that weather observations are indeed exogenous to the processes being studied. Where significant measurement error is suspected, some effort should be made to address that error such as by using RC or multiple imputation. For researchers who want a reasonable measure of temperature that is not systematically affected by coverage problems, our analysis suggests that the BEST data have desirable properties. While the weather station data used by the BEST project are influenced by civil conflict and its risk factors, the network is sufficiently dense that its estimates appear unaffected by coverage gaps.

While this paper has focused on temperature, there has also been interest in the effect of precipitation on conflict risk (e.g., Miguel et al. 2004; Dube and Vargas 2013) and as an instrument for other forms of political activity, such as protests and voting (e.g., Ritter and Conrad 2016; Hansford and Gomez 2010). It is likely that precipitation data suffer from similar issues in regions like SSA. Although there are more stations that measure rainfall than temperature, precipitation anomalies are correlated across smaller distances (450km vs. 1200km), so gaps caused by station loss could in principle be more severe. Unlike with temperature, there is no consistent trend in precipitation levels at regional or local scales, so the

CRU practice of filling gaps with zero anomalies does not introduce a predictable bias; however, measurement error is likely quite large, and there is significant variation across observational data sets due in part to sparse coverage (Sarojini, Stott, and Black 2016).

A second implication of this study is that, if recent trends of station loss continue, the problems identified here are going to get worse. The ideal solution, of course, would be to reverse the trend, which would require investment in new stations, maintenance of existing ones, and harvesting of existing data from stations not currently in the system (United Nations Economic Commission for Africa 2011). Unfortunately, the requirement by CRU that stations have a record of coverage in the 1961-90 period makes meaningful addition of new stations problematic. Moreover, the practice of filling in coverage gaps with zero anomaly readings may make sense as a conservative way to avoid overstating recent global warming; however, from the perspective of the research at grid cell or country-level spatial scale, this practice introduces a downward bias on temperature estimates, a bias that grows more pronounced with warming temperatures.

Going forward, the solution to this problem may be to rely on temperature estimates from sources that are not influenced by political and economic conditions in the places that are the object of study. Two main avenues suggest themselves. First, satellite-based measurements are a natural alternative, since they are not affected by conditions on the ground in the countries being observed. Existing satellite-based data, however, have limitations for the applications such as those considered here. The longest running satellite based temperature dataset, produced by the University of Alabama in Huntsville, goes back to Dec. 1978 and is only available at a 2.5° resolution, which is crude for country-level analysis. Higher resolution observations are available from the Moderate Resolution Imaging Spectroradiometer (MODIS), but data acquisition did not begin until 2000. There are also challenges in deriving surface air temperature estimates from satellites, due to the effects of soil moisture, solar radiation, and cloud cover (Vancutsem et al. 2010). Nevertheless, Heft-Neal, Lobell, and Burke (2017) show that the 1km daily-scale MODIS observations improve as a proxy for surface air temperature at higher levels of temporal aggregation, suggesting that these data may be useful for estimating climate response functions at the monthly or annual levels.

Second, researchers may be able to take advantage of the fact that sea surface temperatures (SST) are observed through a variety of mechanisms that are not directly affected by conditions in the countries of observation, such as shipboard measurements, floats, and satellites. Researchers can then take advantage of links between terrestrial air temperatures and climate patterns in the surrounding oceans or other large-scale variation in ocean temperatures, such the El Niño-South Oscillation (ENSO) in the tropical Pacific. For example, Hsiang, Meng, and Cane (2011) avoid terrestrial temperature observation altogether by showing an increase in civil war risk during El Niño years in countries whose climates are strongly affected by the ENSO cycle. In a recent paper, Linke et al. (2017) simulate the historical temperature in SSA by using the publicly available Community Earth Systems Model (CESM) constrained by global observations of the monthly SSTs. The simulations track the CRU temperature estimates reasonably well and generate similar results in their model of political violence.

Future advancement on these fronts would be helpful in insulating the measurement of temperature from the political, economic, and social conditions that we are trying to understand.

References

Allansson, Marie, Erik Melander, and Lotta Themnér. 2017. “Organized Violence, 1989–2016.” *Journal of Peace Research* 54 (4): 574–87. https://doi.org/10.1177/0022343317718773.

Besley, Timothy, and Tortsen Persson. 2010. “State Capacity, Conflict, and Development.” *Econometrica* 78 (1): 1–34. https://doi.org/10.3982/ECTA8073.

Blackwell, Matthew, James Honaker, and Gary King. 2017. “A Unified Approach to Measurement Error and Missing Data: Overview and Applications.” *Sociological Methods & Research* 46 (3): 303–41. https://doi.org/10.1177/0049124115585360.

Bollfrass, A., and A. Shaver. 2015. “The Effects of Temperature on Political Violence: Global Evidence at the Subnational Level.” *PLoS ONE* 10 (5).

Buhaug, Halvard, and Kristian Skrede Gleditsch. 2008. “Contagion or Confusion? Why Conflicts Cluster in Space.” *International Studies Quarterly* 52 (2): 215–33. https://doi.org/10.1111/j.1468-2478.2008.00499.x.

Burke, M. B., E. Miguel, S. Satyanath, J. A. Dykema, and D. B. Lobell. 2010. “Climate Robustly Linked to African Civil War.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (51): E185.

Burke, Marshall B., Edward Miguel, Shanker Satyanath, John A. Dykema, and David B. Lobell. 2009. “Warming Increases the Risk of Civil War in Africa.” *Proceedings of the National Academy of Sciences* 106 (49): 20670–74.

Carleton, T. A., and S. M. Hsiang. 2016. “Social and Economic Impacts of Climate.” *Science* 353 (6304): aad9837–aad9837. https://doi.org/10.1126/science.aad9837.

Carroll, Raymond J., David Ruppert, Leonard A. Stefanski, and Ciprian M. Crainiceanu. 2006. *Measurement Error in Nonlinear Models: A Modern Perspective*. 2nd ed. Monographs on Statistics and Applied Probability 105. Boca Raton, FL: Chapman & Hall/CRC.

Conley, Timothy G. 2008. "Spatial Econometrics." In *The New Palgrave Dictionary of Economics*, edited by Steven N. Durlauf and Lawrence E. Blume, 2nd ed., 741–47. Basingstoke: Nature Publishing Group. https://doi.org/10.1057/9780230226203.1582.

Couttenier, Mathieu, and Raphael Soubeyran. 2014. "Drought and Civil War In Sub-Saharan Africa." *The Economic Journal* 124 (575): 201–44. https://doi.org/10.1111/ecoj.12042.

Cowtan, Kevin, and Robert G. Way. 2014. "Coverage Bias in the HadCRUT4 Temperature Series and Its Impact on Recent Temperature Trends: Coverage Bias in the HadCRUT4 Temperature Series." *Quarterly Journal of the Royal Meteorological Society* 140 (683): 1935–44. https://doi.org/10.1002/qj.2297.

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken. 2012. "Temperature Shocks and Economic Growth: Evidence from the Last Half Century." *American Economic Journal: Macroeconomics*, 66–95.

———. 2014. "What Do We Learn from the Weather? The New Climate-Economy Literature." *Journal of Economic Literature* 52 (3): 740–98. https://doi.org/10.1257/jel.52.3.740.

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.

Fan, Yun, and Huug van den Dool. 2008. "A Global Monthly Land Surface Air Temperature Analysis for 1948–Present." *Journal of Geophysical Research* 113 (D1). https://doi.org/10.1029/2007JD008470.

Fearon, James D., and David D. Laitin. 2003. “Ethnicity, Insurgency, and Civil War.” *American Political Science Review* 97 (01): 75–90. https://doi.org/10.1017/S0003055403000534.

Gleditsch, Nils Peter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Håvard Strand. 2002. “Armed Conflict 1946-2001: A New Dataset.” *Journal of Peace Research* 39 (5): 615–37. https://doi.org/10.1177/0022343302039005007.

Hansford, Thomas G., and Brad T. Gomez. 2010. “Estimating the Electoral Effects of Voter Turnout.” *The American Political Science Review* 104 (2): 268–88.

Hardin, J. W., H. Schmiediche, and R. J. Carroll. 2003. “The Regression-Calibration Method for Fitting Generalized Linear Models with Additive Measurement Error.” *Stata Journal* 3 (4): 361–72.

Harris, I., P.D. Jones, T.J. Osborn, and D.H. Lister. 2014. “Updated High-Resolution Grids of Monthly Climatic Observations - the CRU TS3.10 Dataset.” *International Journal of Climatology* 34 (3): 623–42. https://doi.org/10.1002/joc.3711.

Heft-Neal, Sam, David B Lobell, and Marshall Burke. 2017. “Using Remotely Sensed Temperature to Estimate Climate Response Functions.” *Environmental Research Letters* 12 (1): 014013. https://doi.org/10.1088/1748-9326/aa5463.

Hendrix, Cullen S. 2010. “Measuring State Capacity: Theoretical and Empirical Implications for the Study of Civil Conflict.” *Journal of Peace Research* 47 (3): 273–85. https://doi.org/10.1177/0022343310361838.

Hendrix, Cullen S. 2017. “The Streetlight Effect in Climate Change Research on Africa.” *Global Environmental Change* 43 (March): 137–47. https://doi.org/10.1016/j.gloenvcha.2017.01.009.

Hollyer, James R., B. Peter Rosendorff, and James Raymond Vreeland. 2014. “Measuring Transparency.” *Political Analysis* 22 (4): 413–34.

Hsiang, S. M. 2010. “Temperatures and Cyclones Strongly Associated with Economic Production in the Caribbean and Central America.” *Proceedings of the National Academy of Sciences* 107 (35): 15367–72. https://doi.org/10.1073/pnas.1009510107.

Hsiang, S. M., K. C. Meng, and M. A. Cane. 2011. “Civil Conflicts Are Associated with the Global Climate.” *Nature* 476 (7361): 438–41.

Hsiang, Solomon M., Marshall Burke, and Edward Miguel. 2013. “Quantifying the Influence of Climate on Human Conflict.” *Science* 341 (6151). https://doi.org/10.1126/science.1235367.

Jackman, Simon. 2000. “In and Out of War and Peace: Transitional Models of International Conflict.” Stanford University. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.200.5895&rep=rep1&type=pdf.

Jerven, Morten. 2013. *Poor Numbers: How We Are Misled by African Development Statistics and What to Do about It*. Ithaca: Cornell University Press.

Kelley, Judith G., and Beth A. Simmons. 2015. “Politics by Number: Indicators as Social Pressure in International Relations: POLITICS BY NUMBER.” *American Journal of Political Science* 59 (1): 55–70. https://doi.org/10.1111/ajps.12119.

Kraay, Aart, Daniel Kaufmann, and Massimo Mastruzzi. 2010. *The Worldwide Governance Indicators : Methodology and Analytical Issues*. Policy Research Working Papers. The World Bank. https://doi.org/10.1596/1813-9450-5430.

Lawrimore, Jay H., Matthew J. Menne, Byron E. Gleason, Claude N. Williams, David B. Wuertz, Russell S. Vose, and Jared Rennie. 2011. “An Overview of the Global Historical Climatology Network Monthly Mean Temperature Data Set, Version 3.” *Journal of Geophysical Research* 116 (D19). https://doi.org/10.1029/2011JD016187.

Lee, Melissa M., and Nan Zhang. 2016. “Legibility and the Informational Foundations of State Capacity.” *The Journal of Politics* 79 (1): 118–32. https://doi.org/10.1086/688053.

Linke, Andrew M, John O’Loughlin, Andrew Gettelman, and Arlene Laing. 2017. “Subnational Violent Conflict Forecasts for Sub-Saharan Africa, 2015–65, Using Climate-Sensitive Models.” *Journal of Peace Research* 54 (2): 175–92. https://doi.org/10.1177/0022343316682064.

Maystadt, J.-F., M. Calderone, and L. You. 2015. “Local Warming and Violent Conflict in North and South Sudan.” *Journal of Economic Geography* 15 (3): 649–71. https://doi.org/10.1093/jeg/lbu033.

Merry, Sally Engle. 2016. *The Seductions of Quantification: Measuring Human Rights, Gender Violence, and Sex Trafficking*. Chicago Series in Law and Society. Chicago: The University of Chicago Press.

Miguel, Edward, Shanker Satyanath, and Ernest Sergenti. 2004. “Economic Shocks and Civil Conflict: An Instrumental Variables Approach.” *Journal of Political Economy* 112 (4): 725–53.

O’Loughlin, J., A. M. Linke, and F. D. W. Witmer. 2014. “Effects of Temperature and Precipitation Variability on the Risk of Violence in Sub-Saharan Africa, 1980-2012.” *Proceedings of the National Academy of Sciences of the United States of America* 111 (47): 16712–17.

O'Loughlin, J., F. D. Witmer, A. M. Linke, A. Laing, A. Gettelman, and J. Dudhia. 2012. "Climate Variability and Conflict Risk in East Africa, 1990-2009." *Proceedings of the National Academy of Sciences of the United States of America* 109 (45): 18344–49.

Prichard, Wilson, Alex Cobham, and Goodhall. 2014. *The ICTD Government Revenue Dataset*. ICTD Working Paper 19. Brighton: International Centre for Tax and Development.

Ritter, Emily Hencken, and Courtenay R. Conrad. 2016. "Preventing and Responding to Dissent: The Observational Challenges of Explaining Strategic Repression." *American Political Science Review* 110 (01): 85–99. https://doi.org/10.1017/S0003055415000623.

Rohde, Robert, Richard Muller, Robert Jacobsen, Saul Perlmutter, and Steven Mosher. 2013. "Berkeley Earth Temperature Averaging Process." *Geoinformatics & Geostatistics: An Overview* 01 (02). https://doi.org/10.4172/2327-4581.1000103.

Sarojini, Beena Balan, Peter A. Stott, and Emily Black. 2016. "Detection and Attribution of Human Influence on Regional Precipitation." *Nature Climate Change* 6 (June): 669–75.

Spiegelman, Donna, Roger Logan, and Douglas Grove. 2011. "Regression Calibration with Heteroscedastic Error Variance." *The International Journal of Biostatistics* 7 (1): 1–34. https://doi.org/10.2202/1557-4679.1259.

Sundberg, Ralph, and Erik Melander. 2013. "Introducing the UCDP Georeferenced Event Dataset." *Journal of Peace Research* 50 (4): 523–32. https://doi.org/10.1177/0022343313484347.

United Nations Economic Commission for Africa. 2011. "An Assessment of Africa's Climate Observing Networks and Data Including Strategies for Rescuing of Climatic Data." Addis Ababa. http://hdl.handle.net/10855/21110.

Vancutsem, Christelle, Pietro Ceccato, Tufa Dinku, and Stephen J. Connor. 2010. "Evaluation of MODIS Land Surface Temperature Data to Estimate Air Temperature in Different Ecosystems over Africa." *Remote Sensing of Environment* 114 (2): 449–65. https://doi.org/10.1016/j.rse.2009.10.002.

Willmott, Cort J., and Kenji Matsuura. 2015. "Terrestrial Air Temperature: 1900-2014 Gridded Monthly Time Series, Version 4.01 (1900 - 2014)." 2015. http://climate.geog.udel.edu/~climate/html_pages/Global2014/README.GlobalTsT2014.html.

World Meteorological Organization. 2008. *Guide to Meteorological Instruments and Methods of Observation.* Geneva, Switzerland: World Meteorological Organization.

———. 2011. *Guide to Climatological Practices*. Geneva, Switzerland: World Meteorological Organization.

Table 1. Correlations between Coverage, Conflict, and State Capacity

	CRU Stations	CRU in 1200km	BEST Stations	BEST in 1200km	Civil Conflict
Civil Conflict	-0.29*	-3.15**	-0.66**	-5.92**	0.40**
N = 2270	(0.13)	(0.96)	(0.21)	(1.33)	(0.00)
Avg. Conflict in 1961-90	-0.44*	-3.75**	-0.71	-3.92	0.17**
N = 2270	(0.20)	(1.28)	(0.36)	(2.85)	(0.03)
Conflict w/in 1200km	-0.31*	-3.27**	-0.61*	-5.89**	0.05**
N = 2270	(0.13)	(1.00)	(0.25)	(1.44)	(0.02)
GDP per capita (logged)	-0.02	0.11	0.34	5.34	-0.05
N = 1986	(0.22)	(1.50)	(0.38)	(3.65)	(0.03)
Infant mortality	0.22	5.64**	-0.35	0.31	-0.00
N = 2084	(0.15)	(0.90)	(0.27)	(2.49)	(0.03)
Revenue/GDP	-0.09	-1.16	0.34	7.29	-0.09**
N = 1266	(0.17)	(1.24)	(0.35)	(4.16)	(0.03)
Bureaucratic quality	0.18	0.66	0.51	4.14	-0.09*
N = 972	(0.21)	(1.45)	(0.41)	(2.98)	(0.03)
Gov. effectiveness	0.23	0.51	0.94*	13.41**	-0.11**
N = 762	(0.17)	(1.03)	(0.39)	(3.28)	(0.03)
Transparency	0.32	0.77	0.77*	5.05*	-0.06*
N = 1178	(0.18)	(1.62)	(0.31)	(2.35)	(0.03)
Population density	-0.03	-3.01**	0.57	-2.71	0.03
N = 2214	(0.22)	(0.85)	(0.44)	(1.79)	(0.02)
Mean Elevation	-0.82**	-4.99*	-0.99	0.41	0.06
N = 2270	(0.28)	(2.29)	(0.49)	(4.04)	(0.04)
Area (logged)	-0.59	-2.20	-1.63*	-3.55	0.10**
N = 2270	(0.39)	(1.51)	(0.64)	(3.15)	(0.03)

Note: This table reports coefficients from bivariate regressions of each independent variable on the coverage and conflict indictors. All independent variables were standardized. Standard errors corrected for clustering by country. ** $p<0.01$, * $p<0.05$

Table 2. The Effect of Civil Conflict on the Risk of Station Death

	(1)	(2)	(3)
Civil Conflict	1.71*		2.17**
	(0.37)		(0.394)
Violent event 0-10km		3.03*	1.78
		(1.36)	(0.76)
Violent event 10-50km		1.96*	1.18
		(0.64)	(0.41)
Violent event 50-100km		1.54	0.97
		(0.37)	(0.215)
Violent event 100km+		1.27	0.85
		(0.27)	(0.15)
Station Age (logged)	0.89*	0.93	0.93
	(0.048)	(0.072)	(0.073)
Observations	307,922	205,883	205,883
Country FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes

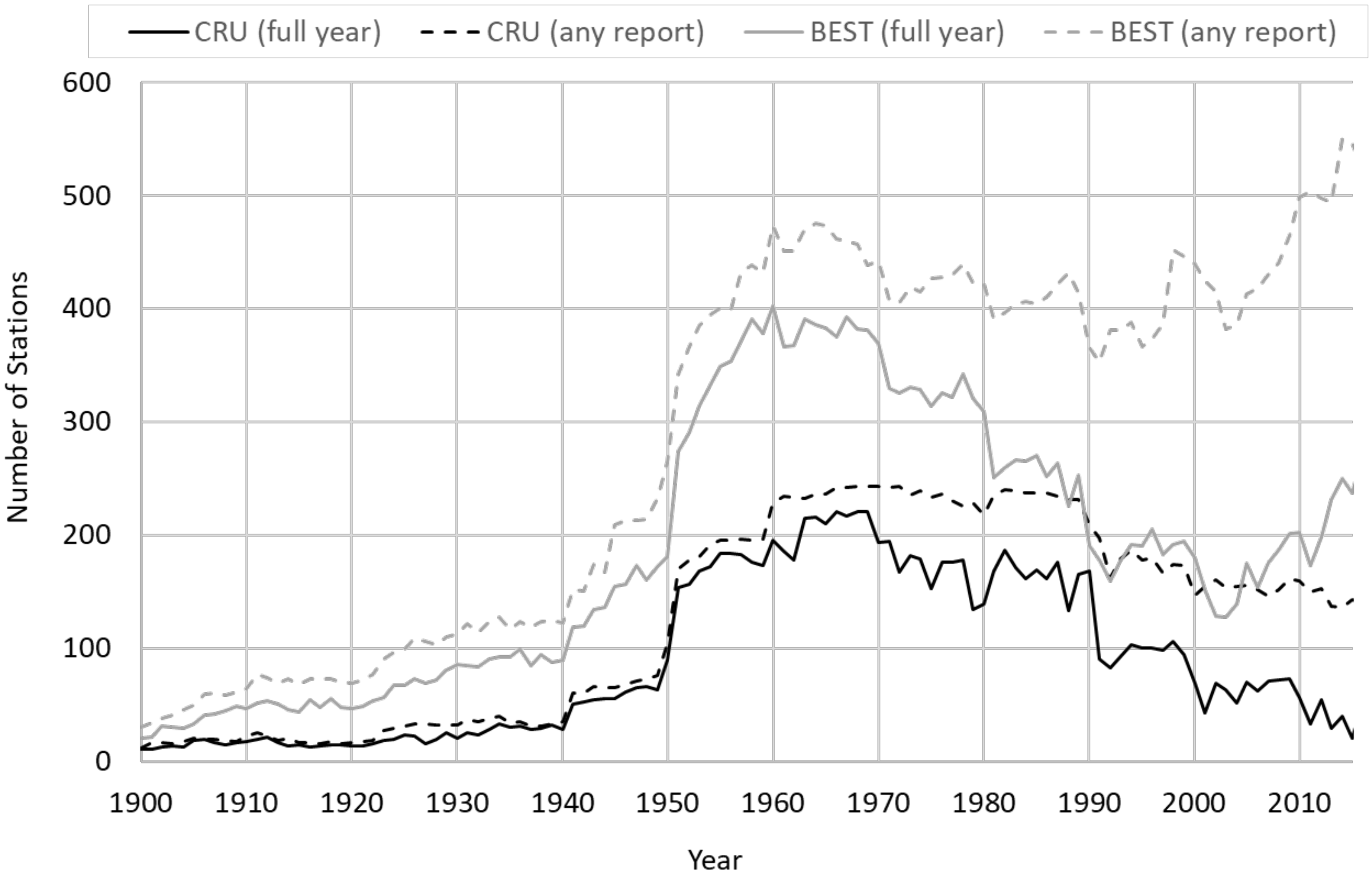
Note: Entries in the table are odds ratios, with robust standard errors, clustered by country, in parentheses. Indicators for violent events are lagged by one month. All models include a constant and a counter for the duration of the current spell of activity, included as a cubic polynomial. The sample in column (1) covers 1973-2016; columns (2)-(3) cover 1989-2016. ** $p<0.01$, * $p<0.05$

Table 3. The Effect of Civil Conflict on Reliability

	(1)
Civil Conflict	-0.114
	(0.112)
Conflict w/in 1200km	-0.037
	(0.115)
Avg. Conflict Incidence	-0.625*
	(0.297)
Avg. Conflict w/in 1200km	-0.554
	(0.412)
Temperature Anomaly	-0.089
	(0.162)
Mean Elevation	-0.001**
	(0.000)
Area (logged)	0.139**
	(0.034)
Latitude	0.010
	(0.005)
Latitude^2	0.002**
	(0.000)
Year FE	Yes
Observations	2,174
R-squared	0.44

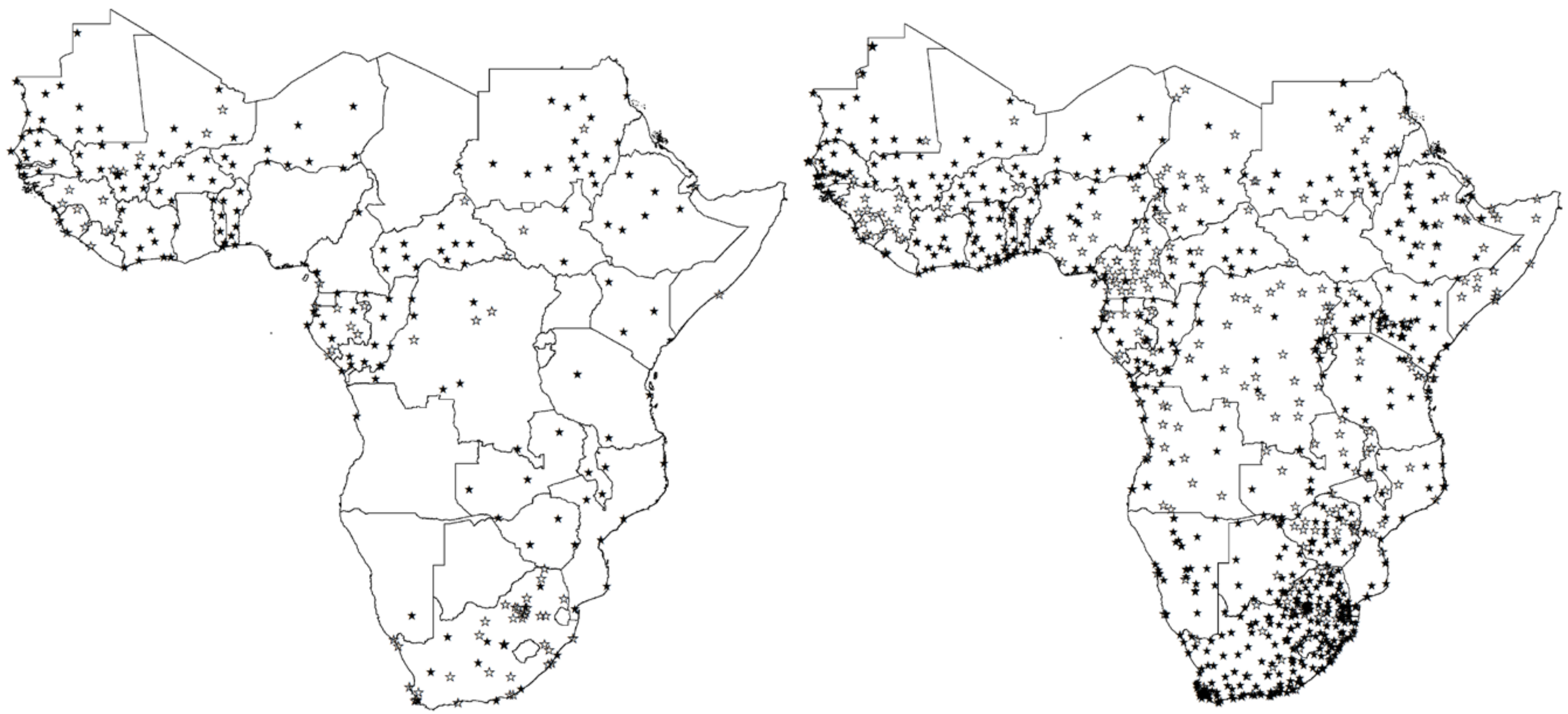
Note: The table reports estimates from a linear regression on the signal-to-noise ratio (logged). Standard errors in parentheses corrected cross-sectional spatial dependence within 1200km and panel-specific serial correlation over 5 years. ** p<0.01, * p<0.05

Figure 1. Weather Stations in Sub-Saharan Africa, 1900-2016



Note: This figures shows the number of weather stations in Sub-Saharan Africa that reported a temperature in all 12 months of the year (solid lines) or at least one month in the year (dashed lines). Weather stations counts come from CRU (black) and BEST (grey) data sets.

Figure 2. Locations of Active and Defunct Weather Stations

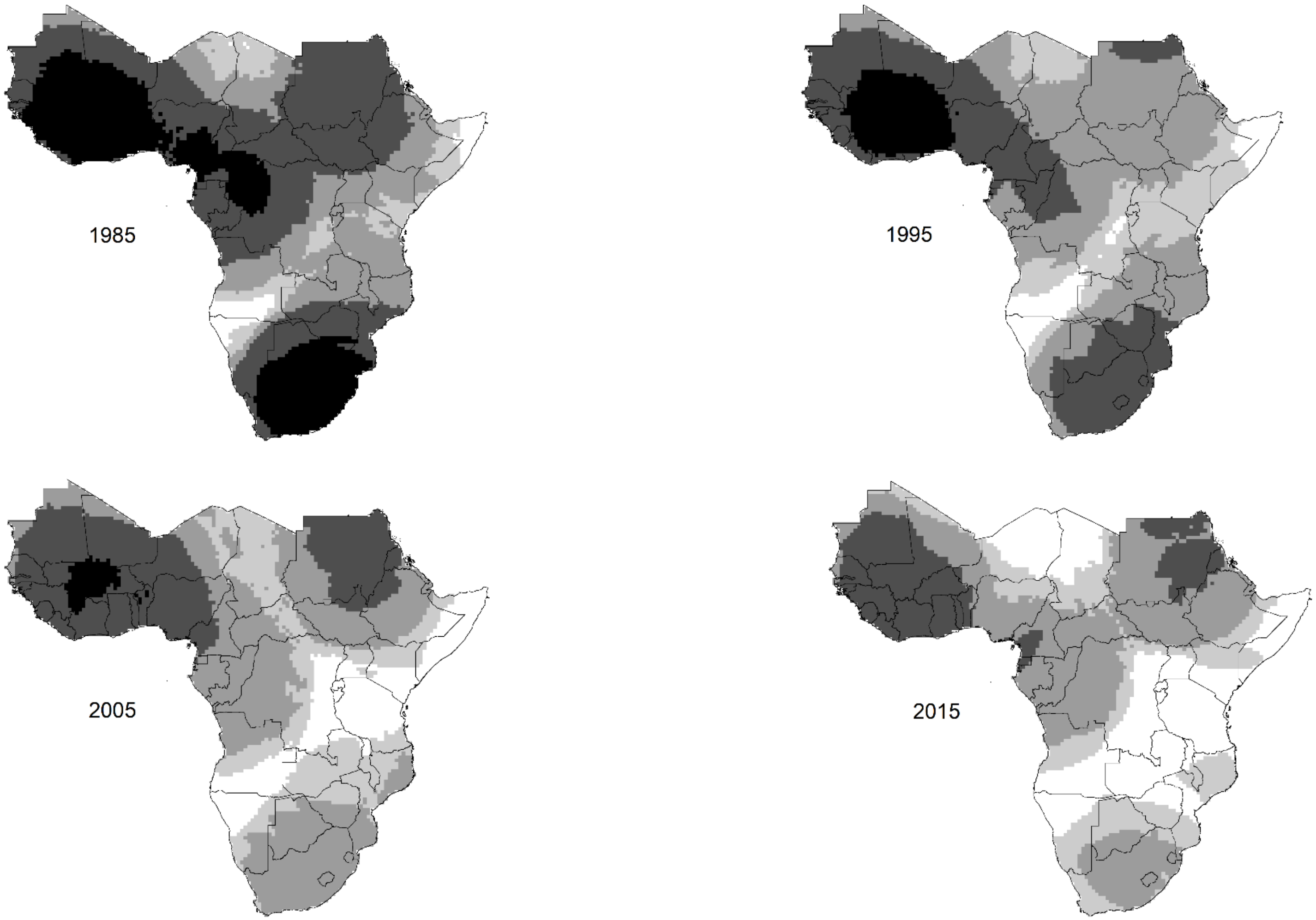


(a) CRU

(b) BEST

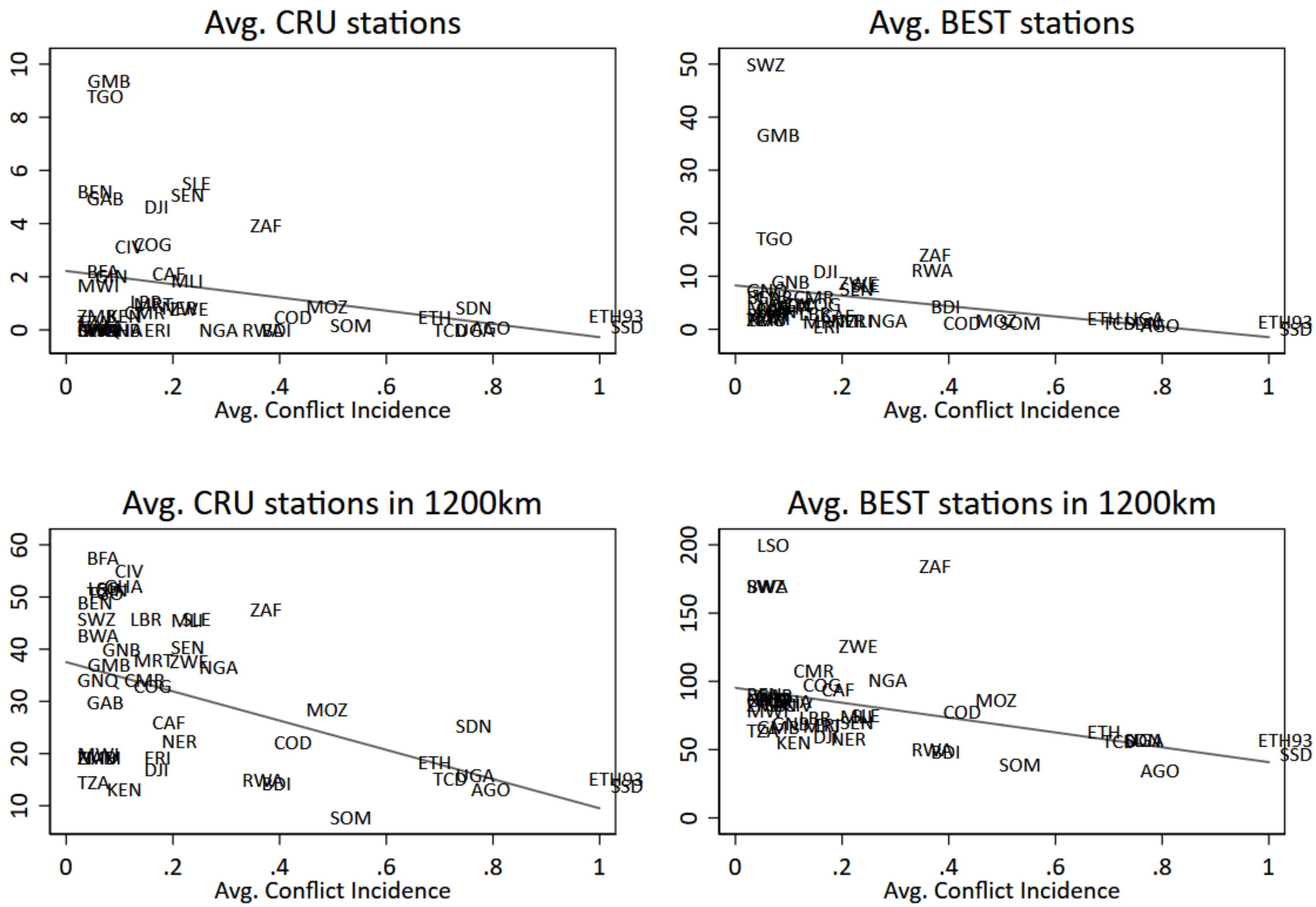
Note: The maps show the location of weather stations that contributed to CRU (panel a) and BEST (panel b) high resolution times series temperature data in the period 1946-2016. Solid stars indicate stations that reported at least once in the period 2010-16; hollow stars show stations that did not report in this period.

Figure 3. The Evolution of Coverage in the CRU Data



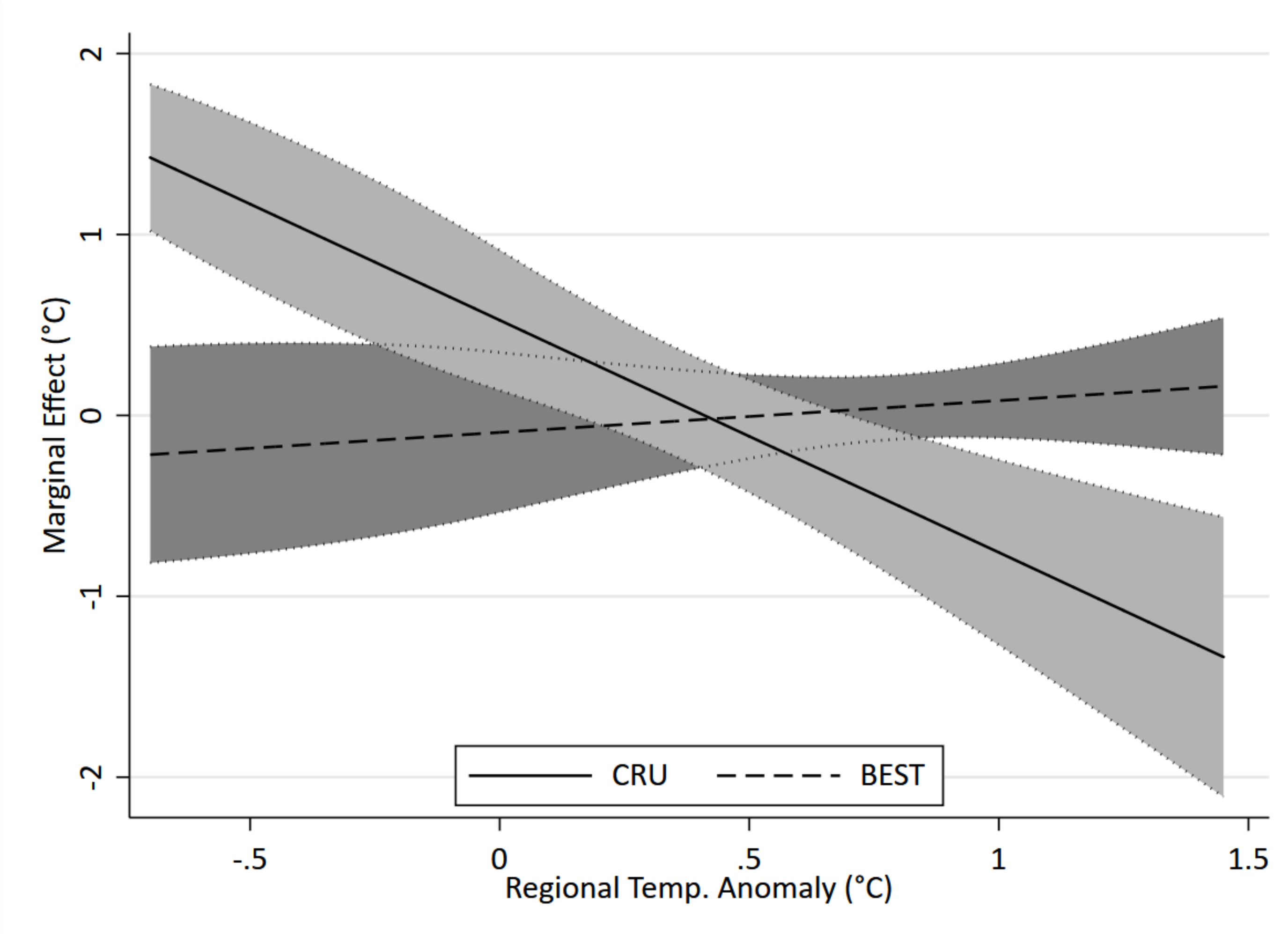
Note: The maps show for each 0.5° grid cell the number of stations with 1200km of the cell that reported a temperature in January of the indicated year.

Figure 4. Average Coverage and Civil Conflict Incidence, 1946-2016



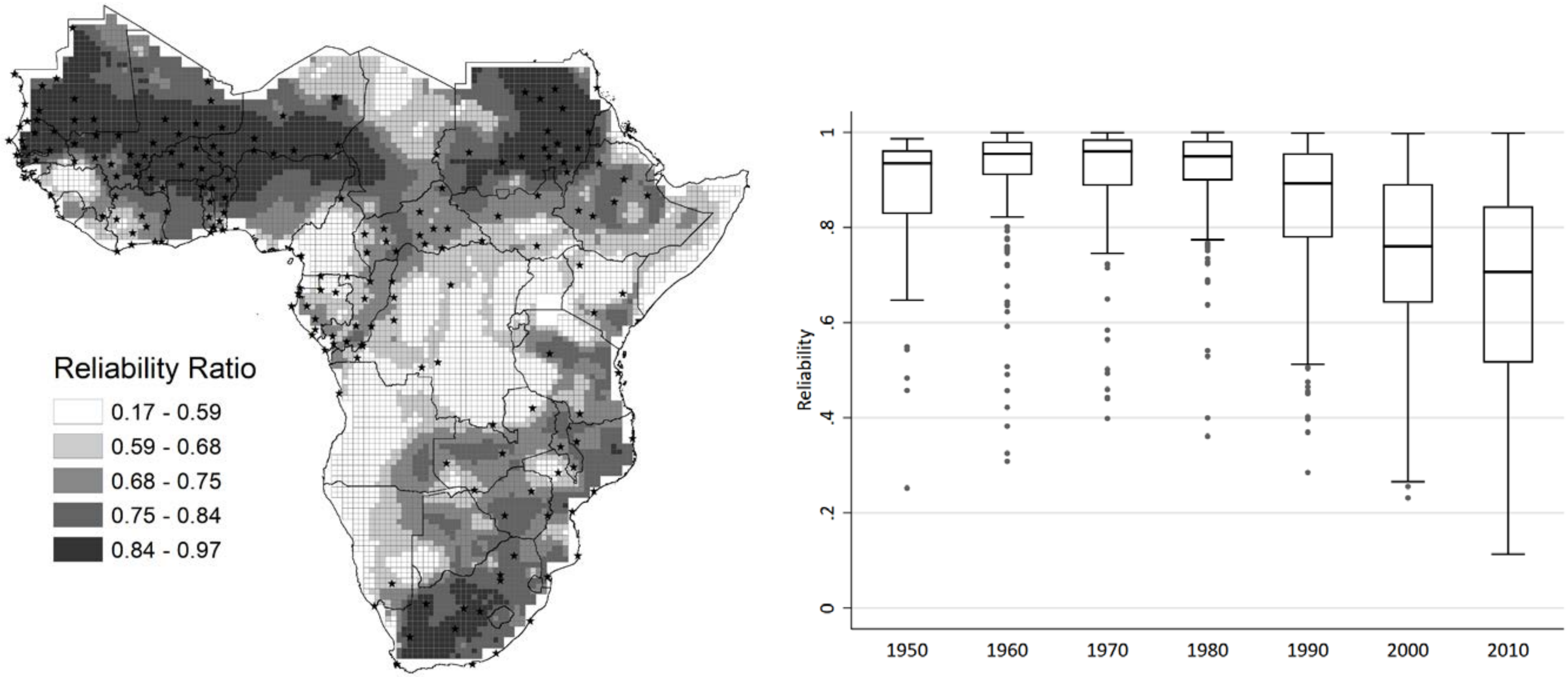
Note: The figures show the cross-sectional relationship between each coverage measure and the proportion of years that a country experienced civil conflict as an independent state in the period 1946-2016. Station counts in the top row are per 100,000 sq. km. of country area. Station counts in the bottom row report the average number of stations with 1200km of each grid cell in the country.

Figure 5. Estimated Effect of Coverage Gaps on Reported Temperatures



Note: The figure shows the marginal effect of CRU coverage gaps on a country's annual temperature anomaly reported by CRU and BEST, respectively, as a function of the regional temperature anomaly in that year, with 95 percent confidence intervals. The estimates come from a regression that controls for the regional temperature anomaly and country fixed effects (N=2270).

Figure 6. The Distribution of Measurement Error in Temperature Observations

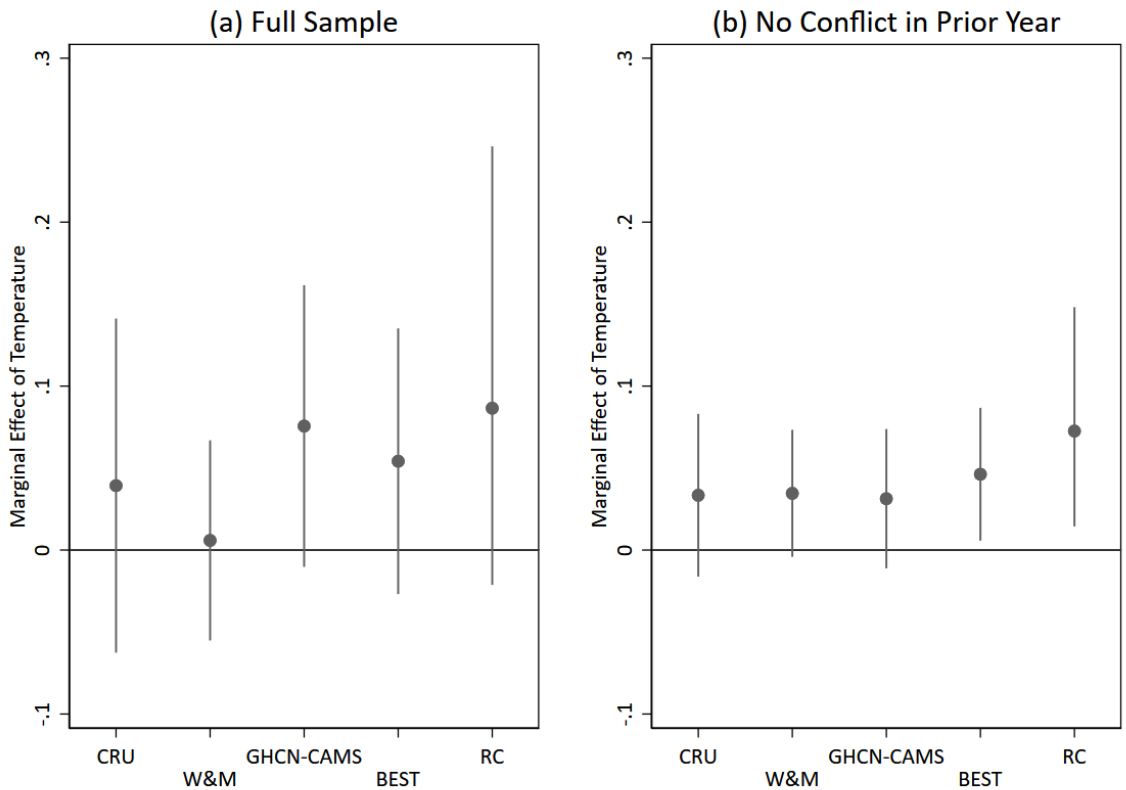


(a) Cell-month observations, 2000-14

(b) Country-year observations

Note: The map shows the reliability ratio for each grid cell averaged across all months 2000-14, divided into quintiles. Stars indicate the locations of CRU weather stations that reported at least once in that period. The box plot shows the distribution of reliability ratios at the country-year level for each decade.

Figure 7. The Estimated Effect of Temperature Anomaly on Civil Conflict Risk, 1960-2014



Note: This figure shows the estimated coefficient of temperature anomaly (°C) on the probability of civil conflict, with 95 percent confidence intervals, using each of the four temperature data sets and regression calibration (RC). Estimates are from linear probability models with country and year fixed effects and standard errors clustered by country. Panel (a) uses the full sample of country years (N=2125); panel (b) conditions on no conflict in the prior year (N=1709). CRU = Climatic Research Unit, W&M = Willmott and Matsuura (2015), GHCN-CAMS = Global Historical Climatology Network-Climate Anomaly Monitoring System, BEST = Berkeley Earth.

Is Temperature Exogenous? The Impact of Civil Conflict on the Instrumental Climate Record in Sub-Saharan Africa

Kenneth A. Schultz

Justin S. Mankin

Supplementary Information

(Intended for on-line publication only)

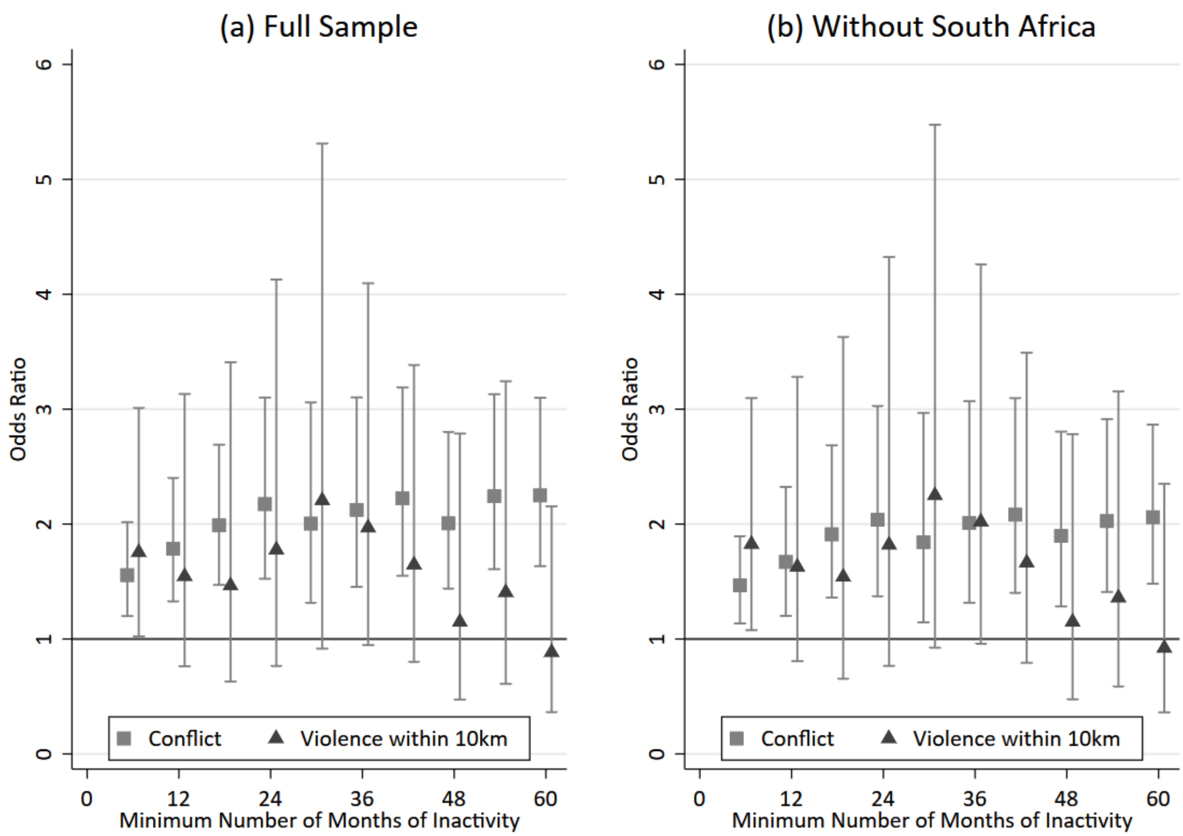
This supplementary appendix consists of four parts. Part 1 presents additional tests of the effect of civil conflict on the probability of station failure. Part 2 presents the analysis of bias in the CRU data using the cell-month level observations. Part 3 estimates the relationship between weather station coverage and the reliability of temperature observations. Part 4 analyzes whether the regression calibration (RC) estimate of the effect of temperature on conflict risk is significantly different from estimates obtained using the observational data sets.

1. Effect of Conflict on Station Loss

Figure A1 reports the results of two robustness checks on the results presented in Table 2 of the text. In both panels, we vary the minimum length of time a station has to be inactive before considering it dead. The figure reports the coefficients, expressed as odds ratios, and associated 95 percent confidence intervals, for two variables: whether there was a civil conflict in the country in which the station was located and whether there was a violent event between the government and rebels within 10km of the station in the prior month. Panel (a) reports results on the full sample, and the estimates reported for 24 months of inactivity correspond to those in column (3) in Table 2. Panel (b) reports results from a sample that excludes stations from South Africa. Inspection of the data revealed that South Africa accounts for 295 of the 1169 stations in the data set, or 25 percent; by comparison, the next most common country, the DRC, has only 60, or 5.1 percent. Given that South Africa is also the most developed country in SSA, it is important to see determine if the results are over influenced by those observations.

Civil conflict has a positive effect (i.e., greater than one) on the risk of station death for all criteria and in both samples, and its effect is largest for deaths that last multiple years. Violent events within 10km only have statistically significant effects when the minimum period of inactivity to consider a station dead is only six months. This suggests that violent events tend to increase the risk of short outages. None of these results change when South African stations are dropped from the sample.

Figure A1. The Estimated Effect of Conflict on the Risk of Station Death



Note: This figure shows the estimated effect of civil conflict and nearby violent events on the risk of station death, along with 95 percent confidence intervals, while varying the number of months of inactivity required to consider a station dead. Panel (a) shows estimates from the full sample of country years, while panel (b) shows estimates from a sample that excludes stations located in South Africa.

2. The Effect of Coverage Gaps on Cell-Month Temperature Estimates

In this section, we assess the bias induced in the CRU temperature data from the use of dummy stations with zero anomaly to interpolate observations when there are fewer than three stations within 1200km. Figure A2 reports the estimated bias in the CRU and BEST temperature anomalies as function of the number of reporting weather stations in the respective data set that was located within 1200km of the cell. The estimates were obtained by regressing the temperature anomaly on a series of dummy variables for each of the indicated station counts. The baseline case consists of cells that contained a weather station in the given month and therefore had their temperatures observed directly.[19] The regression model includes controls for the average temperature anomaly in all cells within 1200km, thereby capturing regional conditions, as well as the cell's mean elevation and its latitude (linear and squared).[20] Standard errors are corrected for cross-sectional spatial dependence within 1200km and panel-specific serial correlation over 5 years using the method proposed by Conley (2008) and implemented by Hsiang (2010).

As the figure shows, there is a negative bias in the CRU estimates for cells with sparse coverage, particularly those that had fewer than three stations within range and therefore were influenced by one or more dummy stations. For context, temperature anomalies in these data

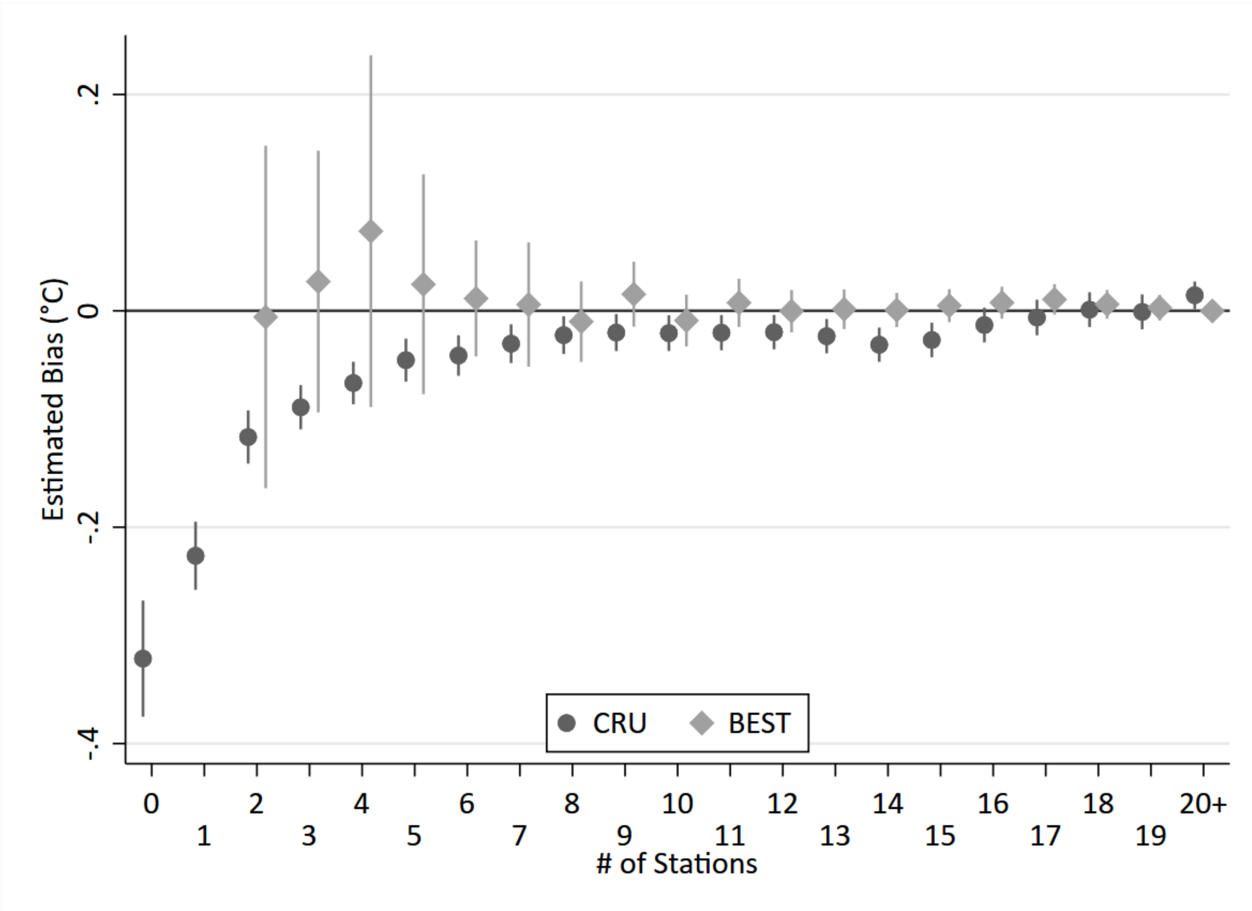
---

[19] Note that the BEST kriging algorithm imputes all temperatures, even when there is a weather station within the cell, but we treat such cells as the baseline case to ensure comparability with the CRU results.

[20] When calculating the average CRU temperature anomaly within 1200km, cells whose temperature were influenced by at least one dummy station were dropped.

range from -7.3 to 6.2, with a standard deviation of 0.91, so the bias in the worst case (no stations) is equivalent to about one-third of a standard deviation. By contrast, the BEST data set has no cells with fewer than two weather stations within 1200km, and there is no systematic effect of station density on temperature estimates.

Figure A2. The Effect of Coverage on Interpolated Temperature Anomalies



Note: This figure shows the estimated bias in the CRU and BEST gridded temperature estimates, respectively, as a function of the number of stations within 1200km of the cell. The baseline case is a cell in which a station was located.

3. The Effect of Coverage on Measurement Error

Table A1 reports estimates from linear regressions of the reliability on CRU and BEST coverage indicators using the grid cell-month (columns 1 and 2) and country-year (columns 3 and 4) as observations. As in Table 3 in the text, the dependent variable in these regressions is $\ln v_i^2 / \sigma_{it}^2$, or the logged signal to noise variance, which maps one-to-one onto reliability ratios. The coverage indicators are the (logged) count the number of reporting stations within 1200km of each grid cell in each month; the country-year data set uses area- and time-weighted averages. All models include a control for the average temperature across the four data sets as well measures of the unit's mean elevation, latitude (linear and squared), year fixed effects, and, in the case of the country-level data, area (logged). Standard errors are corrected for cross-sectional spatial dependence within 1200km and panel-specific serial correlation over five years using the method proposed by Conley (2008) and implemented by (Hsiang 2010). As the table shows, the reliability of the temperature estimates is robustly increasing in the density of weather station coverage.

Table A1. Estimated Effect of Coverage on Reliability, 1948-2014

	(1)	(2)	(3)	(4)
	Cell-month observations		Country-year observations	
	CRU	BEST	CRU	BEST
Coverage	0.460**	0.405**	0.515**	0.465**
	(0.010)	(0.012)	(0.112)	(0.157)
Mean Temp. Anomaly	-0.015	-0.017	-0.154	-0.158
	(0.009)	(0.009)	(0.157)	(0.163)
Mean Elevation	-0.000**	-0.000**	-0.001**	-0.001**
	(0.000)	(0.000)	(0.000)	(0.000)
Area (logged)			0.122**	0.103**
			(0.029)	(0.029)
Latitude	0.017**	0.018**	0.003	0.008
	(0.001)	(0.001)	(0.005)	(0.006)
Latitude^2	0.002**	0.002**	0.002**	0.002**
	(0.000)	(0.000)	(0.000)	(0.000)
Year FE	Yes	Yes	Yes	Yes
Observations	5,843,069	5,853,120	2,174	2,174
R^2	0.28	0.27	0.45	0.44

Note: The table shows the estimated effect of the CRU and BEST coverage indicators on the spread of temperature estimates at the cell-month and country-year levels of observation. Coverage indicator are the logged count of stations within 1200km of each cell, averaged across cells and months for the country-year data (columns 3 and 4). Standard errors are corrected for cross-sectional spatial dependence within 1200km and panel-specific serial correlation over five years. ** p<0.01, * p<0.05

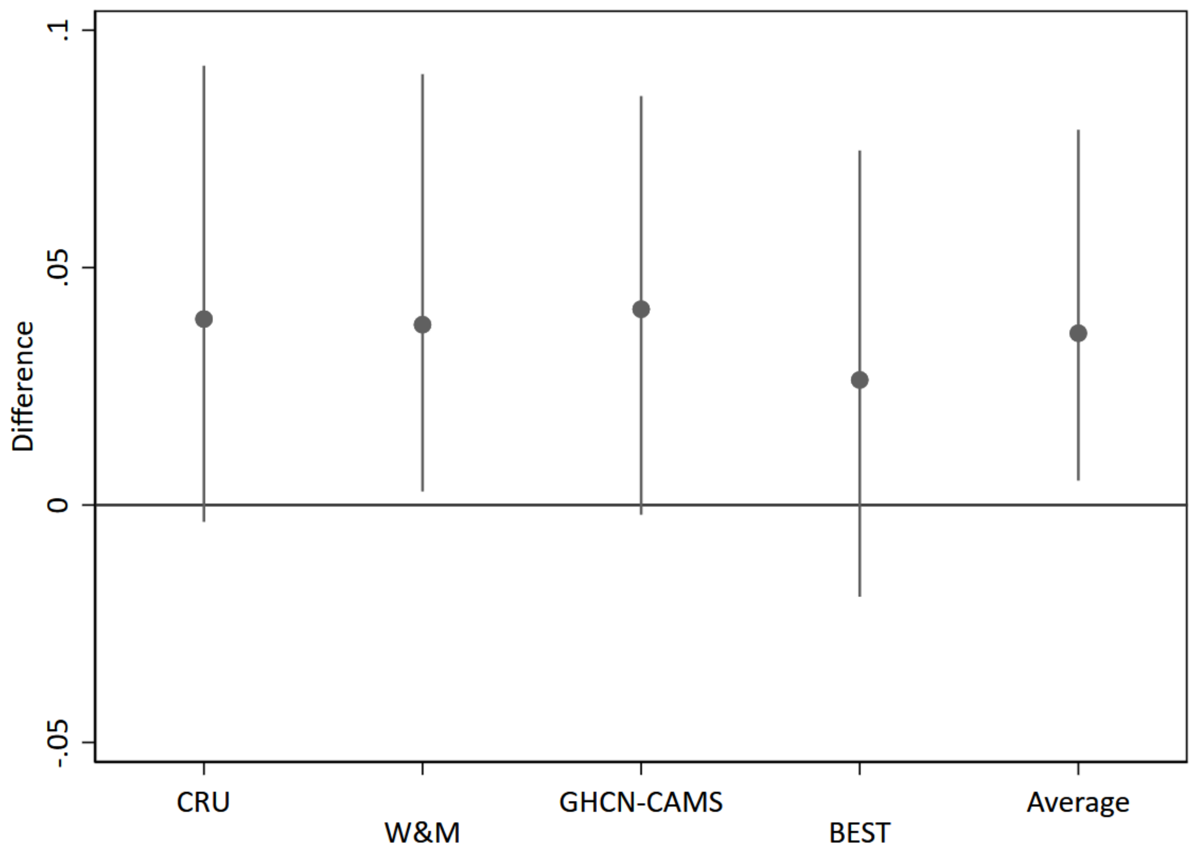
4. The Significance of the Regression Calibration Estimate

In this section, we examine the results from Figure 7b to assess whether the RC estimated of the effect of temperature on conflict risk is statistically different from the estimates obtained using the proxy data sets. There should be some caution in doing so. Even though the RC estimate is larger—as expected given the measurement error we have documented—it also has a larger standard error, due in part to the fact it captures the uncertainty inherent in estimating the "true" temperature from the proxy data sets. Hence, the noisiness of this estimate is to some extent a desirable property. Even so, it is useful to ask whether RC estimate might be larger than the others simply due to chance. To do this, we calculated two sets of quantities: (1) the difference between the RC estimate and each of the four estimates obtained from an observational data set and (2) the difference between the RC estimate and the average of the four other estimates. Standard errors for these quantities were calculated by bootstrapping, with clustering by country.[21]

The results are displayed in Figure A3. When compared to the individual estimates, the RC estimate is significantly different (at the 5 percent level) from the estimate obtained using the Willmott and Matsuura (2015) data. In the case of the CRU and GHCN-CAMS data, the p-value on the test statistic is 0.08. When compared to the average of the individual estimates, the RC estimate is significantly different at the 5 percent level. Thus, while the results from the individual comparisons are mixed, we can reject the null hypothesis that the RC estimate is larger than the average of all four other estimates simply due to chance.

[21] Comparisons of coefficients across models are often done by means of a seemingly unrelated regression; however, that option is not available with the rcal command in Stata.

Figure A3. Comparison of the Regression Calibration and Proxy-Based Estimates



Note: This figure shows the difference between the RC estimate of the effect of temperature on conflict risk from Figure 7b with the estimates obtained using the indicated temperature data sets and the average of those estimates. The 95 percent confidence intervals were calculated by bootstrapping, with clustering by country and bias correction. N=1709. CRU = Climatic Research Unit, W&M = Willmott and Matsuura (2015), GHCN-CAMS = Global Historical Climatology Network-Climate Anomaly Monitoring System, BEST = Berkeley Earth.