Environmental Research Letters

LETTER

Forest loss in Brazil increases maximum temperatures within 50 km

Avery S Cohn[1] , **Nishan Bhattarai**[2] , **Jake Campolo**[3], **Octavia Crompton**[4], **David Dralle**[5] , **John Duncan**[6] **and Sally Thompson**[4,7]

1   Friedman School of Nutrition Science and Policy, Tufts University, 150 Harrison Avenue, Boston, MA, 02155, United States of America
2   School of Environment and Sustainability, University of Michigan, 440 Church St, Ann Arbor, MI, 48109, United States of America
3   Department of Earth System Science, Stanford University, 473 Via Ortega, Stanford, CA 94305, United States of America
4   Department of Civil and Environmental Engineering, Davis Hall, UC Berkeley, Berkeley, CA, 94720, United States of America
5   Department of Geology, California State University, Sacramento, 6000 J Street, Sacramento, CA 95819, United States of America
6   School of Environment and Agriculture, University of Western Australia, 35 Stirling Hwy, Crawley, WA, 6009, Australia
7   Department of Civil, Environmental and Mining Engineering, University of Western Australia, 35 Stirling Hwy, Crawley, WA, 6009, Australia
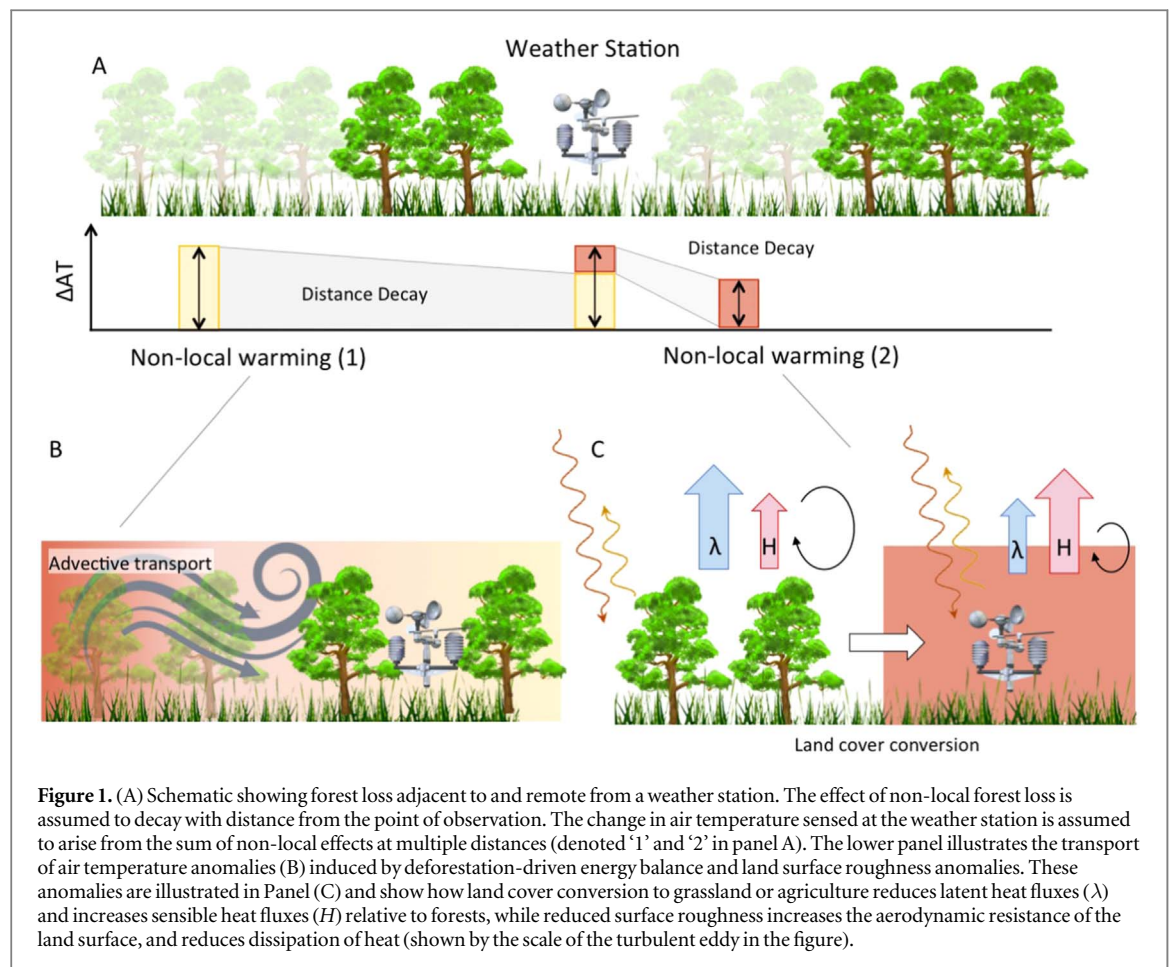
E-mail: avery.cohn@tufts.edu

Abstract

Forest cover loss in the tropics is well known to cause warming at deforested sites, with maximum temperatures being particularly sensitive. Forest loss causes warming by altering local energy balance and surface roughness, local changes that can propagate across a wide range of spatial scales. Consequently, temperature increases result from not only changes in forest cover at a site, but also by the aggregate effects of non-local forest loss. We explored such non-local warming within Brazil's Amazon and Cerrado biomes, the region with the world's single largest amount of forest loss since 2000. Two datasets, one consisting of *in-situ* air temperature observations and a second, larger dataset consisting of ATs derived from remotely-sensed observations of land surface temperature, were used to quantify changes in maximum temperature due to forest cover loss at varying length-scales. We considered undisturbed forest locations (1 km^2 in extent), and forest loss trends in annuli ('halos'), located 1–2 km, 2–4 km, 4–10 km and 10–50 km from these undisturbed sites. Our research finds significant and substantial non-local warming, suggesting that historical estimates of warming due to forest cover loss under-estimate warming or mis-attribute warming to local change, where non-local changes also influence the pattern of temperature warming.

1. Introduction

The expansion of agriculture is causing the loss of vast swaths of forests and woody savannas in the tropics and sub-tropics (Strassburg *et al* 2017, McNicol *et al* 2018, Curtis *et al* 2018). Tropical land cover change leads to considerable disruption of the Earth's climate system through both biogeochemical and biogeophysical pathways (Bonan 2008); where biogeochemical pathways include the direct emissions of greenhouse gases by land cover conversion and changes to carbon cycling, and biogeophysical pathways primarily include changes to the land surface energy balance. These energy balance changes themselves arise from the interactions of numerous phenomena, including

increases in albedo (Culf *et al* 1995, Giambelluca *et al* 1997), reductions in transpiration, and reductions in surface roughness (Davin and de Noblet-Ducoudré 2010). Crop leaf area varies with phenology, limiting transpiration rates in young plants, while crop and pasture species are shallow-rooted relative to trees, limiting plant access to soil moisture reserves and reducing transpiration, particularly during droughts or the dry season (Oliveira *et al* 2005a, 2005b). Lower transpiration implies a greater partitioning of incoming radiative energy to sensible heat, and thus heating of the land surface. Similarly, lower height of the vegetation canopy in agriculture versus forest or savanna reduces land surface roughness, increasing aerodynamic resistance, and impeding turbulent

Figure 1. (A) Schematic showing forest loss adjacent to and remote from a weather station. The effect of non-local forest loss is assumed to decay with distance from the point of observation. The change in air temperature sensed at the weather station is assumed to arise from the sum of non-local effects at multiple distances (denoted '1' and '2' in panel A). The lower panel illustrates the transport of air temperature anomalies (B) induced by deforestation-driven energy balance and land surface roughness anomalies. These anomalies are illustrated in Panel (C) and show how land cover conversion to grassland or agriculture reduces latent heat fluxes (λ) and increases sensible heat fluxes (H) relative to forests, while reduced surface roughness increases the aerodynamic resistance of the land surface, and reduces dissipation of heat (shown by the scale of the turbulent eddy in the figure).

exchange with the atmosphere. Consequently, the loss of natural ecosystems in the tropics results in a land surface that tends to warm faster and dissipate heat less effectively than a comparable area of intact tropical forest or woody savanna (Feddema *et al* 2005, Ban-Weiss *et al* 2011, Devaraju *et al* 2018). Of the myriad consequences such changes have for the climate, we focus on one; changes in daily maximum temperatures. Our interest is motivated in part by the aim to understand the extent to which limiting (accelerating) land cover change in tropical ecosystem frontiers can reduce (amplify) temperature increases in regions that already face disproportionate challenges to maintaining economic productivity due greenhouse gas emission-derived global change (Diffenbaugh and Burke 2019, Masuda *et al* 2019).

Biogeophysical changes to the land surface energy balance cause changes in climate at the site of land cover change (hereafter 'local') and cause effects that propagate over a broad range of spatial scales (hereafter 'non-local' and outlined conceptually in figure 1). General Circulation Models (GCMs) have shown that to date, local climate responses to land cover change can rival greenhouse gas emissions in scale (Lawrence and Vandecar 2015). Additional studies have also used GCMs to understand 'non-local' climate change from land cover change. For example, at mesoscales (10–100 km), differential heating of the

planetary boundary layer due to heterogeneity in land cover can create lateral temperature and pressure gradients and change circulation (Baidya Roy and Avissar 2002). Such changes in mesoscale circulation may be responsible for increased rainfall in deforested areas (D'Almeida *et al* 2007). One recent GCM experiment convincingly decomposed biogeophysical climate change into 'local' and 'non-local' components. The study tracked temperature and precipitation response to 'local' and 'non-local' land cover change, finding each to be of considerable and comparable magnitude (Winckler *et al* 2017). While helpful for process understanding and for tracking changes occurring over large length scales, however, many GCMs represent Earth on a relatively coarse (e.g. ≈50 km) grid, making them poorly suited to explore either 'local' or 'non-local' micro-scale (1–10 km) phenomena.

Micro-scale effects of land cover heterogeneity have thus primarily been documented in observational studies, which suggest that there are indeed important influences of land cover variation at these scales. For instance, when considering the urban heat island effect in 32 metropolitan areas in China, air temperatures (ATs) 3–6 km from the urban boundary were elevated (relative to control sites located away from cities) (Zhou *et al* 2015). Similarly, maximum daily summer ATs within Madison, Wisconsin were found to vary by as much as 3.5 °C (Ziter *et al* 2019).

Proximity to tree cover was found to be the single biggest source of variation in maximum temperatures. It led to cooling of up to 1.5 °C, an effect that dominated impervious surface density in explaining the variation observed (Ziter *et al* 2019). Although the transport mechanisms responsible for these examples of micro-scale non-local warming were not identified, at scales of kilometers and greater, advection of sensible heat fluxes appears to be a probable process (D'Almeida *et al* 2007, Lawrence and Vandecar 2015), and see figure 1. To date, to our knowledge, no similar observational studies have been undertaken to identify the 'non-local', micro-scale influence of forest loss on maximum temperatures. Given the extent of recent land cover change in the tropics and the magnitude of temperature response to forest loss, micro-scale non-local warming may represent an important mechanism for tropical forest loss to shape regional climate.
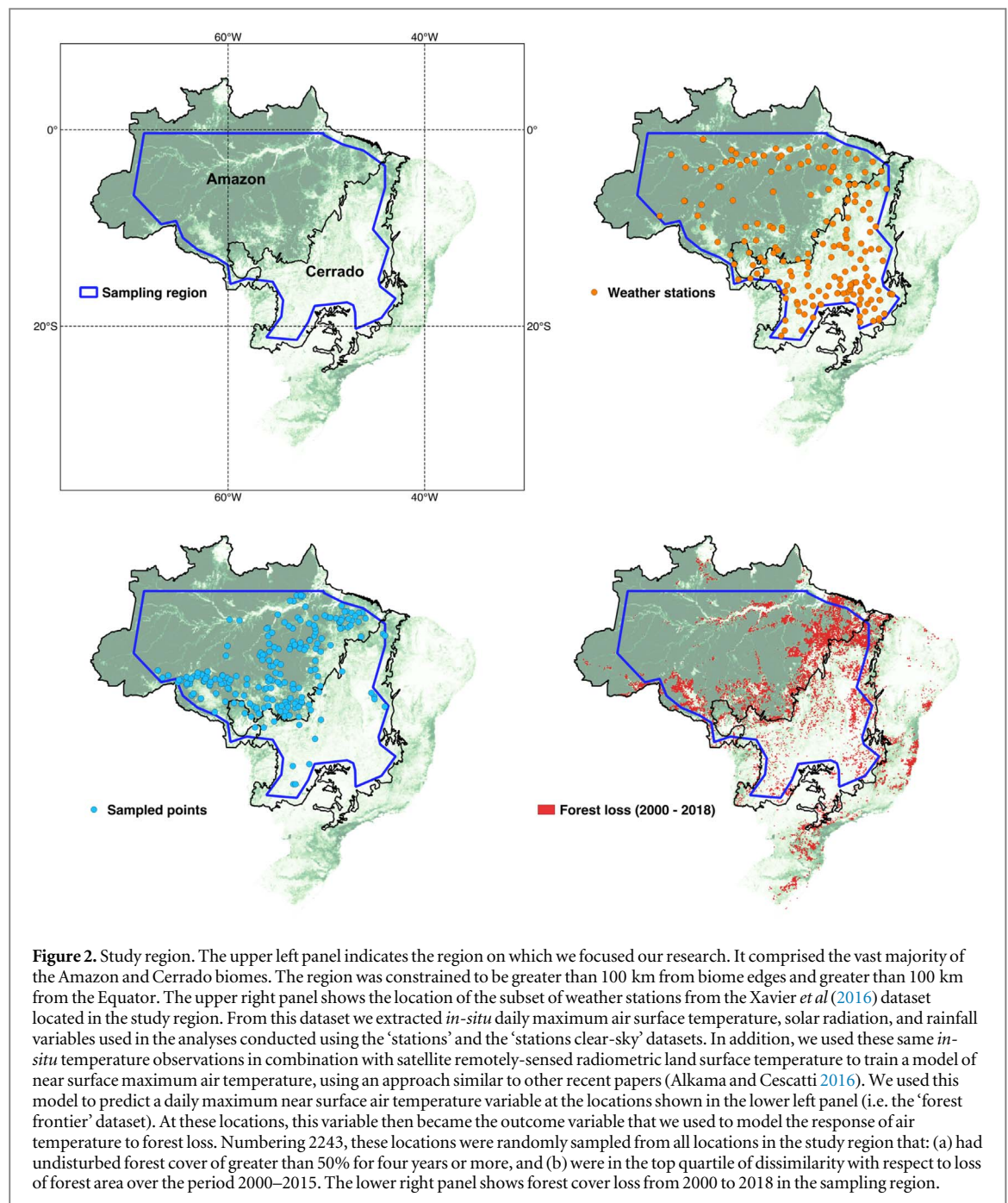
Recently, a number of observational studies (notably Malyshev *et al* 2015, Alkama and Cescatti 2016, Winckler *et al* 2019) robustly identified the local effects of land cover change on temperature while controlling for or removing temperature changes arising from neighboring land cover change. In this study, we aim to provide a similarly robust assessment of the converse, that is of the micro-scale maximum AT response to non-local forest loss occurring at distances from 1 to 60 km from the location of temperature monitoring. Our approach controls for local land cover change by confining the study to temperature monitoring sites where no land cover change occurs. It also makes statistical adjustment for coarser scale temperature trends. Our region of focus, mapped in figure 2 is Brazil's Amazon and Cerrado regions. The region ranges from Longitude $-67.41°$ to $-43.66°$ and Latitude $-20.96°$ to $-0.56°$. We selected this region because it has globally important biodiversity, exerts powerful controls on Earth's climate system, is home to millions, harbors a dynamic and thriving agricultural sector (Gibbs *et al* 2015, Strassburg *et al* 2017), and is experiencing rapid rates of forest cover conversion.

We used a combination of satellite and weather station temperature records for the period 2000–2015 (Xavier *et al* 2016) (see figure 2 for locations), to analyze how maximum AT depended on changing forest cover in a region located 1–60 km from a target site. Weather stations sites tend to be in regions that have less forest cover and experience less forest cover change (see figure S1 is available online at stacks.iop.org/ERL/14/084047/mmedia). In addition, the relatively small number of locations, $n = 209$ in the sampled region, limits the amount of variation in non-local forest cover change across length scales. This therefore limits the potential to attribute temperature change to forest loss by distance from the location of temperature observation. To address representativeness and to be able to better understand how the influence of forest loss on temperature varies with distance,

we created a second dataset of maximum ATs at sites that were (1) undisturbed forest during the entirety of the study period, and (2) were selected because they experienced a range of changes in 'non-local' ($\leqslant 50$ km) forest cover. We generated this dataset by training a model of near surface AT at weather stations as a function of remotely sensed forest radiometric land surface temperature (LST) and using the model to predict near surface AT at millions of forested locations across the study region where remotely sensed LST was available (Liu *et al* 2017). This approach is preferable to alternatives such as using gridded products derived by interpolating weather station data (Mildrexler *et al* 2011), which are sparse and provides particularly poor coverage of areas of rapid agricultural expansion in the tropics (Pielke *et al* 2011, Oliveira *et al* 2013, Cohn *et al* 2016). Our study region, the combination of the Brazilian Amazon and Cerrado region, constitutes approximately 4% of the Earth's ice free land surface. But during our study period, it contained fewer than 200 weather stations per year; i.e. just 0.5% of the 40 000 weather stations used for constructing gridded global temperature records (Rhode *et al* 2013)—or one weather station for each 22 000 km² . Clearly this is insufficient to represent effects arising at <60 km as targeted in this analysis.

We first statistically modeled the response of temporal anomalies in maximum daily AT to temporal anomalies in non-local forest cover, adjusting for antecedent soil moisture and rainfall anomalies as well as anomalies in two indicators of solar radiation, two sea surface temperature (SST) indices predictive of the region's climate, and region-wide annual temperature anomalies. In this initial approach, we employed a single non-local forest cover index, which we generated by weighting forest cover based on distance from the location of temperature observation using exponential decay functions with varying length scales, and integrating the result. This approach revealed that AT increased as the non-local forest cover index decreased, but did not provide insights into the relative contribution of forest loss at different distances to AT changes. To discriminate effects occurring on different length scales, we binned forest cover in four annuli ('halos') located 1 and 2 km, 2 and 4 km, 4 and 10 km, and 10 and 50 km from the location of the temperature observations, and performed a set of regressions that were identical to the first set except that they substituted out the single non-local forest cover index for various combinations of the halos.

The results indicate that non-local forest cover loss caused increases in maximum AT at distances of 1–60 km. In our forest frontier dataset, these changes were considerable and robust to a variety of model specifications. The magnitude of the effect per km² forest loss decays rapidly with increasing distance from the site of interest. However, because the area of land cover change able to influence temperature at any given location increases with the square of the distance

Figure 2. Study region. The upper left panel indicates the region on which we focused our research. It comprised the vast majority of the Amazon and Cerrado biomes. The region was constrained to be greater than 100 km from biome edges and greater than 100 km from the Equator. The upper right panel shows the location of the subset of weather stations from the Xavier *et al* (2016) dataset located in the study region. From this dataset we extracted *in-situ* daily maximum air surface temperature, solar radiation, and rainfall variables used in the analyses conducted using the 'stations' and the 'stations clear-sky' datasets. In addition, we used these same *in-situ* temperature observations in combination with satellite remotely-sensed radiometric land surface temperature to train a model of near surface maximum air temperature, using an approach similar to other recent papers (Alkama and Cescatti 2016). We used this model to predict a daily maximum near surface air temperature variable at the locations shown in the lower left panel (i.e. the 'forest frontier' dataset). At these locations, this variable then became the outcome variable that we used to model the response of air temperature to forest loss. Numbering 2243, these locations were randomly sampled from all locations in the study region that: (a) had undisturbed forest cover of greater than 50% for four years or more, and (b) were in the top quartile of dissimilarity with respect to loss of forest area over the period 2000–2015. The lower right panel shows forest cover loss from 2000 to 2018 in the sampling region.

from that location, under scenarios of uniform loss of forested area, the increase in temperature due to forest loss at long distances is nevertheless substantial. Overall, we find that maximum AT can be expected to increase from 0.5 °C to 0.95 °C in response to a uniform 25 percentage point decline in forest cover within a 50 km. This increase is comparable in magnitude to estimates of maximum temperature change due to local tropical forest loss (approximately 0.5 °C ± 0.02 °C, a value obtained by applying the empirical results of Alkama and Cescatti (2016) to a similar hypothetical uniform, 25 percentage point forest cover decline). That our non-local findings are similar in magnitude to these local findings, combined with the fact that both methodologies deliberately exclude the

other driver of warming, suggests that using either approach in isolation would underestimate the warming attributable to tropical forest loss.

2. Methods

2.1. Sampling

We used statistical modeling to estimate the response of maximum AT to non-local forest loss at two sets of locations: (i) weather stations included in the Brazil-wide network described by Xavier *et al* (2016) (referred to as the 'stations' dataset), and (ii) forested sites selected to exhibit a high degree of dissimilarity in forest cover change across different spatial scales (defined by annuli, or halos, of 1–2 km, 2–4 km,

4–10 km, and 10–50 km distance from each location of temperature observation). This latter dataset centers on the epicenter of conversion of natural ecosystems to agriculture in Brazil (see figure 2, bottom left panel), and is referred to as the 'forest frontier' dataset.

We generated the forest frontier dataset by identifying and then sampling from a set of locations that exhibited high dissimilarity in forest loss over the period 2000–2015 across halos at distances 1–2 km, 2–4 km, 4–10 km, and 10–50 km from the study site. We prepared the sample by first using Google Earth Engine (Gorelick *et al* 2017) to flag all forested sites of 1 km or greater within our study region (see figure 2) that maintained >50% cover for at least a four year period between 2000 and 2015 according to the Global Forest Cover dataset (Hansen *et al* 2013). These sites numbered in the tens of millions. Next, we computed a continuous measure of forest loss, ranging from 0% to 100% over the study period for each of the four halos surrounding each study site; 1–2 km, 2–4 km, 4–10 km, and 10–50 km. Following that, we classified the continuous halo forest loss metrics into quintiles ranking the quantity of forest loss occurring at a given location in a given halo relative to other locations for that same halo. We reference these categorical variables as C_i^r where i indicates the location and r indicates the halo. We next computed a version of the Multivariate Ordinal Getis dissimilarity Index D (Meng *et al* 2006) to estimate the degree of dissimilarity in loss categories across halos at each location. The index is a continuous measure of unevenness in the makeup of a population, and is computed here as:

$$D_i = 1 - \frac{\sum_{j=1}^{m}|\Delta C_i^{r_o} - \Delta C_i^{r_p}|}{\max_{j}(\sum_{j=1}^{m}|\Delta C^{r_o} - \Delta C^{r_p}|)}, \qquad (1)$$

where D_i is an index of dissimilarity in forest change across halos at each location, i, in the 'forest frontier' dataset. Across all combinations of halos, j from 1 to m, we sum the absolute value of the difference between $C_i^{r_o}$ and $C_i^{r_p}$, where o is the first halo in the combination and p is the second halo in the combination. We divide this sum by the maximum possible value of this same sum across all locations. Therefore the lower the value of D_i, the more dissimilarity at the location. Next we grouped the continuous values of D_i into quartiles. Finally, we sampled 2243 points from the lowest (i.e. most dissimilar) quartile. These points are the locations contained in the 'forest frontier' dataset (see figure 3 for a generalized workflow).

Statistics summarizing non-local forest cover and forest cover change in both the 'stations' and 'forest frontier' datasets are presented in table 1, while figure S1 illustrates the differences in the distribution of forest cover and forest change across these datasets. Table 1 also contains summary statistics describing the regression variables described in sections 2.4 and 2.5. Summary statistics for forest cover and forest loss for each of the halos used to calculate the dissimilarity

index and used in the regressions described in 2.5 are provided in table S1.

2.2. Daily maximum temperature

Our primary response variable was daily maximum temperature, AT, which we obtained from weather stations or predicted from remotely sensed radiometric LSTs. For the station dataset we used AT observations from Xavier *et al* (2016) (for the years 2000–2015), subject to quality control procedures as described therein. We also created a dataset consisting of the subset of station observations that had clear-sky conditions at the time of either the MODIS Aqua or MODIS Terra overpass. The clear-sky days were selected as those when either MODIS Aqua or MODIS Terra reported a radiometric LST with quality flag 0. This quality flag signified that the data were of highest quality with respect to a range of quality assurance metrics including low errors in both emissivity and LST—a situation that only occurs under clear-sky conditions. The purpose of the station clear-sky dataset was to facilitate comparison between *in-situ* temperature measurements at weather stations, and the AT values in the forest frontier dataset. This subset of station AT data are denoted AT|clearsky. Temperature data for the stations and clear-sky datasets, and their anomalies during the study period, are summarized in table 1.

In the forest frontier dataset, we predicted AT from remotely sensed LST, available only on days that met the same quality flagging criteria as AT|clearsky. Predicting AT from remotely sensed LST raises several challenges: (i) relationships between satellite observations of LST and AT are not necessarily consistent across land cover types (Mildrexler *et al* 2011, Li *et al* 2015), (ii) due to the position of the MODIS sensor and the curvature of earth, wide zenith angles and the gridding process can mean that as much as 50% of the signal assigned to a MODIS pixel can stem from locations outside the pixel (Peng *et al* 2015), and (iii) observations are only available under clear-sky conditions (motivating the comparison with the AT|clearsky dataset). With respect to the first challenge, we focused our analysis on relatively undisturbed forested locations within 2 km of *in-situ* weather stations, recognizing that ATLST relationships are most reliable at forests (Li *et al* 2015). Undisturbed forested locations were identified as those that: (i) had at least 50% forest cover in the year 2000 (based on a resampling of 30 m forest cover from the Global Forest Change dataset) and (ii) were observed to have less than 1% change in forest cover for at least three consecutive years during the period 2000–2015. To address the second challenge, we also required that at least a one grid cell forested buffer was present around these forested locations. The third challenge has implications for both interpretation and analysis of the effects of forest loss using the AT dataset. Our analysis
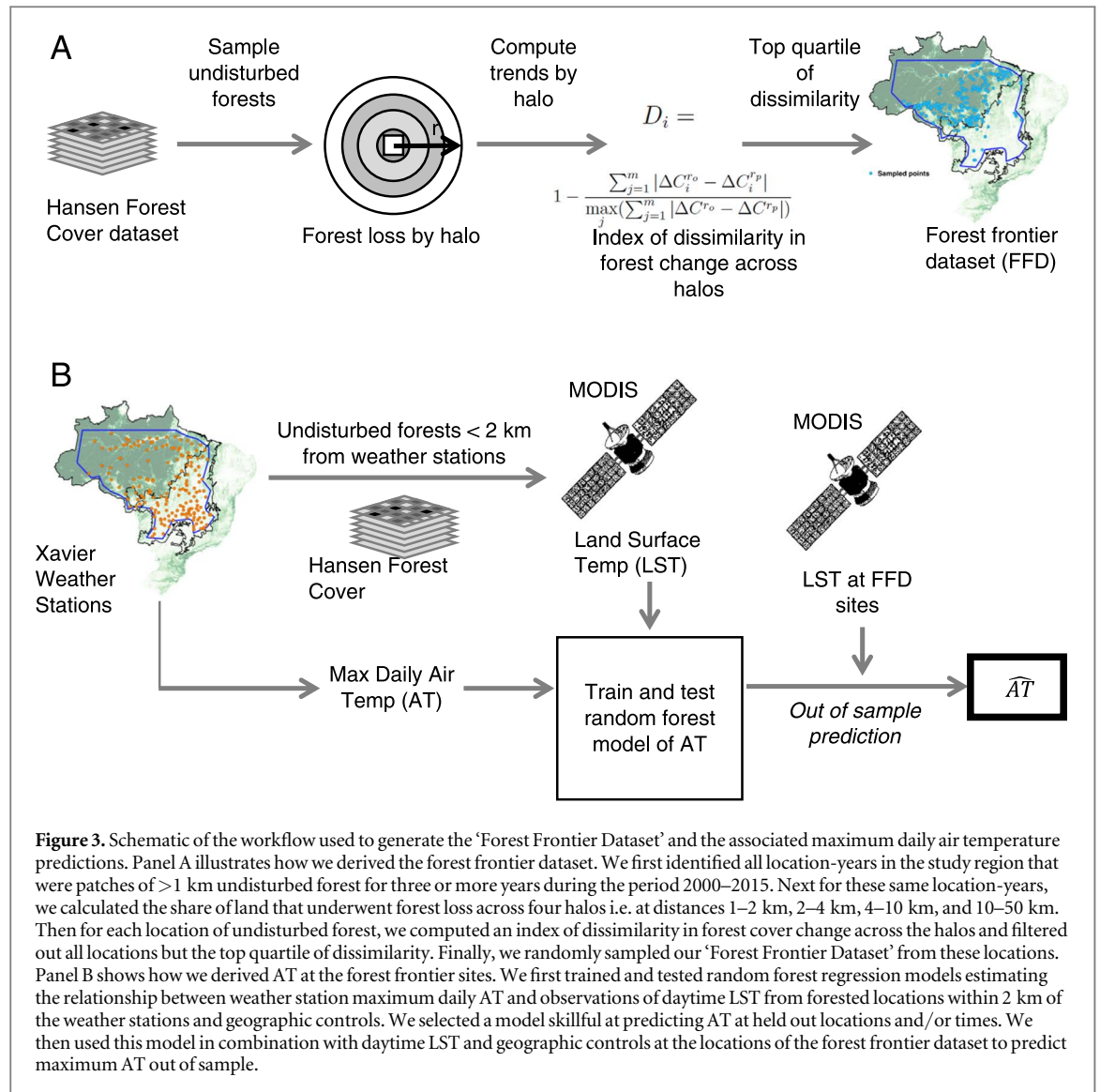
Figure 3. Schematic of the workflow used to generate the 'Forest Frontier Dataset' and the associated maximum daily air temperature predictions. Panel A illustrates how we derived the forest frontier dataset. We first identified all location-years in the study region that were patches of >1 km undisturbed forest for three or more years during the period 2000–2015. Next for these same location-years, we calculated the share of land that underwent forest loss across four halos i.e. at distances 1–2 km, 2–4 km, 4–10 km, and 10–50 km. Then for each location of undisturbed forest, we computed an index of dissimilarity in forest cover change across the halos and filtered out all locations but the top quartile of dissimilarity. Finally, we randomly sampled our 'Forest Frontier Dataset' from these locations. Panel B shows how we derived AT at the forest frontier sites. We first trained and tested random forest regression models estimating the relationship between weather station maximum daily AT and observations of daytime LST from forested locations within 2 km of the weather stations and geographic controls. We selected a model skillful at predicting AT at held out locations and/or times. We then used this model in combination with daytime LST and geographic controls at the locations of the forest frontier dataset to predict maximum AT out of sample.

features a number of choices that account for the clear-sky day bias, as outlined in section 2.5. Interpretation issues are revisited in the Discussion.

2.3. Forest cover

Forest cover was treated in two ways: (i) as an aggregated non-local forest cover metric NL^L, or (ii) as the total forested area within prescribed annular bands ('halos') spanning the distance 1–50 km from the site. To generate the non-local forest cover metric, we firstly defined a radial exponential kernel with the function:

$$NL^L = \sum_j X_j C e^{-r_j/L}, \qquad (2)$$

where X_j is the percent cover at pixel j, r_j is the Euclidean distance in km between the location of the target pixel x, y and a non-local pixel x_j, y_j, and L is a length scale determining how rapidly the influence of non-local forest cover decays. To ensure the weights add to 1 requires that:

$$C = \frac{1}{\sum_j e^{-r_j/L}}. \qquad (3)$$

We used four decay length scales: $L = 2, 5, 10,$ and 20 km. This kernel was used to weight the forest cover surrounding each pixel in the forest frontier dataset for each year from 2000 to 2015, and the weighted forest cover was then summed within a $3L$ radius of the target pixel to generate a non-local forest cover metric.

The forest cover within the halos was computed for each year from 2000 to 2015 from the Global Forest Cover dataset. Forest cover was always computed using four 'halos' (1–2 km, 2–4 km, 4–10 km, and 10–50 km), however we also aggregated the four halos in various ways (e.g. 1–10 km, 1–50 km, 10–50 km, 4–50 km), to enable us to explore the sensitivity of our results to the granularity with which forest cover is described by distance. We only prepared forest cover halos for the forest frontier dataset, as there was too much collinearity in trends in forest loss at all scales in the stations datasets to allow a balanced analysis across halos for those sites.

Table 1. Summary statistics for the datasets used in the analysis: station, clear-sky station and forest frontier. The forest cover change metric reported here shows the difference between forest cover in the initial and final years of the dataset, 2000 and 2015. The forest cover dataset is common to the station and clear-sky datasets, the forest frontier dataset contains separate forest cover attributes.

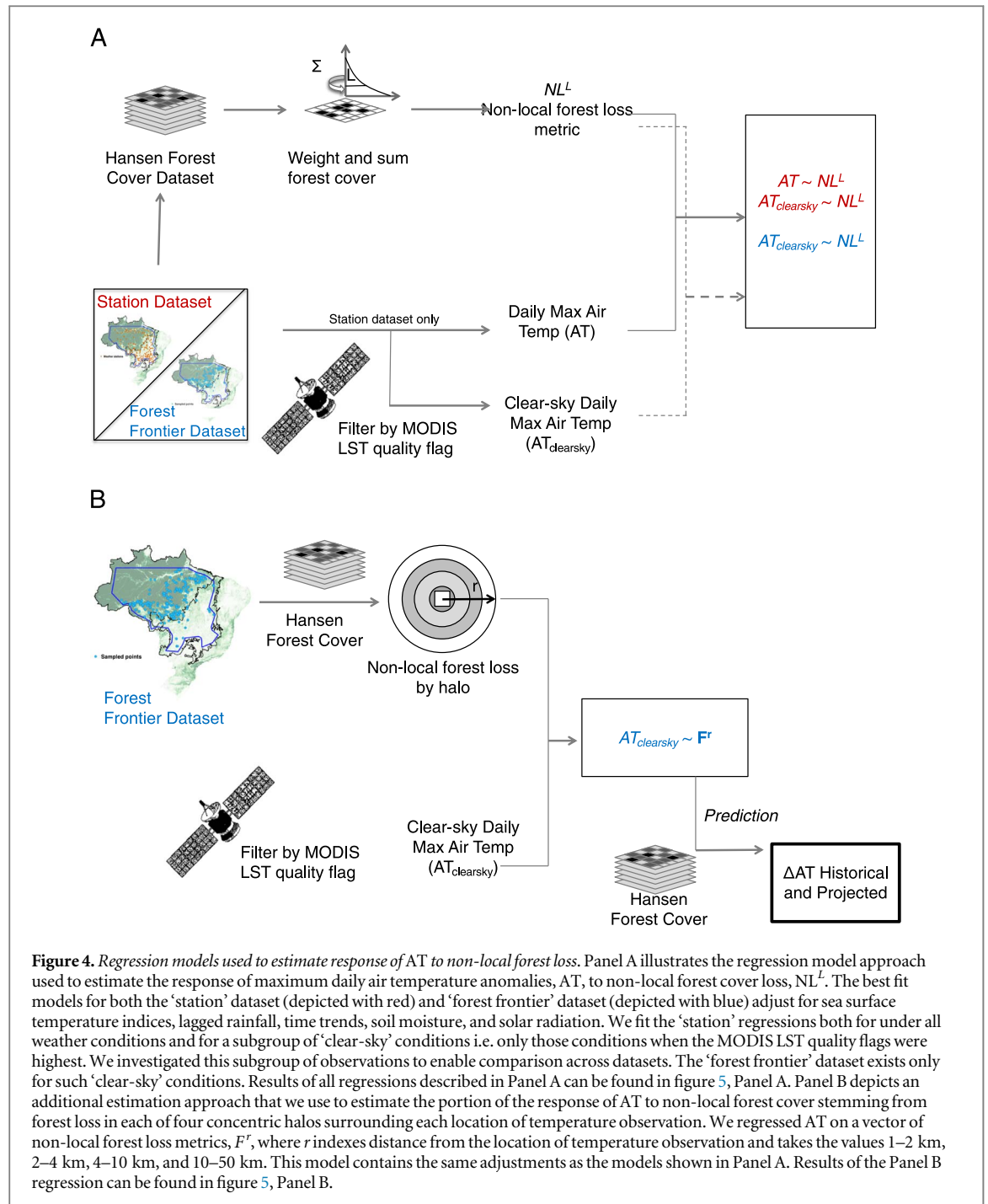|	Mean	SD	Min	Max
(A) SST indices (common between datasets)				
Niño 4 SST	28.65	0.79	26.92	30.13
TNA	0.40	0.37	−0.37	1.40
(B) Stations dataset				
AT (°C)	31.82	3.18	4.20	45.00
AT anomaly (°C)	0.00	3.18	−27.62	13.18
F(1–50 km) (km^2)	2985.00	2045.00	472.90	7649.00
F(1–50 km) change (km^2)	102.70	158.20	1.61	1026.00
P(lag = 1 d) (mm)	4.62	8.84	0.00	153.00
P(lag = 2 d) (mm)	4.62	8.84	0.00	153.00
P(lag = 3 d) (mm)	4.62	8.85	0.00	235.10
P(lag = 4 d) (mm)	4.62	8.85	0.00	235.10
P(lag = 5 d) (mm)	4.62	8.84	0.00	235.10
Soil moisture (kg m^{-2})	27.96	9.51	3.84	45.31
TOA radiation (W m^{-2})	33.91	4.20	22.63	40.58
Net surface radiation (W m^{-2})	18.41	4.18	1.12	34.62
(C) Stations dataset (clear-sky)				
AT (°C)	32.69	3.48	4.20	42.50
AT anomaly (°C)	0.00	3.48	−28.49	9.81
P(lag = 1 d) (mm)	0.35	2.41	0.00	66.71
P(lag = 2 d) (mm)	0.55	3.04	0.00	76.39
P(lag = 3 d) (mm)	0.72	3.56	0.00	101.30
P(lag = 4 d) (mm)	0.83	3.92	0.00	101.30
P(lag = 5 d) (mm)	0.83	3.80	0.00	84.48
Soil moisture (kg m^{-2})	18.70	8.34	4.39	43.27
TOA radiation (W m^{-2})	30.10	3.59	22.63	40.55
Net surface radiation (W m^{-2})	20.78	2.84	4.54	34.62
(D) Forest frontier				
AT (°C)	32.59	1.69	14.36	40.94
AT anomaly (°C)	0.00	1.69	−18.23	8.35
F(1–50 km) (km^2)	5857.00	1459.00	847.70	7744.00
F(1–50 km) change (km^2)	676.00	447.00	18.34	2649.00
P(lag = 1 d) (mm)	1.37	4.64	0.00	151.00
P(lag = 2 d) (mm)	1.66	5.14	0.00	151.00
P(lag = 3 d) (mm)	1.77	5.33	0.00	151.00
P(lag = 4 d) (mm)	1.84	5.45	0.00	154.50
P(lag = 5 d) (mm)	1.90	5.56	0.00	151.00
Soil moisture (kg m^{-2})	25.09	8.23	5.06	44.99
TOA radiation (W m^{-2})	31.23	3.17	22.63	40.58
Net surface radiation (W m^{-2})	19.24	2.68	4.78	34.27

In summary, the primary outcome and predictor variables used for analysis were: for the station sites, the two daily maximum AT datasets, AT and AT|clearsky, and four non-local forest cover metrics NL^L. For the forest frontier sites, one maximum AT dataset AT, four non-local forest cover metrics NL^L, and forest cover within four halos (and their aggregations). These datasets, as well as descriptive statistics for all ancillary variables used as covariates in the regression models, are summarized for the stations, clear-sky and forest frontiers datasets in table 1.

2.4. The response of AT to non-local forest loss
We estimated the response of temporal anomalies in AT and AT|clearsky at the weather stations to changes in non-local forest cover metrics NL$_L$ and the response of AT in the forest frontier dataset to these same non-local forest cover metrics. We schematically show the approach in figure 4. We used Weighted Least Squares regression to perform this modeling. Data points were weighted based on average forest cover over the study period such that the mean of the product of the weights with forest cover was 0.45, which is approximately equal to the mean forest cover in the study region for the period of study. The goal of this weighting was to correct for the over-sampling of heavily forested regions in the forest frontier dataset relative to the study region as a whole and the over-sampling of the sparsely forested regions in the station dataset (see figure S1 for histograms of forest cover in the forest frontier dataset versus the stations dataset). Because both the station dataset and the station clear-sky dataset comprise the same locations and differ only in dates of observation, they have the same weights. The locations of temperature observations in the forest frontier dataset differ from the locations of temperature observation in the station dataset and therefore the weights differ. Weighting allows us to interpret the resulting coefficient estimates as the average influence of forest loss on temperature across the study region.

The regression approach adopted in this task attempts to replicate the treatment-control nature of a clinical trial and follows a widely used statistical approach in the climate impacts literature (Hsiang 2016). It assumes that the locations in the study have been randomly 'assigned' particular forest loss profiles across years. The anomaly on anomaly structure of our model controls for confounding factors that vary uniformly in space or time. We are thus treating each location as both treatment and control; in effect comparing the temperature at each location under more forest loss to its temperature under less forest loss.

The regression equations used to estimate the coefficients of forest loss on temperature reported in the main text were of the form below (where AT is replaced by AT|clearsky for the clear-sky dataset, and

Figure 4. *Regression models used to estimate response of* AT *to non-local forest loss.* Panel A illustrates the regression model approach used to estimate the response of maximum daily air temperature anomalies, AT, to non-local forest cover loss, NL^L. The best fit models for both the 'station' dataset (depicted with red) and 'forest frontier' dataset (depicted with blue) adjust for sea surface temperature indices, lagged rainfall, time trends, soil moisture, and solar radiation. We fit the 'station' regressions both for under all weather conditions and for a subgroup of 'clear-sky' conditions i.e. only those conditions when the MODIS LST quality flags were highest. We investigated this subgroup of observations to enable comparison across datasets. The 'forest frontier' dataset exists only for such 'clear-sky' conditions. Results of all regressions described in Panel A can be found in figure 5, Panel A. Panel B depicts an additional estimation approach that we use to estimate the portion of the response of AT to non-local forest cover stemming from forest loss in each of four concentric halos surrounding each location of temperature observation. We regressed AT on a vector of non-local forest loss metrics, F^r, where r indexes distance from the location of temperature observation and takes the values 1–2 km, 2–4 km, 4–10 km, and 10–50 km. This model contains the same adjustments as the models shown in Panel A. Results of the Panel B regression can be found in figure 5, Panel B.

the regressions were run separately for AT at the stations and AT at the forest frontier dataset):

$$\mathrm{AT}_{it} = \gamma_0 + \gamma_1 \mathrm{NL}^L_{iy} + \gamma_2 \mathbf{P}_{\mathbf{i}, t-z} + \gamma_3 \mathbf{R}_{\mathbf{i}t}$$
$$+ \gamma_4 \mathbf{SST}_t + \gamma_5 \mathrm{soil}_{it} + t_y + c_i + \epsilon_{it}. \quad (4)$$

In this expression, i indexes the weather station (or equivalently, the locations at which we predicted AT in the forest frontier dataset), t indexes the day of the time-series, y is the year, and NL^L is the non-local forest cover computed with decay length-scale L. The regression adjusts for a number of variables which would be expected to influence maximum daily temperature independently of forest cover condition. First is a vector of antecedent rainfall, $P_{i,t-z}$. We adjust for

rainfall from each day in a five day period (indexed by z) prior to the temperature observation. These data were obtained from Xavier *et al* (2016) for the stations dataset and from CHIRPS (Funk *et al* 2015) for the forest frontier dataset. We also adjusted for two metrics of solar radiation $\mathbf{R}_{\mathbf{i}t}$, a measure of net solar radiation collected at weather stations on the day of the temperature observation and tested for quality by Xavier *et al* (2016), and top of atmosphere radiation calculated using a formula based on latitude and day of year (Allen *et al* 1998). Additionally, we adjusted for $\mathbf{SST}_t$, a vector of two SST indices, the Tropical North Atlantic SST index (Enfield *et al* 1999) and the Niño 4 index which captures SST anomalies in the central equatorial

Pacific associated with El Niño (Rayner *et al* 2003). Next, we adjusted for soil$_{it}$, an estimate of soil moisture in the top 10 cm of the soil obtained from GLDAS 2.1 (Rodell *et al* 2004). The variable t_y is a dummy of space-invariant year fixed effects; and c_i is a time-invariant spatial unit fixed effect. Visual inspection of NL^L versus AT plots suggested a linear relationship, motivating the linear form of equation (4). Errors (ϵ_{it}) were clustered at level of locations, because we wish to use our findings to estimate the influence of forests on temperature across the Cerrado-Amazon region, but there are locations in that region that are not represented in the sample (Abadie *et al* 2017). The model results can be interpreted as providing a regionally representative estimate of the response of AT to forest loss.

We tested the robustness of the regressions to different specifications of the model, including different treatments of year, clear-sky classification criteria, and decay length scale. As shown in the supplementary information (figures S3 and S4), the regression results are robust to these changes of specification. We also explored the impact of removing weighting and error clustering from the stations and stations clear-sky models. The results of these model versions tell us about the forest-temperature relationship at weather stations and at weather stations on clear-sky days, but they do not tell us about the forest-temperature relationship in the wider region.

The outcome of these regressions allows us to observe the overall effects of change in spatially-weighted forest cover on maximum daily temperatures. By comparing the effect sizes for AT and AT|clearsky at the stations, we gain insight into whether the effects of non-local forest cover on temperature vary with the degree of solar heating. By comparing the effect sizes for AT|clearsky and AT in the forest frontier dataset we can learn about the degree of bias that may be occurring in estimates made at weather stations due to the tendency for weather stations to be located in regions with generally low forest cover.

2.5. Distance decay in the response of AT to non-local forest loss

Having modeled the response of daily maximum temperatures to non-local forest cover in aggregate, we next implemented an approach for estimating the sensitivity of the influence of forest loss on temperature to the distance between the location of forest loss and the location of the temperature observation. We did this by disaggregating non-local forest cover into four halos located at progressively increasing distances from the temperature observation (shown schematically in figure 4). We estimated multiple model specifications that varied according to which halos we included in the model and which sets of halos (if any) we aggregated. These regressions, shown in

equation (5) below therefore differed from those described in equation (4) in their representation of forest cover. They also differed in that they were estimated only on the forest frontier dataset to avoid issues with multi-collinearity in the stations dataset

$$
\begin{aligned}
AT_{it} = \; & \beta_0 + \beta_1 F_{iy}^{r_1} + \beta_2 F_{iy}^{r_2} + \beta_3 F_{iy}^{r_3} + \beta_4 F_{iy}^{r_4} \\
& + \beta_2 \mathbf{P}_{i,t-z} + \beta_{10} \mathbf{R}_{it} + \beta_{11} \mathbf{SST}_t \\
& + \beta_{12} \text{soil}_{it} + t_y + c_i + \epsilon_{it}.
\end{aligned}
\tag{5}
$$

Here, the terms $F_{iy}^{r_1}$ through $F_{iy}^{r_4}$ represent total forest cover located in the halos at 1–2 km, 2–4 km, 4–10 km, and 10–50 km from the observation, for location i and year y. We clustered the error term ϵ_{it} at each observation. We again performed a number of robustness checks to determine that the results were broadly insensitive to the choice of e.g. time fixed effects versus time trends, and various halo groupings.

To investigate whether models were overfit, we performed k-fold cross validation using two ways to divide the data into test and training subsets: (i) by year and (ii) by location. For each division of the data, we followed a hierarchical approach where we firstly estimated temperature response to forest cover change with a model consisting only of forest cover and location and year fixed effects terms We then progressively added all other independent variables from equation (5). We started by adding single sets of independent variable (e.g. rainfall), then pairs of sets (e.g. rainfall and solar radiation), and then groups of three (e.g. rainfall, solar radiation, and SST). We compared the forest change estimates across these specifications and found them to be largely insensitive to alternative configurations of independent variables. We also tracked model R^2 and Root Mean Square Error (RMSE) calculated by comparing temperature predictions, $\widehat{AT}$, with held out values of AT. Further details about this process are provided in the supplementary information (S7.2 and table S8). The full model achieves the lowest RMSE (1.134 °C for train/test sets grouped by year, and 1.135 °C for train/test sets grouped by location). This suggests that the model is not overfit.

2.6. Predicting warming ($\Delta\widehat{AT}$) from non-local forest loss

Following estimation of the both the length scale and halo models in equations (4) and (5), we used the forest cover coefficient estimates obtained in combination with forest cover change scenarios to predict the marginal change in AT, $\Delta\widehat{AT}$, from non-local forest cover change (see figure 4). These predictions were the source of the estimates displayed in figures 5 and 6(B), and table 2, row 2. Figures 5 and 6(B), and table 2, row 2 each show predicted warming (and 95% Confidence Intervals) from a uniform 25 percentage point decline in non-local forest cover within 50 km of the location of temperature observation. Table 2 row 1 shows the results of predicting warming using historical forest
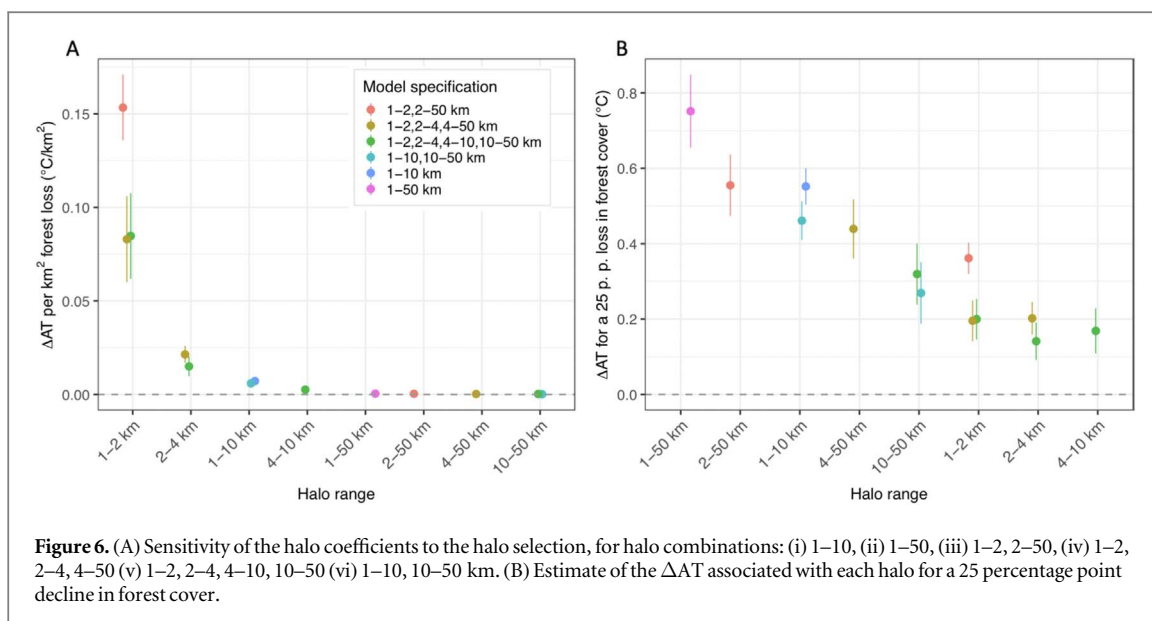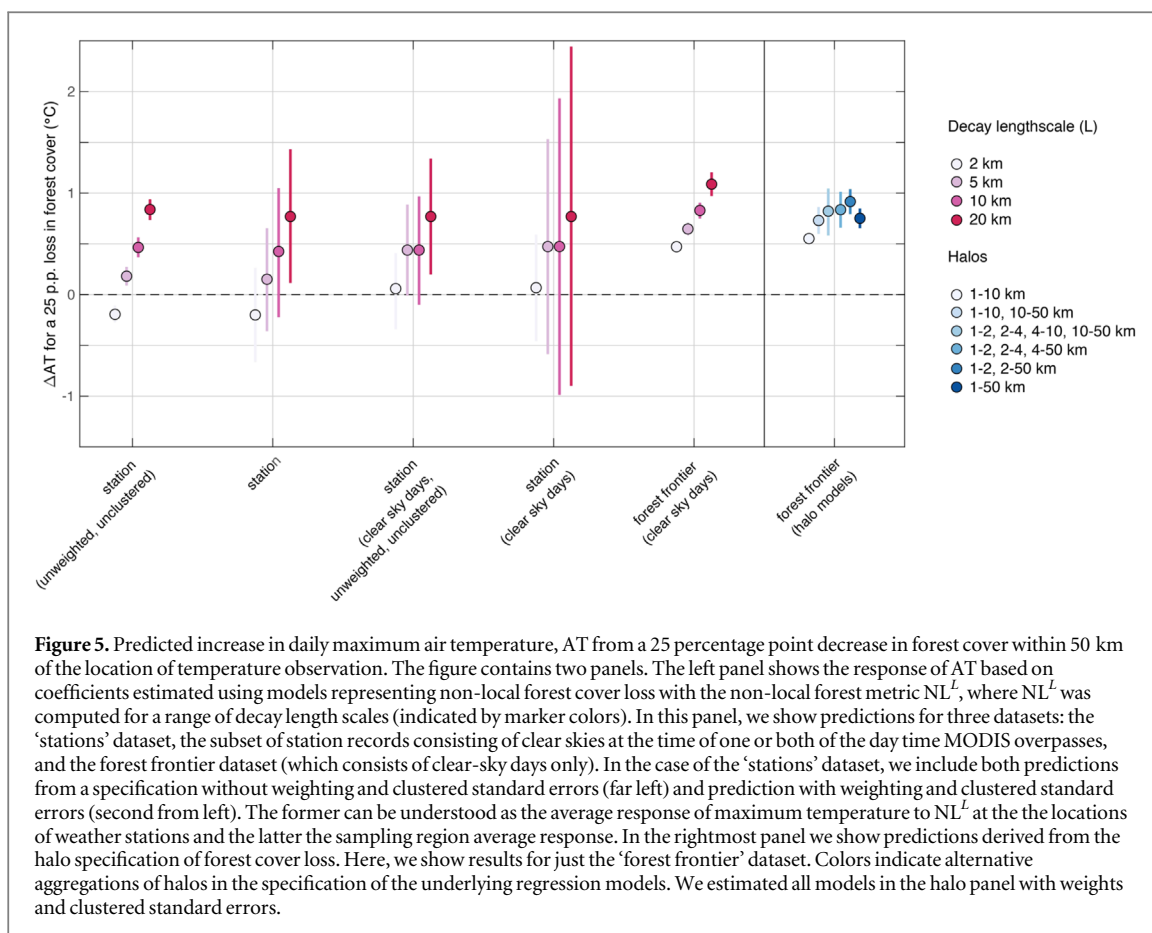
Figure 5. Predicted increase in daily maximum air temperature, AT from a 25 percentage point decrease in forest cover within 50 km of the location of temperature observation. The figure contains two panels. The left panel shows the response of AT based on coefficients estimated using models representing non-local forest cover loss with the non-local forest metric NL^L, where NL^L was computed for a range of decay length scales (indicated by marker colors). In this panel, we show predictions for three datasets: the 'stations' dataset, the subset of station records consisting of clear skies at the time of one or both of the day time MODIS overpasses, and the forest frontier dataset (which consists of clear-sky days only). In the case of the 'stations' dataset, we include both predictions from a specification without weighting and clustered standard errors (far left) and prediction with weighting and clustered standard errors (second from left). The former can be understood as the average response of maximum temperature to NL^L at the the locations of weather stations and the latter the sampling region average response. In the rightmost panel we show predictions derived from the halo specification of forest cover loss. Here, we show results for just the 'forest frontier' dataset. Colors indicate alternative aggregations of halos in the specification of the underlying regression models. We estimated all models in the halo panel with weights and clustered standard errors.

Figure 6. (A) Sensitivity of the halo coefficients to the halo selection, for halo combinations: (i) 1–10, (ii) 1–50, (iii) 1–2, 2–50, (iv) 1–2, 2–4, 4–50 (v) 1–2, 2–4, 4–10, 10–50 (vi) 1–10, 10–50 km. (B) Estimate of the ΔAT associated with each halo for a 25 percentage point decline in forest cover.

Table 2. Predicted temperature change associated with four different forest loss scenarios. Two historical cases were analyzed: one for the 2000–2015 period using the forest frontier AT dataset (scenario 1), and one for the 1985–2017 period using MapBiomas (scenario 3). Scenario 2 considers the regional impact of a hypothetical 25 percentage point reduction in forest cover applied uniformly through space.

	Forest loss scenario	Domain	ΔAT (°C)	95 CI % (°C)
1	Historical (2000–2015)	Locations of forest frontier dataset	0.20	±0.06
2	25 percentage point decline in forest cover	Sample region	0.81	±0.23
3	Historical (1985–2017)	Sample region	0.35	±0.10

loss (from Hansen GFC) at the locations of the forest frontier dataset. Table 2 row 3 shows warming predicted from the mean amount of net forest cover change in the Sample Region and over the period 1985–2017 as reported in the MapBiomas dataset (Souza and Azevedo 2017).

For halo models, to predict $\widehat{\Delta \text{AT}}$ from ΔF we used the equation:

$$
\begin{aligned}
\frac{\widehat{\Delta \text{AT}}}{\Delta F} = \hat{\beta}_1 {}^* \Delta F^{r_1} + \hat{\beta}_2 {}^* \Delta F^{r_2} \\
+ \hat{\beta}_3 {}^* \Delta F^{r_3} + \hat{\beta}_4 {}^* \Delta F^{r_4},
\end{aligned} \tag{6}
$$

where $\Delta F^r = F^r_{final} - F^r_{initial}$ was a vector of the change in forest cover (F) in halo r in km^2, and the coefficient estimates, $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ and $\hat{\beta}_4$ were the change in temperature per km^2 forest loss and were obtained from the estimation of regression equation (5).

For length scale models to predict $\widehat{\Delta \text{AT}}$ from ΔF we used the equation:

$$
\frac{\widehat{\Delta \text{AT}}}{\Delta \text{NL}^L} = \hat{\gamma}_1 {}^* \Delta \text{NL}^L. \tag{7}
$$

In this case, ΔNL^L was the change in forest cover and $\hat{\gamma}_1$ was the estimated response of AT from the loss of forest cover within up to 60 km.

2.7. Comparison to estimates of warming due to local forest loss

Alkama and Cescatti (2016) estimated biogeophysical warming in response to forest loss across all global regions for the 2003 and 2012 period. Similarly to our analysis, this study used remotely sensed radiometric LST to predict ATs. But in contrast to our approach, it estimated the difference over time in maximum AT between intact forest and paired deforested locations within 50 km, placing the highest weighting on the locations closest to the forested sites. This approach thus subtracts away as much temperature variation from non-local causes possible, including the non-local forest loss that is the focus of our study. In the Alkama and Cescatti (2016) study, the estimated influence of complete (i.e. 100%) local forest loss on AT for all tropical regions was 2.03 °C $\pm$ 0.06 °C. This would scale to a 0.22 °C $\pm$ 0.01°C increase in temperature due to the 11 percentage point historical forest cover loss in the forest frontier dataset, or a 0.50 °C $\pm$ 0.02 °C increase for a uniform 25 percentage point decline in forest cover.

2.8. Replication of results

We performed all analysis with R (R Core Team 2019). Code for the replication of analyses presented in the main text of this study can be found at https://doi.org/10.5281/zenodo.3267328. The code uses a dataset available for download at https://doi.org/10.17605/OSF.IO/WVR93.

3. Results

The results presented firstly consider the response of maximum daily AT to changes in the non-local forest cover metrics NL^L followed by the response of the maximum daily AT to forest cover delineated by the halos at different scales. We then use each set of models to predict temperature change under a variety of forest loss scenarios, and benchmark these predictions against the response of maximum AT to local forest loss. In the main text, we discuss the magnitude and uncertainty of the estimated influence of non-local forest cover on maximum temperature for individual non-local forest cover variables as well as groupings of forest cover variable. Full regression results for all non-local forest cover regressions and halo regressions can be found in the SI, tables S3–S6 and figures S3, and S4.

Figure 5 shows the response of maximum daily temperature, AT, to a 25 percentage point change in both NL^L and forest cover disaggregated into halos, estimated for three separate datasets, four different decay length scales L, and with and without weighting of datasets by forest cover and clustering of standard errors. For each dataset and regression type, dots show the mean marginal effect on AT of a 25 percentage point decline in forest cover at distances 1–50 km from the location of temperature observation. Bars in the figure indicate 95% confidence intervals around these predicted means.

Results in the leftmost column are from an Ordinary Least Squares regression of the response of AT to NL^L at weather stations. Across every length scale we find significant effects ($p < 0.001$). In all cases except the 2 km length scale we find that forest loss increases temperature at the observation point. The unexpected 2 km finding that forest cover loss may result from omitted variable bias associated with excluding forest loss at greater than 6 km, which, from other regressions is demonstrably influential on temperatures. The apparent cooling effect could also stem from the high degree of inverse correlation between coarse scale forest area near stations and the time trend in warming. Figure S3 demonstrates this, showing that in specifications of this regression omitting time trends or year fixed effects that this coefficient is significant and positive. The results point to the difficulty associated with clearly teasing apart time trends and forest cover change effects on thermal responses in areas with low forest cover.

The results of the leftmost regression in figure 5 can be interpreted as an estimate of the response of AT to non-local forest loss in locations across the study region with similar characteristics to weather stations. However, because stations are in locations with a disproportionately small amount of forest cover and forest cover change (see figure S1) they are not representative of the sample region in general and are particularly not representative of areas undergoing

large scale forest loss. To employ this dataset to explore the region-wide response of AT to forest loss it is necessary to perform Weighted Least Squares to weight more forested observations higher and to cluster standard errors. When we do this we fail to identify a significant impact of forest cover loss on maximum daily AT at the stations for either AT or AT|clearsky for length scales of <20 km.

Conversely, non-local forest cover loss has a statistically significant ($p < 0.001$) warming effect on AT in the forest frontier dataset, regardless of how forest cover is weighted by distance from the target location. The effect sizes are broadly comparable for all decay length scales, with a 25 percentage point decline in any non-local forest cover metric NL^L (corresponding to a 25 percentage point decline in forest cover from 1 to 50 km) resulting in $0.19\,°C \pm 0.04\,°C$ ($L = 2$ km) to $0.84\,°C \pm 0.05\,°C$ ($L = 20$ km) degrees of warming at the target pixel (weighted and clustered errors).

Uncertainty in the estimated effect sizes is relatively large within these regressions, particularly for the stations and especially the clear-sky days at the weather stations. These large confidence intervals may be a simple consequence of sample size, as evidenced by the much lower uncertainty bounds emerging in the larger and more spatially representative forest frontier dataset.

The remaining results concern the fitting of the 'halo' models to the forest frontier dataset. Regression tables detailing the halo model results are provided in table S7. Figure 5 shows the changes in AT that arise at the target pixel due to a uniformly applied 25 percentage point forest loss at scales of 1–50 km distance, using models estimated for various combinations of the four halos. As shown in the figure, the estimated impact on AT ranges from an increase of $0.55\,°C \pm 0.02\,°C$ to approximately $0.92\,°C \pm 0.06\,°C$, depending on the aggregation of the halos. The estimated temperature change is larger when the effects of the 1–2 km halo are computed separately from other halos, which is likely indicative of a reduction in non-local temperature influences with distance from the target pixel. The uncertainty in the estimate increases with the number of halos for which effects are independently estimated. Overall, the results suggest that the uncertainty in the mean estimated temperature effect associated with the spatial discretization of the halos is on the order of 50% variation in the means (e.g. from $0.5\,°C$ to $0.95\,°C$ for a 25 percentage point reduction in forest cover, see figure 5).

The effects of a uniform forest loss at any given length scale on temperature at the target pixel arise from the combination of two factors: (i) the effect size associated with a unit of forest loss at a given length scale, and (ii) the amount of forest loss at that length scale. Although the non-local effects of a unit of forest loss decline with distance from the target pixel, the area of forest available to be lost increases with

distances from the target pixel. Thus, under forest loss scenarios that are uniform in distance, the effect of forest loss on temperature declines more slowly with distance than it does when we consider the effect per unit area of forest loss. We show this contrast in figure 6(A) versus figure 6(B). Panel A shows the effect size associated with a km^2 of forest loss by halo grouping. Panel B scales these effects by uniform forest loss areas. In both panels A and B, the marker color indicates model specification and results are presented in order of declining effect size, from left to right on the *x*-axis. This figure shows a rapid decline in effect size as halos become more remote from the target pixel and slower decline when effects are scaled by area. It also demonstrates that on a per-halo basis, the estimation of the per-unit-area effect sizes is relatively consistent across the different combinations of halos used to fit the model, as shown in the supplementary information (table S7). The effect sizes used for prediction are, respectively $\beta_1 = -8.47 \times 10^{-2} \pm 1.17 \times 10^{-2}$, $\beta_2 = -1.50 \times 10^{-2} \pm 2.69 \times 10^{-3}$, $\beta_3 = -2.56 \times 10^{-3} \pm 4.63 \times 10^{-4}$ and $\beta_4 = -1.69 \times 10^{-4} \pm 2.02 \times 10^{-5}$. Figure 6(B) shows the estimated ΔAT for a 25 percentage point uniform forest loss scenario, for the same halo groupings as in figure 6(A). Note that the *x* axis, which is organized from highest to lowest ΔAT, is different from that in figure 6(A). These results suggest that whether unit change or uniform loss scenarios are considered, most of the effects of non-local temperature change area associated with length scales occurring within 10 km from the location of temperature observation.

The robust finding that non-local forest cover change impacts temperature at forested pixels at length scales of less than 50 km allows us to project the overall impacts of forest loss on temperature change at remaining forested sites in the Brazilian Amazaon-Cerrado region. As shown in table 2, we estimate that to date, forest loss within the study region for the 1985–2017 period resulted in $0.35\,°C \pm 0.12\,°C$ warming, with the majority of that ($0.20\,°C \pm 0.06\,°C$) arising from forest loss in the 2000–2015 period. Were the study region to experience a 25 percentage point decline in forest cover, non-local effects would result in 0.81 ± 0.23 warming.

These results are of similar magnitude to recent estimates of the effects of forest loss on the temperature of land where forest is lost at the Southern fringe of the Amazon biome, which, as described above, would suggest a $0.2\,°C$ increase in temperature due to local biogeophysical warming from forest loss, or a $0.5\,°C$ warming due to a hypothetical 25 percentage point decline in forest cover in the region. These two components of biogeophysical warming should be combined to estimate total biogeophysical warming, and suggest that, total warming due to forest cover conversion of the Amazon-Cerrado biome to date is on the order of $0.6\,°C$.

4. Discussion and conclusions

The research presented here aims to add to the growing understanding of biogeophysical climate change associated with forest cover loss in the tropics using the Brazilian Amazon-Cerrado region, the single largest concentration of forest loss in the past 20 years as a case study. Our approach is complementary to but distinct from several recent studies that demonstrated substantial temperature changes due to *local* forest cover loss (Li *et al* 2015, Silvério *et al* 2015, Alkama and Cescatti 2016) on LSTs, but which deliberately differenced out impacts of mesoscale or neighboring forest cover changes on temperature using paired location approaches. By focusing on temperature changes in undisturbed forested locations, we control for such local changes, and isolate the influence of spatially lagged forest cover changes. The results here are therefore not confounded by local forest cover losses and instead isolate the non-local effects of neighboring and regional forest cover changes.

That such non-local effects arise is strongly supported by the range of analyses undertaken in this study. We find a warming effect of non-local forest cover loss on undisturbed forested locations regardless of how we specify land cover change (e.g. as a weighted non-local index, or via a series of spatially discrete zones around the target pixel). The warming effect is robust to multiple model specifications and arises in both *in-situ* observations (albeit at spatially biased locations) and in remote-sensing derived but forest frontier datasets. Although there is non-trivial uncertainty regarding the magnitude of temperature change associated with specified land cover changes, this uncertainty is relatively small compared to the estimated effect sizes.

Further, the non-local warming identified in this study is non-trivial. It is comparable in magnitude to local warming due to forest loss, and it amplifies the effects of such local warming. Biogeophysical warming due to forest losses in Brazil will therefore occur due to both local warming, and to the influence of local warming on neighboring sites (non-local warming). This may mean that historical estimates of warming due to forest cover under-estimate warming (in studies where non-local effects are robustly controlled for), or mis-attribute warming to local change where non-local changes also influence observe warming signatures (in studies which do not robustly control for local versus non-local attribution of warming effects).

Non-local warming effects varied with the length scales considered. A precise identification of how effect size depends on distance is complicated by the finite set of halos explored in this study, the collinearity in forest cover loss change across these halos, and the confounding impact of increased total area versus decreased effect size with increasing distance from the target pixel. Nonetheless, whether focusing on absolute temperature change or effect size, the most important scales in this analysis seem to be 1–10 km from the target pixels, which is commensurate with studies regarding the scales of influence of urbanization on temperature (Zhou *et al* 2015). This further suggests that the truncation of the study's exploration of effects at 50 km from the target pixel, which was initially primarily based on unavoidable, excessive multicollinearity in forest loss at larger scales than 50 km, has presumably not strongly impacted the estimates of non-local warming. Over the studied period, 2000–2015, forest loss and agriculture were heavily concentrated in regions of low weather station density, particularly along the southern and eastern fringes of the Amazon Basin. Given that the 1–10 km scale may be particularly important for understanding non-local temperature change effects, the current *in-situ* monitoring stations and gridded products derived from them are too coarsely spaced to facilitate detection of non-local warming due to forest loss—particularly in the regions where forest loss has proliferated. This suggests an important role for remote sensing.

Nonetheless, the use of remotely sensed data to estimate AT engenders a number of limitations in the study. For example, the use of such data constrains the dataset to observations of clear-sky days, which are likely to be hotter and drier than the full daily maximum temperature distribution. This analysis may therefore be better suited to exploring changes in extreme heat, rather than the full AT distribution. Extreme heat, however, is an important constraint on ecosystem and agricultural productivity in the tropics (Challinor *et al* 2014), and impacts of increased extreme heat on productivity scale nonlinearly (Schlenker and Roberts 2009, Gourdji *et al* 2013, Lobell *et al* 2013).

The use of linear models to estimate temperature responses to non-local forest loss may be confounded by differential buffering of temperature extremes in areas with more forest cover (Winckler *et al* 2019), leading to potentially nonlinear non-local forest cover—temperature relationships. For these reasons we have avoided extrapolating the findings to extreme forest loss scenarios (e.g. to 100% forest removal), and confined prediction to changes that are broadly compatible with the range of forest cover change observed in the dataset, when linearization of the forest cover—temperature relationship should be less problematic.

Similar nonlinearities may impact the translation of AT change due to forest loss in forested locations to agricultural settings, as forested sites may alter evapotranspiration in response to changing temperatures, buffering observed changes relative to those that might be observed in more water limited (e.g. agricultural) conditions. There is also an important temporal component to this problem, as forest loss may be succeeded over time by land cover types (for example sugarcane) which may serve as substantial sources of evapotranspiration and thus cooling (Loarie *et al* 2011). In some cases (i.e. especially within short

periods of the year), agricultural land cover may fully offset the effect of local forest loss on local temperature (Luyssaert *et al* 2014). Thus, the magnitude of non-local temperature change due to forest cover losses may well be modulated by the rate and type of post-forest land cover transitions. Our approach measures the effect on temperature of forest loss given the post-forest land cover that resulted in the study region over the period of analysis. But if this composition changes over time then the effect size of forest loss on AT would change with it. Further considerations such as the spatial characteristics of forest loss, for example patch size (Pitman and Lorenz 2016), are similarly omitted from the present analysis, but may well influence the relationships under consideration. Finally, the study also avoids consideration of how changes in variability of access to soil moisture and groundwater, and changes in the dynamics of these hydrological resources under different land cover scenarios, might impact evapotranspiration and the land surface energy balance (Maxwell and Condon 2016). Therefore, future research should seek to quantify additional land use and land cover change influences on AT including non-forest vegetation, temporal and spatial patterns in forest loss, and hydrology.

ORCID iDs

Avery S Cohn https://orcid.org/0000-0002-6906-0488
Nishan Bhattarai https://orcid.org/0000-0003-2749-3549
David Dralle https://orcid.org/0000-0002-1944-2103

References

Abadie A, Athey S, Imbens G W and Wooldridge J 2017 When should you adjust standard errors for clustering? *When should you adjust standard errors for clustering? Working Paper No. 24003* National Bureau of Economic Research (http://nber.org/papers/w24003) (https://doi.org/10.3386/w24003)

Alkama R and Cescatti A 2016 Biophysical climate impacts of recent changes in global forest cover *Science* **351** 600–4

Allen R G, Pereira L S, Raes D and Smith M 1998 *Crop Evapotranspiration-Guidelines for Computing Crop Water Requirements* (*Fao Irrigation and Drainage* vol 56) (Rome: FAO)

Baidya Roy S and Avissar R 2002 Impact of land use/land cover change on regional hydrometeorology in Amazonia *J. Geophys. Res.: Atmos.* **107** LBA–4

Ban-Weiss G A, Bala G, Cao L, Pongratz J and Caldeira K 2011 Climate forcing and response to idealized changes in surface latent and sensible heat *Environ. Res. Lett.* **6** 034032

Bonan G B 2008 Forests and climate change: forcings, feedbacks, and the climate benefits of forests *Science* **320** 1444–9

Challinor A J, Watson J, Lobell D B, Howden S, Smith D and Chhetri N 2014 A meta-analysis of crop yield under climate change and adaptation *Nat. Clim. Change* **4** 287–91

Cohn A S, VanWey L K, Spera S A and Mustard J F 2016 Cropping frequency and area response to climate variability can exceed yield response *Nat. Clim. Change* **6** 601–4

Culf A, Fisch G and Hodnett M 1995 The albedo of Amazonian forest and ranch land *J. Clim.* **8** 1544–54

Curtis P G, Slay C M, Harris N L, Tyukavina A and Hansen M C 2018 Classifying drivers of global forest loss *Science* **361** 1108–11

D'Almeida C, Vörösmarty C J, Hurtt G C, Marengo J A, Dingman S L and Keim B D 2007 The effects of deforestation on the hydrological cycle in Amazonia: a review on scale and resolution *Int. J. Climatol.* **27** 633–47

Davin E L and de Noblet-Ducoudré N 2010 limatic impact of global-scale deforestation: radiative versus nonradiative processes *J. Clim.* **23** 97–112

Devaraju N, de Noblet-Ducoudré N, Quesada B and Bala G 2018 Quantifying the relative importance of direct and indirect biophysical effects of deforestation on surface temperature and teleconnections *J. Clim.* **31** 3811–29

Diffenbaugh N S and Burke M 2019 Global warming has increased global economic inequality *Proc. Natl Acad. Sci.* **20** 9808–13

Enfield D B, Mestas-Nuñez A M, Mayer D A and Cid-Serrano L 1999 How ubiquitous is the dipole relationship in tropical Atlantic sea surface temperatures? *J. Geophys. Res.: Oceans* **104** 7841–8

Feddema J J, Oleson K W, Bonan G B, Mearns L O, Buja L E, Meehl G A and Washington W M 2005 The importance of land-cover change in simulating future climates *Science* **310** 1674–8

Funk C, Peterson P, Landsfeld M, Pedreros D, Verdin J, Shukla S and Michaelsen J 2015 The climate hazards infrared precipitation with stations-a new environmental record for monitoring extremes *Sci. Data* **2** 150066

Giambelluca T W, Hölscher D, Bastos T X, Frazão R R, Nullet M A and Ziegler A D 1997 Observations of albedo and radiation balance over postforest land surfaces in the eastern Amazon basin *J. Clim.* **10** 919–28

Gibbs H K, Rausch L, Munger J, Schelly I, Morton D C, Noojipady P and Walker N F 2015 Brazila's soy moratorium *Science* **347** 377–8

Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D and Moore R 2017 Google earth engine: planetary-scale geospatial analysis for everyone *Remote Sens. Environ.* **202** 18–27

Gourdji S M, Sibley A M and Lobell D B 2013 Global crop exposure to critical high temperatures in the reproductive period: historical trends and future projections *Environ. Res. Lett.* **8** 024041

Hansen M C, Potapov P V, Moore R, Hancher M, Turubanova S, Tyukavina A and Townshend J 2013 High-resolution global maps of 21st-century forest cover change *Science* **342** 850–3

Hsiang S 2016 Climate econometrics *Annu. Rev. Resour. Econ.* **8** 43–75

Lawrence D and Vandecar K 2015 Effects of tropical deforestation on climate and agriculture *Nat. Clim. Change* **5** 27

Li Y, Zhao M, Motesharrei S, Mu Q, Kalnay E and Li S 2015 Local cooling and warming effects of forests based on satellite observations *Nat. Commun.* **6** 6603

Liu S, Su H, Tian J, Zhang R, Wang W and Wu Y 2017 Evaluating four remote sensing methods for estimating surface air temperature on a regional scale *J. Appl. Meteorol. Climatol.* **56** 803–14

Loarie S R, Lobell D B, Asner G P, Mu Q and Field C B 2011 Direct impacts on local climate of sugar-cane expansion in Brazil *Nat. Clim. Change* **1** 105–9

Lobell D B, Hammer G L, McLean G, Messina C, Roberts M J and Schlenker W 2013 The critical role of extreme heat for maize production in the United States *Nat. Clim. Change* **3** 497–501

Luyssaert S, Jammet M, Stoy P C, Estel S, Pongratz J, Ceschia E and Dolman A J 2014 Land management and land-cover change have impacts of similar magnitude on surface temperature *Nat. Clim. Change* **4** 389–93

Malyshev S, Shevliakova E, Stouffer R J and Pacala S W 2015 Contrasting local versus regional effects of land-use-change-induced heterogeneity on historical climate: analysis with the GFDL earth system model *J. Clim.* **28** 5448–69

Masuda Y J, Castro B, Aggraeni I, Wolff N H, Ebi K, Garg T and Spector J 2019 How are healthy, working populations affected by increasing temperatures in the tropics? Implications for climate change adaptation policies *Glob. Environ. Change* **56** 29–40

Maxwell R M and Condon L E 2016 Connections between groundwater flow and transpiration partitioning *Science* **353** 377–80

McNicol I M, Ryan C M and Mitchard E T 2018 Carbon losses from deforestation and widespread degradation offset by extensive growth in African woodlands *Nat. Commun.* **9** 3045

Meng G, Hall G B and Roberts S 2006 Multi-group segregation indices for measuring ordinal classes *Comput. Environ. Urban Syst.* **30** 275–99

Mildrexler D J, Zhao M and Running S W 2011 A global comparison between station air temperatures and MODIS land surface temperatures reveals the cooling role of forests *J. Geophys. Res.: Biogeosci.* **116** G03025

Oliveira L J, Costa M H, Soares-Filho B S and Coe M T 2013 Large-scale expansion of agriculture in Amazonia may be a no-win scenario *Environ. Res. Lett.* **8** 024021

Oliveira R, Bezerra L, Davidson E, Pinto F, Klink C, Nepstad D and Moreira A 2005a Deep root function in soil water dynamics in cerrado savannas of central Brazil *Funct. Ecol.* **19** 574–81

Oliveira R S, Dawson T E, Burgess S S and Nepstad D C 2005b Hydraulic redistribution in three Amazonian trees *Oecologia* **145** 354–63

Peng J, Liu Q, Wang L, Liu Q, Fan W, Lu M and Wen J 2015 Characterizing the pixel footprint of satellite albedo products derived from MODIS reflectance in the Heihe River Basin, China *Remote Sens.* **7** 6886–907

Pielke R A Sr, Pitman A, Niyogi D, Mahmood R, McAlpine C, Hossain F and de Noblet N 2011 Land use/land cover changes and climate: modeling analysis and observational evidence *Wiley Interdiscip. Rev. Clim. Change* **2** 828–50

Pitman A and Lorenz R 2016 Scale dependence of the simulated impact of Amazonian deforestation on regional climate *Environ. Res. Lett.* **11** 094025

Rayner N, Parker D E, Horton E, Folland C K, Alexander L V, Rowell D and Kaplan A 2003 Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century *J. Geophys. Res.: Atmos.* **108** 4407

R Core Team 2019 A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria (www.R-project.org/)

Rhode R, Muller R, Jacobsen R E M, Permutter S, Rosenfeld A and Wickham C 2013 A new estimate of the average earth surface land temperature spanning 1753 to 2011 *Geoinform. Geostat.* **1** 1–7

Rodell M, Houser P, Jambor U, Gottschalck J, Mitchell K and Meng C D T 2004 The global land data assimilation system *Bull. Am. Meteorol. Soc.* **85** 381–94

Schlenker W and Roberts M J 2009 Nonlinear temperature effects indicate severe damages to US crop yields under climate change *Proc. Natl Acad. Sci.* **106** 15594–8

Silvério D V, Brando P M, Macedo M N, Beck P S, Bustamante M and Coe M T 2015 Agricultural expansion dominates climate changes in southeastern Amazonia: the overlooked non-GHG forcing *Environ. Res. Lett.* **10** 1–9

Souza C and Azevedo T 2017 *MapBiomas General Handbook* (São Paulo: MapBiomas) pp 1–23

Strassburg B B, Brooks T, Feltran-Barbieri R, Iribarrem A, Crouzeilles R, Loyola R and Balmford A 2017 Moment of truth for the cerrado hotspot *Nat. Ecol. Evol.* **1** 1–3

Winckler J, Lejeune Q, Reick C H and Pongratz J 2019 Nonlocal effects dominate the global mean surface temperature response to the biogeophysical effects of deforestation *Geophys. Res. Lett.* **46** 745–55

Winckler J, Reick C H and Pongratz J 2017 Robust identification of local biogeophysical effects of land-cover change in a global climate model *J. Clim.* **30** 1159–76

Xavier A C, King C W and Scanlon B R 2016 Daily gridded meteorological variables in Brazil (1980-2013) *Int. J. Climatol.* **36** 2644–59

Zhou D, Zhao S, Zhang L, Sun G and Liu Y 2015 The footprint of urban heat island effect in China *Sci. Rep.* **5** 11160

Ziter C D, Pedersen E J, Kucharik C J and Turner M G 2019 Scale-dependent interactions between tree canopy cover and impervious surfaces reduce daytime urban heat during summer *Proc. Natl Acad. Sci.* **15** 7575–80